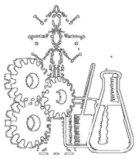
Modelling and Control of Activated Sludge Processes

Michela Mulas



Dottorato di Ricerca in Ingegneria Industriale
Università degli Studi di Cagliari
XVIII Ciclo

Modelling and Control of Activated Sludge Processes

Michela Mulas

Supervisors:
Prof. Roberto Baratti
Ing. Stefania Tronci

Dottorato di Ricerca in Ingegneria Industriale
Università degli Studi di Cagliari
XVIII Ciclo

Acknowledgements

This work was carried out in the *Laboratory of process modelling and control* in the Chemical and Materials Engineering Department, in Cagliari, between 2002 and 2005.

I wish to thank my supervisor Prof. Roberto Baratti for the opportunity he gave me to work in the wastewater treatment field, for the freedom he gave me to participate to conferences and for everything I learned during these years. I wish to express my gratitude to Stefania Tronci for sharing her insight and experience and being a friend. I wish also to thank the present and former members of the *Laboratory* for making a nice working environment.

A warm thank you to Prof. Sigurd Skogestad and all the people in his group for the interesting discussions and the kindness during my stay in Norway.

Finally I would like to thank my dearest parents for their encouragement and Francesco for his patience and support during writing this thesis.

Cagliari, January 2006

Michela Mulas

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The progressive deterioration of water resources and the large amount of polluted water generated in industrialized societies gives Wastewater Treatment (WWT) processes a fundamental importance in the water prevention. New guidelines and regulations (the Directive 91/271/CEE referring to the European countries) enforce the adoption of specific qualility indexes for the treated wastewater. Taking into account current environmental problems, it is not unrealistic to believe that this trend will continue. At the same time loads on existing plants are expected to increase due to growth of urban areas. This situation demands more efficient treatment procedures for wastewater.

Inside a biological wastewater treatment plant, the *Activated Sludge Process* (ASP) is the most commonly used technology to remove organic pollutant from wastewater, even if the process was developed in the early 20^{th} century. This is because it is the most cost-effective, it is very flexible (it can be adapted to any kind of wastewater), it is reliable and has the capacity of producing high quality effluent. For all these reasons, our interest during this work will focus on demonstrating that a better understanding and management of the process can lead to benefits for the overall wastewater treatment plant.

1.1 Motivations

Several motivations may be found to explain why we are interested in activated sludge processes. We can classify them in two main categories:

- Environmental motivation. Because water is something special and we want to preserve it!

- Economic motivations. Because a wastewater treatment plant can be considered as the largest industry in terms of raw material treated. Therefore, we want such an industry to work always near to its maximum efficiency. Mostly because, in some countries recent evolution of the legislation concerning surface or groundwater use is such that total recycling of process water has become an issue.

In such a context, the wastewater treatment becomes part of a production process where the quality control of the effluent is very important since poor operation of the treatment process can lead to important production losses and environemental problems.

1.2 Objectives

In order to explain the objectives of this work, we consider the *building blocks* idea exposed by Vanrolleghem in [112]. The objective is to introduce benefits to the ASP improving each block in the general control loop structure in Figure 1.1.

The first block represents the *process*. The knowledge about the physical and biochemical behavior of the activated sludge process can be summarized in a proper process model. In order to better understand the process itself and also to obtain a good platform to explore future developments on the process, as an improve we should develop an accurate process model.

The second block regards *sensors*. Until very recently this was one of the main bottlenecks in improving wastewater treatment automation. We can think to improve this block introducing *soft-sensors*, that use the available (few) measures to reconstruct the time evolution of the unmeasured variables.

The third block is the *control system*. This block can be improved by defining an adequate control structure before designing the controller itself. With simple considerations on the control structure design, we might obtain the controlled variables that lead to an optimal plant operation.

The last block represents the *actuators* which implement the controller outputs on the plant. A limited choice of control actuators is available: most of them are valves, pumps and so on. As a matter of fact, we can improve this block only improving the controller itself.

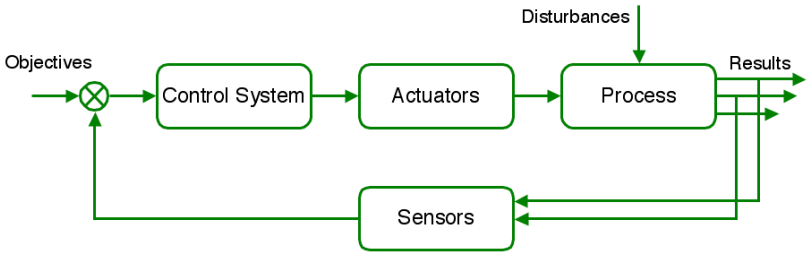


Figure 1.1: General control loop structure

1.3 Thesis Overview

This thesis deals with modelling and control of activated sludge processes and the associated issues. The thesis contains two main parts: the first part deals with the controlled variables selection on the activated sludge process, and the second part deals with model order reduction and software sensor design. Referring to Section 1.2, the third block is considered in the first part, and the second block in the second part. The first block is analyzed in the introductory Chapter 2.

A brief overview of each Chapter is given below.
In **Chapter 2** the ASP model is illustrated. Since the process is essentially composed of two main units (a bioreactor and a settler linked together with a recycle line), a model for each unit is defined. An overall process model is developed by using the commercial software GPS-X^{TM}, and Matlab^{TM}/Simulink^{TM}.

In **Chapter 3** the idea of self-optimizing control and controlled variable selection is illustrated.

In **Chapter 4**, the controller structure design is applied to an ASP. On this chapter is also based a paper presented at the ICheaP-7 Conference in May 2005 ([80], [81]).

In **Chapter 5**, the model reduction techniques are exposed focusing on those that are applied to obtain a reduced order model of the ASP.

In **Chapter 6** some applications with existing models found in literature are exposed. Furthermore, reduced models based on modal and balanced reduction techniques are proposed.

In **Chapter 7** the reduced models are used in the design of different software sensor: a Luenbergher-like observer, an extend Kalman filter and a nonlinear geometric observer. Several papers were based on this subject [71], [110], [82], [79].

Chapter 2

ASP Models and Simulations

In this Chapter we describe the mathematical modelling approach used to represent an Activated Sludge Process (ASP) by means of computer simulations. The ASP is the most widely used biological treatment of liquid waste, essentially because it is a cheap technology which can be adapted to any kind of wastewater. In the activated sludge process, a bacterial biomass suspension (the activated sludge) is responsible for the removal of pollutants. Depending on the design and the specific application, an activated sludge wastewater treatment plant can achieve biological nitrogen removal and biological phosphorus removal, plus the removal of organic carbon substances. Many different activated sludge process configurations have evolved during the years: Jeppsson [52] provides an exhaustive review on the historical evolution of the activated sludge process. In this Chapter and also in our entire work the traditional ASP configuration, involving a bioreactor followed by a settler with recycle, is considered.

This Chapter is organized as follows. In Section 2.1 a brief description of the activated sludge process is given, and because we consider only nitrogen removal phenomena an explanation of that process is also given (Section 2.1.1). This provides the basis to understand the process model formulation for the bioreactor part, and the secondary setter part. Since, the biological reactor (Section 2.2.1) and the settler tank (Section 2.2.2) are interacting, because of the recycle flow, Section 2.2.3 shows how to integrate the two units in the activated sludge process. The overall goal of this Chapter is to implement a computer simulation of this biological process. To achieve this two different approache are taken: the first one employs a commercial software, the GPS-X^{TM}(Section 2.3.1), and the second one uses Matlab^{TM}/Simulink^{TM}(Section 2.3.2).

2.1 The Activated Sludge Process

In Figure 2.1 the basic layout for the considered activated sludge process is shown: from the secondary settler, the sludge is partially recirculated to the bioreactor (*Returned Activated Sludge*, RAS) and partially wasted as excess sludge (*Waste Activated Sludge*, WAS). The ASP is a biological process in which microorganisms oxidize and mineralize organic matter. The microorganisms in the activated sludge are mainly bacteria, which can be found also in the raw wastewater incoming into the plant. The composition and the species depend not only on the influent wastewater but also on the design and operation of the wastewater treatment plant.

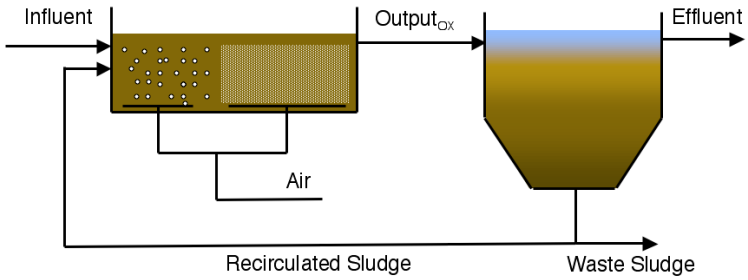


Figure 2.1: Simple configuration for the activated sludge process

Bacteria constantly need energy in order to grow and to support essential life activities. Growing cells utilize substrate and nutrients located outside the cell membrane for growth and energy in a process. Oxygen is used by microorganisms to oxidize organic matter. Some bacteria can use oxygen either as dissolved oxygen or not: these bacteria are called *heterotrophs*. They represent the major part of bacteria in activated sludge and use organic carbon in the form of small organic molecules as substrate. Other essential bacteria for the activated sludge process are *autotrophs*. They can growth only with dissolved oxygen and use inorganic carbon as substrate. To maintain the microbiological population, sludge from the settler is recirculated to the aerated tank. The bacteria growth and particulate inert matter is removed from the process as *waste sludge*.

In order to schematically show the biological process renewal, we refer to Figure 2.2 [68]. Organic matter enters the plant in several different forms and is converted to other forms by biological processes. Firstly, the *hydrolysis* process transforms larger organic matter into more easily accessible molecules (readily biodegradable matter). The biomass growth rate depends

on many variables, such as the amount of biomass, substrate, temperature, pH, etc. During the microorganisms decay, biologically inert (nonbiodegradable matter) is produced: this is also present in the incoming wastewater and such matter remains unaffected through the process to be collected and removed in the settler.

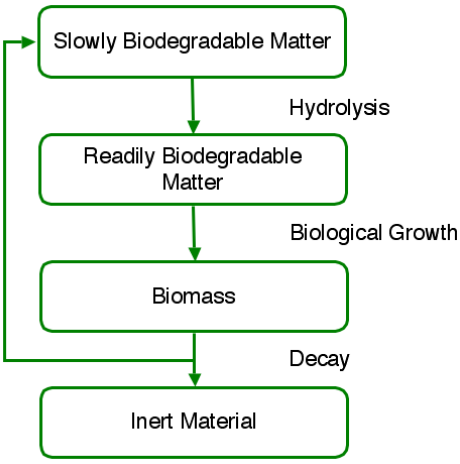


Figure 2.2: Biological process renewal scheme

2.1.1 Biological Nitrogen Removal

Nitrogen appears in wastewater in several forms e.g., as free and ionized ammonia (NH_3 and NH_4^+), nitrate (NO_3^-), nitrite (NO_2^-) and as organic compounds. The different forms constitute the total nitrogen content. Although nitrogen is an essential nutrient for biological growth and is one of the main constituents in all living organism, an excessive presence of it in the effluent wastewater should be avoided for several reasons. Both because ammonia is toxic for aquatic organisms such as fishes and because as nitrate, it can produce an excessive oxygen consumption in the receiving water. Being nitrogen a nutrient, aquatic plants can also growth "without" limit when its level is so high to cause eutotrophic phenomena.

When untreated wastewater arrives to the plant, most nitrogen is present in the ammonia form, which can be removed in a two-step procedure. In the first step ammonia is oxidized to nitrate in *aerobic* conditions, this process is called *nitrification* and can be described by the following simplified chemical

reacting scheme:

$$NH_4^+ + 1.5O_2 \rightarrow NO_2^- + H_2O + 2H^+ \quad (2.1)$$
$$NO_2^- + 0.5O_2 \rightarrow NO_3^- \quad (2.2)$$

That is, the ammonia is firstly oxidized by nitrite and then to nitrate. The bacteria involved in these reactions require oxygen to perform the process.

The nitrates are then converted to nitrogen by *denitrification*. This process occurs in *anoxic* environment: i.e., oxygen is mainly present as nitrate, and the bacteria responsible for it "respire" oxygen as nitrate instead of dissolved oxygen. The following simplified reaction scheme shows the basic process:

$$2NO_3^- + 2H^+ \rightarrow N_2 + H_2O + 2.5O_2$$

By nitrification and denitrification, nitrogen is removed from wastewater biologically. This means that anoxic zones are necessary for denitrification, whereas aerobic zones are necessary for nitrification. Anoxic zones can be placed either at the beginning of the tank (*pre-denitrification*) or at the end of the process (*post-denitrification*). During this work, an anoxic zone followed by an aerobic zone is considered (Figure 2.1).

2.2 Modelling ASP

As reported in the previous section, the activated sludge process is composed of two main units: a biological reactor and a settler. In this section the models employed for the simulation are illustrated.

2.2.1 Bioreactor Model

There are several models describing the biological process in the activated sludge plant, the developments in the family proposed by the International Water Association (IWA) represent a major contribute:

- **ASM1**, the Activated Sludge Process Model No.1 [43] can be considered as the reference model since this model triggered the general acceptance of the biological process modelling. ASM1 was primarily developed to describe the removal of organic compounds and nitrogen with simultaneous consumption of oxygen and nitrate as electron acceptor. The model, furthermore, aims at yielding a good description of the sludge production. COD (*Chemical Oxygen Demand*) was adopted as the measure of the concentration of organic matter.

- **ASM2**, the Activated Sludge Process Model No.2 [41] extends the capabilities of the ASM1 to the description of bio-phosphorus.
- **ASM2d**, the Activated Sludge Process Model No.2d [42] is built on the ASM2 model adding the denitrifying activity of PAOs[1] to allow a better description of the dynamics of phosphate and nitrate.
- **ASM3**, the Activated Sludge Process Model No.3 [36] was also developed for biological nitrogen removal, with basically the same goal as the ASM1. The major difference between the ASM1 and the ASM3 models is that the latter recognizes the importance of storage polymers in the heterotrophic activated sludge conversion.

In this work the adopted model is the ASM1, which will be briefly explained in this section. In Appendix A a complete *Petersen Matrix* for the ASM1 model is reported. For a full description of the model, the orginal IWA report [43] should be consulted. Generally speaking, the model consists of 13 components or state variables (Table 2.1) involved in 8 reactions (Table 2.2).

State Variables

as mentioned, the ASM1 model allows us to describe phenomena of organic matter and nitrogen removal. In fact, the main classification in the model state variables is in *organic matter*, expressed in terms of COD, and *nitrogen compounds* (Table 2.1).

The organic matter is further divided into biodegradable COD (S_S and X_S), nonbiodegradable COD (S_I and X_I) and active biomass (X_{BH} and X_{BA}). The readily biodegradable substrate is assumed to be made up of simple soluble molecules that can be easily absorbed by the organisms and metabolized for energy and synthesis. In contrast, slowly biodegradable substrate consists of relatively complex molecules that require enzymatic breakdown prior to absorption and utilization. Nonbiodegradable organic matter is biologically inert and passes through the system without change in form. It can be soluble (S_I) which leaves the process at the same concentration as it enters and particulate (X_I), becoming enmeshed in the activated sludge and leaving the system mainly as the wastage flowrate. Moreover, the biomass is divided into heterotrophic (X_{BH}), and autotrophic biomass (X_{BA}). As an extra component X_P is included to take into account the inert particulate arising from cell decay.

[1] Polyphosphate Accumulating Organisms

State Variable	ASM1 Notation	
Soluble inert organic matter	S_I	$gCOD/m^3$
Readily biodegradable substrate	S_S	$gCOD/m^3$
Particulate inert organic matter	X_I	$gCOD/m^3$
Slowly biodegradable substrate	X_S	$gCOD/m^3$
Active heterotrophic biomass	X_{BH}	$gCOD/m^3$
Active autotrophic biomass	X_{BA}	$gCOD/m^3$
Part. prod. from biomass decay	X_P	$gCOD/m^3$
Dissolved Oxygen	S_O	gO_2/m^3
Nitrite and Nitrate Nitrogen	S_{NO}	gN/m^3
Free and Ionized Ammonia	S_{NH}	gN/m^3
Soluble biodegr. organic N	S_{ND}	gN/m^3
Part. biodegr. organic N	X_{ND}	gN/m^3
Alkalinity	S_{ALK}	Molar units

Table 2.1: ASM1 State Variables

As for the organic part, nitrogenous matter can be divided into two categories: nonbiodegradable and biodegradable. With respect to the biodegradable part, the particulate portion is associated to the nonbiodegradable particulate COD. The soluble portion is usually negligible and is not incorporated into the model. The biodegradable matter is divided into free and ionized ammonia (S_{NH}), soluble organic nitrogen (S_{ND}), and particulate organic nitrogen (X_{ND}). The last is hydrolyzed to soluble organic nitrogen in parallel with the hydrolysis of slowly biodegradable organic matter, whereas, the soluble organic nitrogen is acted by heterotrophic material and converted into ammonia. For the sake of simplicity, the autotrophic conversion of ammonia to nitrate is considered to be a single step process which requires oxygen. This means that a unique state variable (S_{NO}) represents nitrate/nitrogen compounds in the activated sludge system.

Furthermore, one variable is included to represent the dissolved oxygen consumption in the activated sludge system, S_O. Even if inclusion of alkalinity, S_{ALK}, in the conversion process is not essential, its inclusion in the model is also desiderable because it provides information whereby undue changes in pH can be predicted.

The state variables included in the ASM1 are the fundamental components that act upon the process, but they are not always measurable or interpretable in many practical applications. Therefore, some *composite vari-*

ables can be calculated from the state variables in order to combine them into forms that are typically measured in reality, such as COD (*Chemical Oxygen Demand*), TSS (*Total Suspended Solids*) and TN (*Total Nitrogen*), as reported below:

$$\begin{aligned} COD =& S_I + S_S + X_I + X_S + X_{BH} + X_{BA} + X_P \ [gCOD/m^3] \\ TSS =& 0.75(X_S + X_P + X_I) + 0.9(X_{BH} + X_{BA}) \ [gSS/m^3] \\ TN =& S_{NO} + S_{NH} + S_{ND} + X_{ND} + i_{XB}(X_{BH} + X_{BA}) \\ & + i_{XP}(X_P + X_I) \ [gN/m^3] \end{aligned} \tag{2.3}$$

The conversion coefficients 0.75 $[gSS/gCOD]$ for the inert and particulate material and 0.9 $[gSS/gCOD]$ for the heterotrophic and autotrophic biomass, have been proposed in [43] and also reported in [54]. The parameters i_{XB} and i_{XP} $[gN/gCOD]$ are reported in Table 2.3.

ASM1 Processes

Two types of microorganisms carry out the reactions: heterotrophs and autotrophs. Here, we briefly describe the different mechanisms (Table 2.2) incorporated in the ASM1.

- The *aerobic growth of heterotrophs* occurs at expense of soluble substrate utilizing oxygen and results in a production of heterotrophic biomass. The growth is modelled using Monod kinetics, which are assumed to be subject to double nutrient limitation their rate depending on the concentration of both S_S and S_O. This process is, generally, the main contributor to the production of new biomass and removal of COD. Ammonia is used as nitrogen source for synthesis and is incorporated into the cell mass.

- The *anoxic growth of heterotrophs* occurs in absence of dissolved oxygen with nitrate as the terminal electron acceptor, with S_S the substrate and resulting in heterotrophs biomass. The same Monod kinetics used in the aerobic growth are applied, except that the maximum rate of substrate is less under anoxic conditions. For this reason, the kinetic rate expression is multiplied by a factor $\eta_g < 1$. Ammonia serves as nitrogen source for cell synthesis.

- In *aerobic growth of autotrophs*, S_{NH} serves as the energy source for growth of the nitrifiers, resulting in autotrophic cell mass and nitrate nitrogen as products. This process is associated to the oxygen demand and once again the growth rate is modelled using Monod kinetics.

- The *decay of heterotrophs* is modelled on the death-regeneration approach proposed by Dold *et al.* [25]. The organisms die at a certain rate and a portion of the material is considered to be nonbiodegradable adding up to the X_P fraction. The remainder adds up to X_S. Organic nitrogen associated with X_S becomes available as particulate organic nitrogen.

- The *decay of autotrophs* takes exactly the same modelling approach as the decay of the heterotrophs.

- The *ammonification of soluble organic nitrogen* regards the conversion of S_{ND} into S_{NH} by a first order process mediated by active heterotrophs.

- In the *hydrolysis of entrapped organics*, slowly biodegradable substrate trapped in the sludge mass is broken down, producing S_S for the organisms to growth. The process is modelled on the basis of reaction kinetics and occurs in aerobic and anoxic environments. The rate of hydrolysis is reduced under anoxic conditions compared to aerobic conditions by a factor $\eta_h < 1$.

- In the *hydrolysis of entrapped organic nitrogen*, X_{ND} is broken down to soluble organic nitrogen at a rate defined by the hydrolysis reaction for entrapped organics.

Process	Basic Reaction
Aerobic growth of heterotrophs	$S_S + S_O + S_{NH} \rightarrow X_{BH}$
Anoxic growth of heterotrophs	$S_S + S_{NO} + S_{NH} \rightarrow X_{BH}$
Aerobic growth of autotrophs	$S_O + S_{NH} \rightarrow X_{BA} + S_O$
Decay of heterotrophs	$X_{BH} \rightarrow X_P + X_S + X_{ND}$
Decay of autotrophs	$X_{BA} \rightarrow X_P + X_S + X_{ND}$
Ammonification of soluble organic N	$S_{ND} \rightarrow S_{NH}$
Hydrolysis of entrapped organics	$X_S \rightarrow S_S$
Hydrolysis of entrapped organic N	$X_{ND} \rightarrow S_{ND}$

Table 2.2: ASM1 Basic Processes

It should be noted that S_I and X_I are not included in any conversion process. Nevertheless they must be considered because important to the performance of the process, being included in the COD computation.

As useful feature in the ASM1 model, we refer to the introduction of the *switching functions* concept that allows to turn process rate equations on and off as environmental conditions are changed. This was particularly necessary for processes that depend upon the type of electron acceptor present. For example, the bacteria that are responsible for nitrification are capable to grow only under aerobic conditions and their rate of growth falls to zero as the dissolved oxygen concentration approaches zero (regardless to the concentration of their energy yielding substrate). This is modelled in the ASM1 by including an oxygen *switch* in the process rate equations. Whenever the aerobic conditions need to be modelled, the oxygen switching function adopted is:

$$\frac{S_O}{K_{OH} + S_O}$$

where small value of K_{OH} (explained in the Table 2.3) implies that the value of the switching function is near unity for moderate dissolved oxygen concentrations and decreases to zero as the oxygen approaches zero. Similarly, processes which occur only when dissolved oxygen is absent may be turned on by a switching function in the form:

$$\frac{K_{OH}}{K_{OH} + S_O}$$

The coefficient K_{OH} has the same value as in the expression for aerobic growth so that as the aerobic growth declines, the anoxic growth increases.

Parameters

In Table 2.3 the kinetic and stoichiometric parameters are reported. The parameters selection of a mathematical model is known as model calibration, and as consequence of high interdependence of the state variables, troublesome nonlinearities, lacking identifiably and verifiability, the calibration of the model can be difficult and laborious. The calibration task becomes very hard especially because the data collectable from wastewater treatment plants are generally very sparse and not always reliable.

2.2.2 Secondary Settler Model

Activated sludge plants transform organic matter into biomass. The effective operation of the process requires the biomass to be removed from the liquid stream (in the secondary settler) prior to being discharged in the receiving waters. The sedimentation of the particles in the liquor is achieved by gravity along with the density differences between the particles and the liquid. Part of the biomass is purged, while a large fraction is returned

ASM1 parameter	Symbol	20 ^0C	10 ^0C	*Literature*	Unit
Heterotrophic Yield	Y_H	0.67	0.67	0.38-0.75	$g(cellCOD\,formed)/g(COD\,oxidized)$
Autotrophic Yield	Y_A	0.24	0.24	0.07-0.28	$g(cellCOD\,formed)/g(N\,oxidized)$
Fraction of biomass yielding part. prod.	f_P	0.08	0.08	–	dimensionless
(Mass N)/(Mass COD) in biomass	i_{XB}	0.086	0.086	–	$gN/gCOD$
(Mass N)/(Mass COD) prod. from biomass	i_{XP}	0.06	0.06	–	$gN/gCOD$
Heterotrophic max. specific growth rate	μ_H	6.0	3.0	0.6-13.2	$1/d$
Half Saturation Coeff. (hsc) for heterotrophs	K_{SH}	20.0	20.0	5-225	$gCOD/m^3$
Oxygen hsc for heterotrophs	K_{OH}	0.20	0.20	0.01-0.20	gO_2/m^3
Nitrate hsc for heterotrophs	K_{NO}	0.50	0.50	0.10-0.50	gNO_3-N/m^3
Heterotrophic decay rate	b_H	0.62	0.20	0.05-1.60	$1/d$
Correction factor for growth for het.	η_g	0.80	0.80	0.60-1.0	dimensionless
Autotrophic max. specific growth rate	μ_A	0.80	0.30	0.20-1.0	$1/d$
Ammonia hsc for autotrophs	K_{NH}	1.0	1.0	–	gNH_3-N/m^3
Oxygen hsc for autotrophs	K_{OA}	0.40	0.40	0.40-2.0	gO_2/m^3
Autotrophic decay rate	b_A	0.20	0.10	0.05-0.20	$1/d$
Ammonification rate	k_a	0.08	0.04	–	$m^3/gCOD/d$
Max. specific hydrolisys rate	k_h	3.0	1.0	–	$g(slowly\,biodegr.COD)/g(cellCOD)/d$
Hsc for hydrolysis of slowly biodegr. sub.	K_X	0.03	0.01	–	$g(slowly\,biodegr.COD)/g(cellCOD)/d$
Correction factor for anoxic hydrolysis	η_h	0.40	0.40	–	dimensionless

Table 2.3: Stoichiometric and kinetic parameters in the ASM1 model [43]

to the biological reactor to maintain the appropriate substrate-to-biomass ratio. This means that the settler combines functions of clarification and thickening into one unit, as shown in Figure 2.3. In some cases the settler model can perform more important tasks, such as *sludge storage*[2] or *reactions*[3]. However, those task have not been considered in this study.

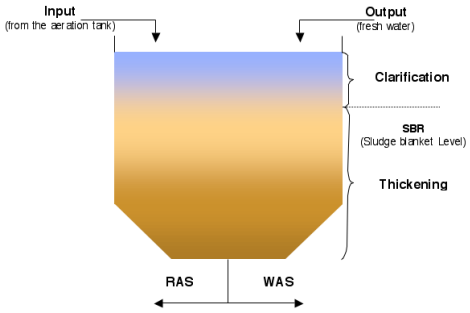


Figure 2.3: Secondary settler processes

The complex behavior of the secondary settler and its importance for the successful operation of the ASP have made the settling process a great challenge for researchers working in the field of mathematical modelling. For that reason, different models are present in literature and a good review on the different modelling approaches is given in [52]. To investigate the functioning of the settler, the IWA report [26] should also be consulted.

For the secondary settler a one-dimensional model approach has been considered in this study. It is assumed that in the settler the profiles of horizontal velocities are uniform and that horizontal gradients in concentration are negligible. Consequently, only the processes in the vertical dimension are modelled. The model considers only one state variable for the all particulate components (the solids concentration) and all the soluble state variables, leaving the settler without settling. In Figure 2.4, the simplified flow scheme is reported. As we can note, at the inlet section the inflow and the solids concentration are homogeneously spread over the horizontal cross section, and the incoming solids are distribuited uniformly and instantaneously across the entire cross-sectional area. The flow is divided into a downward flow towards the underflow outlet at the bottom, and an upward

[2] In the bottom part of the settler, sludge is stored for subsequent use under high waste load conditions.

[3] Where additional aerobic conversion can occur or where denitrification may take place [95].

flow towards the effluent exit at the top.

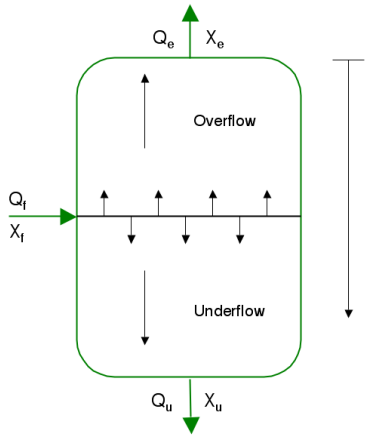


Figure 2.4: Flow-scheme of 1D continuous-flow settling tank approach [26]

The used model is based on the Vitasovic's approach [114] where the settler is divided into a number of layers of constant thickness (as shown in Figure 2.5) and a solids balance is performed on each layers. The model is based on the solid flux concept, which states that the solids entering the secondary settler are carried to the bottom via the gravity settling flux (J_s) and the bulk flux (J_b). The first result from the sludge settling downward through the water, whereas the second results from the water moving downward in the settler owing to the underflow sludge recycle pump. The total flux is given by:

$$J = J_s + J_b \tag{2.4}$$

Five different groups of layers are represented in the Vitasovic's model depending on their position relative to the feed point: the top layer, the layers above the feed point, the feed layer, the layers below feed point and the bottom layer. The solid flux due to bulk movement of the liquid is a straightforward calculation based on the solids' concentration times the liquid bulk velocity, which may be either upward or downward depending on its relative position to the feed layer. The solids flux is due to a specified exponential settling function applicable to both hindered sedimentation and flocculant sedimentation conditions. This means that several things need to be considered for the settler. Of course, we write the mass balances for each layer, but we also consider the solid fluxes between them and the general upward and downward flows.

The solids flux due to the bulk movement of the liquid is also straightforward to asses, being equal to the product of the solids concentration, X, and the bulk velocity of the liquid. The velocity can be downward (v_{dn}) or upward (v_{up}) depending on the position of the layer with respect to the feed point:

$$\begin{aligned} v_{dn} &= \frac{Q_u}{A} = \frac{Q_r + Q_w}{A} \\ v_{up} &= \frac{Q_e}{A} \end{aligned} \tag{2.5}$$

where A is the settler cross-sectional area, Q_u is the under flowrate (with Q_r and Q_w as recycled and wasted flowrates, respectively) and Q_e is the effluent flowrate.

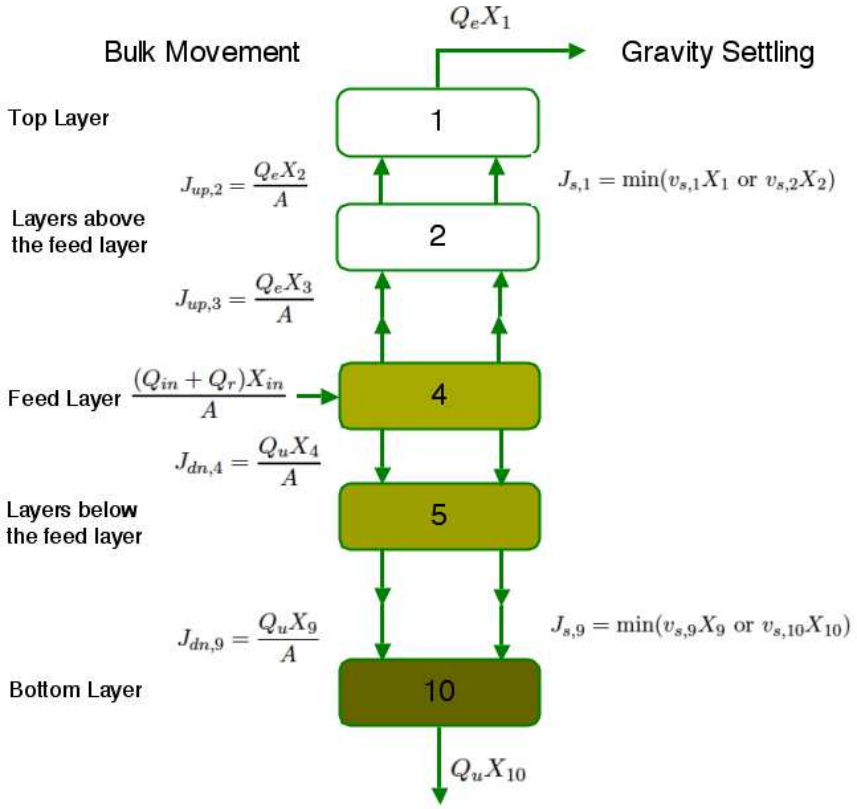


Figure 2.5: Layered settler model

The solids flux, due to gravity settling of the solids particles, is given as the product of the concentration, X and the settling velocity of the solids

particles, v_s. The determination of an appropriate settling velocity is indispensable to model the secondary settler. Several models have been proposed relating v_s to the solids concentration either by a power function or more often by an exponential function of X. In the Vesilind expression $v_s = v_0 e^{-cX}$ the constants v_0 (the maximum settling velocity) and c (a model parameter) need to be calibrated individually for each layer. Moreover, Cho *et al.* [18] derived their settling function from the analytical description of flow in porous media; they also give an exstensive review and comparison between different models in their paper.

In this work, the considered settling velocity is the double exponential velocity expression proposed by Takács *et al.* [107], that can be computed for each layer j as follows:

$$\begin{aligned} & v_{sj} = v_0 e^{-r_h(X_j - X_{min})} - v_0 e^{-r_p(X_j - X_{min})} \ [m/d] \\ & 0 \leq v_{sj} \leq v_0^{'} \end{aligned} \tag{2.6}$$

where:

- v_0 is the maximum theoretical settling velocity in $[m/d]$;
- $v_0^{'}$ is the maximum practical settling velocity in $[m/d]$;
- r_h is the settling parameter characteristic of the hindered settling zone, in $[m^3/d]$;
- r_p is the settling parameter characteristic of low solids concentrations, expressed in $[m^3/d]$;
- $X_{min} = f_{ns} X_{in}$ is the minimum attainable suspended solids concentration, in $[gSS/m^3]$, with X_{in} is the mixed liquor solids entering the settler and f_{ns} is the non settleable fraction of X_{in}.

The expression allows to represent both thickening and clarification phenomena, since the term $v_0 e^{-r_h(x_j - X_{min})}$ reflects the settling velocity of the large flocculating particles and the term $v_0 e^{-r_p(x_j - X_{min})}$ takes into account the smaller settling particles velocity. Once, the velocities have been defined the solids mass balances around each layers come straightforward.

In this approach also the number of layers becomes an important parameter, especially when diffusion phenomena are considered: for instance, Jeppsson and Diehl [53] recommended a model with 30-50 layers for diffusion. However, the introduction of a diffusion term changes the ordinary differential equations to partial differential equations, increasing the computational complexity. In fact, we found that for our purpose a 10 layers Tackas model gives a good representation of the settler behavior.

2.2.3 Coupling Bioreactor and Secondary Settler

A significant problem in the evaluation of the activated sludge process is the difficulty to separating the dynamics of the biological reactor from the settler, because of the recycle flow [23]. From a modelling point of view, the components of the wastewater are described differently for the biological reactor and the secondary settler. The ASM1 is based on 13 types of components: the settler model only considers the total concentrations of the particulate and soluble material (which leaves the settler without settling). Therefore, all particulate components of the ASM1 are lumped into a single variable X, when entering the settler and the reversed process is performed at the outlets. A problem with this amalgamation arises from the different units used in the ASM1 and the settler model. In the settler, the unit for describing the material is $g(SS)/m^3$. In the ASM1, the density unit of all organic material is $g(COD)/m^3$ and the overall fraction of nitrogen is gN/m^3. The particulate material is affected by gravity and the true mass for all components must be known. The conversion coefficients for the inert particulate and the slowly biodegradable material and also for the heterotrophic and autotrophic biomass are reported in the TSS expression 2.3.

It should be noted that the particulate biodegradable organic nitrogen, X_{ND} $[gN/m^3]$, should not be included in the transformation into X even if it is modelled as a particulate material in the ASM1 model. This is because it is a subset of the other particulate components and is already included in their concentration [43].

No conversion factors are needed for the soluble components, since they only follow the water flows and no biological reactions are assumed to occur in the settler.

2.3 Simulating ASP

The acquired knowledge about the activated sludge process can now be used to implement an appropriate dynamical simulations of it. By using simulations, we can study the effect of different environmental conditions, test the system sensitivity to different parameters and apply different control configurations and so on. Information about specific ASP simulators are given by Olsson and Newell in [86] and also by Copp in [20].

From a practical stand point, a simulation can be obtained in different ways either using commercial software or implementing the ASP model in

programming environment like Matlab/Simulink. In the following of this section a description of the two approaches used in this work is given.

2.3.1 Commercial Software

Specific commercial environments usually contain extended libraries of predefined process models offering the representation of the whole wastewater treatment plant. The process configuration to be simulated can be easily-constructed by connecting process units blocks and pop-up windows allow modifying the model parameters. Many examples of commercial simulators exist. Among them we can list, for instance:

- *WEST* (Wastewater treatment plant Engines for Simulation and Training): an interactive dynamic simulator. It is developed mainly at the University of Gent, Belgium and current information about the software can be found on `http://www.hemmis.com/`.

- *SIMBA* (SIMulation programms für die Biologische Abwasserreinigung): developed at the Institut für Automation und Kommonikation (IFAK) in Germany. It can be considered a custom made version of Simulink for wastewater treatment applications. A more extensively compend about the simulator can be found on `http://simba.ifak-md.de/simba/`.

- *EFOR* is a stand-alone software package for the simulation of complete wastewater treatment plant. It is developed mainly at the Danish Technical University. The present progress can be found on `http://www.dhisoftware.com/efor/`.

GPS-X™

We mainly want to focus on the GPS-X, since Version 4 has been exstensively used in this work. GPS-X is a modular multi-purpose modelling environment for the simulation of municipal and industrial wastewater treatment plant. It uses an advanced graphical user interface to facilitate dynamic modelling and simulation. It can virtually cover all of the unit processes found in a wastewater treatment plant, including advanced nutrient removal models, fixed-film operations, anaerobic reactors, secondary settler and so on. More information about the current development in the software can be found on the website: `http://www.hydromantis.com/`.

With regard to the bioreactor in the activated sludge process, the whole ASM family is included in the GPS-X library. Furthermore, some advanced

models are also available. For example, the `mantis` model reproduces the ASM1 model with the following modifications:

- two additional growth process are introduced (one for the autotrophic and one for the heterotrophic organisms) in order to represent the growth of organisms during conditions of low ammonia and high nitrate;
- the kinetic parameters are temperature dependent;
- aerobic denitrification is introduced, according to the Munch *et al.* modifications[4] [83].

Moreover, the `twostepmantis` model allows to describe nitrification phenomena by a two-step process, as represented in the reaction scheme in 2.1. As a consequence, some modifications in the ASM1 model follow:

- The anoxic reactions are duplicated in order to consider also nitrite as electron acceptor;
- Two more state variables are included: `xbai` as nitrite-producers and `xbaa`, nitrate-producers;
- Hydrolysis of rapidly biodegradable substrate is also introduced.

As for the `mantis` model, the kinetic parameters are considered temperature dependent.

With regard to the secondary settler model in the GPS-X library, several models can be found both for the zero and the monodimensional, and for the reactive and the nonreactive cases.

The GPS-X owns two important modules that have been very useful during this work: the `analyzer` and the `optimizer` module. The former has been used to conduct sensitivity analyses on our process layout. The objective of a sensitivity analysis (in the context of simulation) is to determine the sensitivity of the simulation model's output variables to change in the input parameters. The results of this analysis have been very helpful to setting up the model parameters. Moreover, the optimization module has been used to fit the model to experimental data and calibrate the model kinetic and stoichiometric parameters. This allows us to achieve the best possible fit between the model responses and measured data.

[4]The authors demonstrated that simultaneous nitrification (upon the sludge floc surface) and denitrification (in the sludge floc bulk) can take place in the same reactor vessel under identical overall operating conditions.

2.3.2 Matlab™/Simulink™

Matlab is a general high-level language for technical computing. It includes a large library of predefined mathematical functions. Furthermore, it features a family of specific toolboxes that extent the Matlab environment to solve particular classes of problem (there are approximately 40 toolboxes available).

Simulink is an add-on software product to Matlab for modelling, simulating and analyzing any type of dynamic system. Matlab and Simulink are fully integrated, meaning that all functionalities of the Matlab toolboxes are available in the Simulink environment as well. Simulink provides a graphical user interface for building models as block diagrams and manipulating these blocks dynamically. A large number of predefined building blocks are included and it is easy to extend the functionality by customizing blocks or creating new ones. The capabilities of Simulink may be further extended by using the S-functions (system functions), which can be written in Matlab language, C++ or Fortran using predefined syntax. Consequently, S-functions can be easily incorporated and a dynamical system can be described as a mathematical set of equations instead of using predefined block diagrams.

2.4 Case Study

In this entire work, we have considered an activated sludge process operated at the TecnoCasic wastewater treatment plant located near Cagliari (Italy).

In the simulated case the wastewater treatment includes a mechanical treatment to remove floating and settleable solids, then a biological treatment with activated sludge for removal of nitrogen and organic pollutants, and after that other operations such as sludge treatment and water chemical treatment. The liquid waste collected derives from municipalities (30%) and industries (70%).

The removal of nitrogen and organic matter is obtained with an ASP, where a pre-nitrification (supplying a low air flow needed just for mixing purposes) is obtained in the first half basin followed by a nitrification (with higher air flow) in the last half. The aeration is obtained with fine pore air diffusers, located at the bioreactor bottom. Two oxygen sensors located in the anoxic and aerobic zone give the controller values to maintain the desired oxygen set point by manipulating of the aeration supply. The TecnoCasic activated sludge configuration is the same showed in Figure 2.1. In Figure 2.6a the

aeration basis in showed, whereas the secondary settler is depicted in Figure 2.6b. The global process is considered isothermal (around 20 °C).

(a)

(b)

Figure 2.6: TecnoCasic ASP

The process layout has the following characteristic features:

- total biological volume (2000 m^3), with an anoxic zone followed by an aerobic zone. The aeration is obtained with fine pores diffuser on the bottom of the basin;
- non-reactive secondary settler with a surface of 707 m^2 and a depth of 4 m;
- RAS recycle from the underflow of the secondary settler to the front end of the plant at the constant flow rate of 7000 m^3/d;
- WAS is pumped intermittently from the secondary settler underflow;
- DO saturation of 8.88 $gO2/m^3$.

The simulation procedure involves simulations to steady state followed by dynamic simulations using the data available from the TecnoCasic plant. Data provided by the plant were the following:

- Influent flow rates;
- Dissolved oxygen concentration in the basin;
- Daily COD and nitrogen (nitrite, nitrate, TKN, ammonia) concentrations in inflow and outflow streams available every two or three days;
- Daily SVI (*Sludge Volume Index*) data from the settler and suspended solids (in inflow and outflow).

COD and nitrogen measurements were obtained off-line in the TecnoCasic laboratory.

The starting point for any model development project is the description of unit process objects that make up the plant as well as the flow path between unit processes. This information is usually contained in the plant flowsheet. The aim is to create a plant flowsheet in our software environment using the above information about the process. Firstly, this is built choosing the process units to be represented. In our case we have: a bioreactor and a settler, but we also need an influent unit. The later one becomes necessary in order to collect and transformate the data from the plant into state variables for the bioreactor model.

2.4.1 ASP using the GPS-X

The model building with GPS-X implies the selection of the treatment units from the system library. For each process units many different attributes and characteristics that uniquely describe the object must be specified. For this reason, physical parameters like the real dimension of the unit and kinetic and stoichimetric parameters for the biological reactor have been provided to the simulator. It should be also noted that the aeration basin model can be represented with different configurations, and we have chosen to represent it as a plug-flow tank, that consist (in the simulation environment) of 6 continuous stirred tank reactors in series. Other attributes like the model type associated to each unit and the dissolved oxygen controller in the aeration basin have been specified.

As a first attempt, the simulation procedure have been implemented using the GPS-X software with the `twostepmantis` model for the biological reactor. However, our final aim in using the commercial software was to have a good reference to built ASM1 model in Matlab/Simulink. For this reason also a simulation using the ASM1 model have been performed with the GPS-X and the calibration procedure gave the model parameter values. The

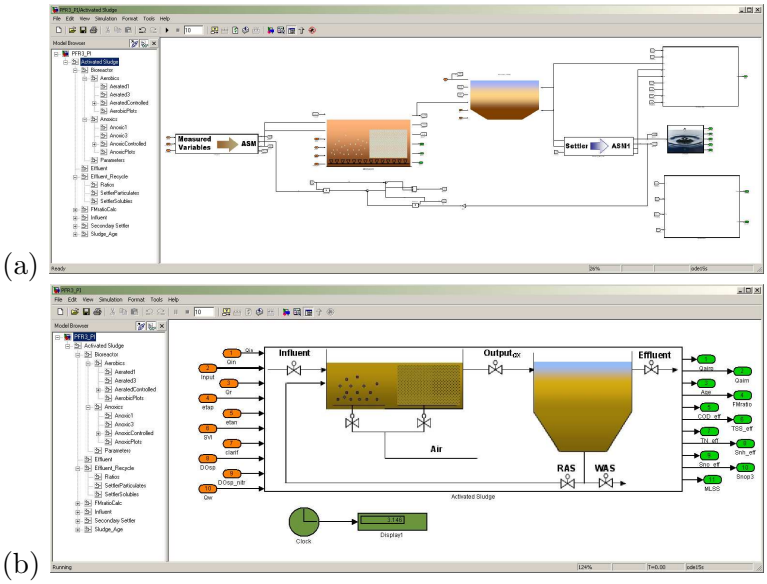


Figure 2.7: Activated sludge process in Simulink environment

analizer module has been used to make sensitivity analysis on the resulting model in order to define the parameters (kinetic and stoichiometric) which have a greater influence on the process. In conclusion, the optimization procedure has been performed to obtain a calibrated model.

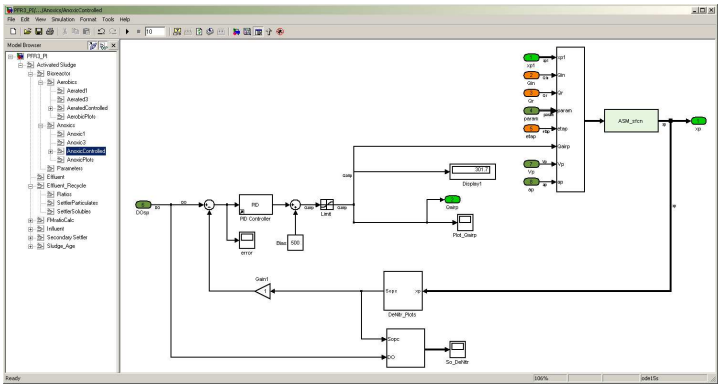
2.4.2 ASP using the Simulink

The required models describing the process in the biological reactor, in the settler and also in the influent units have been implemented in the following way. Firstly, the graphical user interface of Simulink has been used to build the blocks and also the connection between them. Secondly, since there is not model representing the activated sludge process, the Matlab language has been used to build the models using a mathematical notation. They were then incorporated into the Simulink environment by the *S-functions*. There resulting final model is shown in Figure 2.7a and in Figure 2.7b. In the former the activated sludge layout with influent, biological reactor and settler models is shown, whereas the final masked system is reported in the latter.

Some considerations were given before proceeding. Using the GPS-X environment it was noted that data provided have to be related to the ASM1 state variables before implementing the model in Simulink environment. This has been done considering the fractions reported in Appendix B, to

design the influent model. Furthermore, the biological reactor has been represented with different zones, in order to represent in some way the components diffusion through the system. If has been tested with two zones (one anoxic parts and one aerated part), with six parts (3+3) and eventually with ten (5+5) parts. It has been found that a good comprise between CPU time consumtion and real improvement in the system representation is obtained giving the aeration basin 3 zones for denitrification reactions and 3 zones for nitrification. Furthermore, since in the real plant the bioreactor is only virtually divided between anoxic and aerobic part, we have assumed that the last represents 2/3 of the total volume.

(a)



(b)

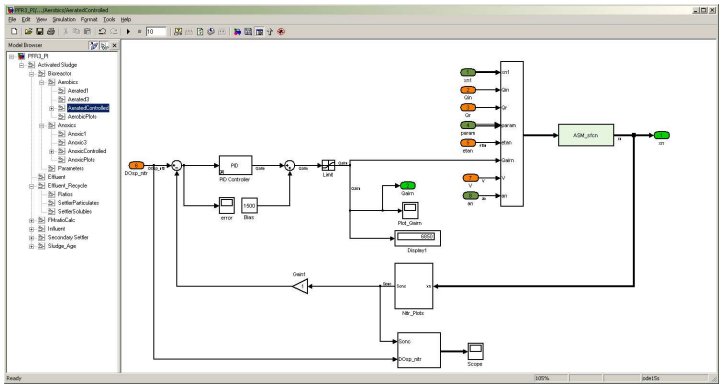


Figure 2.8: Dissolved oxygen controllers in Simulink environment

The dissolved oxygen controller have also been represented in Simulink environment and the resulting screenshot is reported in Figure 2.8a for the anoxic part, and in Figure 2.8b for the aerobic one. It has been assumed that only the central part is controlled in each bioreactor section. The PI controller receives the error between the dissolved oxygen setpoint (given as input data) and returns the corresponding airflow. The air flowrate value

is then used, by each zone, to compute the oxygen mass transfer coefficient according to the following relationship [45]:

$$K_{La} = \frac{1333.3\,\alpha\,\eta\,f_{O_2}}{V S_{Ost}} Q_{air}$$

where, 1333.3 is the unit conversion factor $[gO_2/m^3]$, α is the (wastewater K_{La} to clean water K_{La}) ratio, η is the standard oxygen transfer coefficient, f_{O_2} is the fraction of oxygen in air (i.e., 0.21), V is the aeration volume $[m^3]$, S_{Ost} is the saturated oxygen concentration $[gO_2/m^3]$ and Q_{air} is the air flowrate.

In Appendix B the used stoichiometric and kinetic parameter values for the ASM1 are reported. Appenix B also reports the settler model parameter values for the steady state simulation, whereas in the dynamic simulation they are calculated as function of the SVI data collected from the real plant.

2.4.3 Simulation Results

In Figure 2.9, the comparison between simulation results and experimental data is shown. As we can note the agreement between the simulators is rather good, whereas the agreement with the experimental data is not always so high.

This is not surprising, in fact the data furnished from the plant are sampled every two or three days with respect to the nitrogen compounds and these demonstates not to be enough to calibrate the model especially the ammonia concentration.

In Figure 2.9a, the effluent COD concentration is shown, as we can note the agreement is good, even if some outliers are present. The off-line COD experimental data are sampled once per day, this means that we assumed that the influent COD concentration stays constant during the whole day.

The total suspended solids behavior is shown in Figure 2.9b, as we can note the agreement is also in this case acceptable.

The nitrate/nitrite concentration (Figure 2.9c) shows a good tendency motions, even if it presents offset with the (few) experimental data. Figure 2.9d shows the ammonia behavior, we notice that a good agreement exists between the two simulations. However, both of them suffer from the low data accurancy.

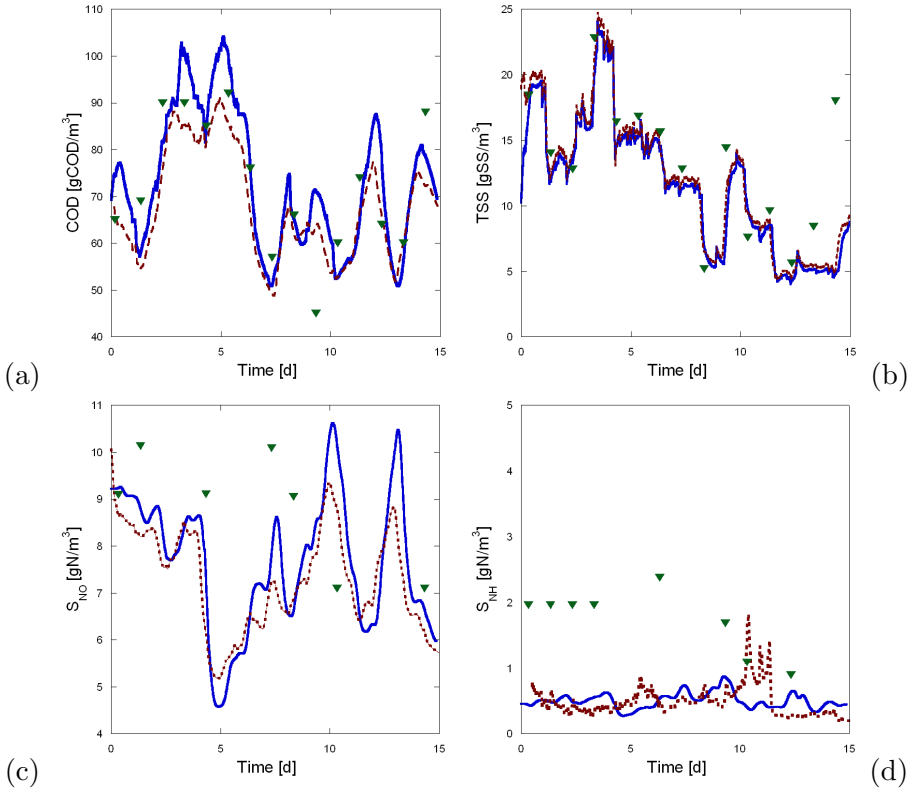


Figure 2.9: Main concentrations in the effluent flowrate [Experimental data (triangle); GPS-X (dashed); Simulink (solid)]

.

In order to improve the simulation results a experimental campaign to collect on-line data should be planed. However this is not the aim of this work and for this reason the obtained behaviors are considered representative of the real plant.

Part I

Controlled Variables Selection

In the following Chapters, the best operating conditions for the activated sludge process will be defined. The aim is to demonstrate how, with simple considerations on the control structure design, the overall efficiency of a wastewater treatment plant can be improved. Of course, one way to improve efficiency could be to construct new and larger basins, but this is expensive and often impossible since the land required may not be available. A sustainable approach is the introduction of more advanced control and operating systems. This is expected to reduce the need for larger volumes, improve the effluent water quality, decrease the use of chemicals and save energy and operational costs.

A suitable solution to the wastewater treatment plant is the development of adequate information systems to control and supervise the process. However, a closer look at the current operation of wastewater treatment plant reveals that automation is still minimal even if in the scientific community and in process industries the importance of automation and control in these processes has been recognized by almost 30 years [87].

Several reasons for this lack in wastewater treatment plant can be found: i) the insight in the process is still marginal, ii) reliable technologies are still unsatisfactory or not existing, iii) the possibilities to act on the process are still inapt or insufficient and, most importantly, iv) wastewater treatment plant is considered as a non-profit industry. Automation has been considered costly and has not been part of the process design. For this reason, the following Chapters deal with the definition of an adequate control structure design focusing on minimizing operational costs in the plant, while keeping it running optimally and satisfying the effluent requirement. This is achieved according to the control structure design proposed by Skogestad [98].

This part of the thesis is organized as follows. In Chapter 3, the basic theory of the Skogestad's procedure is described and some considerations are given for the decentralized control structure approach. Chapter 4 illustrate the proposed applications of such a technique to the WWT process.

Chapter 3

Issues on Control Structure Design

Generally speaking, the objective of a control system is to make the process output behave in a desired way by manipulating the plant's inputs. Considering a full scale plant where different control possibilities exist, the first question that must arise is: *which variables should be measured, which inputs should be manipulated and which link should be made between them?* (Foss, 1973, in [98]). An answer to this question is to select those variables that "*optimize the process*". There may be many issues involved, and to trade them off against each other in a systematic manner we will follow the control structure design proposed by Skogestad [98]. The decision problem is turned into an optimization problem, according to the original ideas proposed by Morari *et al.* [78].

This concept is to mathematically define the quality of operation in terms of a scalar cost function, J, to be minimized. The resuting dynamic optimization problem is to be solved on-line. Of course, this is unrealistic but we might try to achieve a satisfactory operation even with an acceptable loss. In order to obtain this, we could select controlled variables in such a way that acceptable operation with constant setpoints is obatined. Thus, effectively turning the complex optimization problems into a single feedback problem enforcing *self-optimizing control* (Section 3.1).
This procedure is divided in two main parts:

1. **Top-Down Analysis**: including the definition of the operational objectives, the identification of manipulates variables and degrees of freedom and controlled variables and considerations on degrees of freedom available to meet them;

2. **Bottom-Up Design**: of the controlled system, starting with the stabilizing control layers.

Hereafter, we will refer to the first part as the *Controlled Variables Selection* task as explained in Section 3.2 and Section 3.3 outlines the concept of controllability. The second one will be the *Controlled Design* part. This part refers to the practical applicability of an optimum policy by means of a control system: obviously, this involves several tasks depending on the choice made upon the controller. In Section 3.4 the possibility to decentralize the controller is considered. Of course, for an exhaustive compend on self-optimizing control, the Skogestad and Postlethwaite book [101] and also the many articles of Skogestand and coworkers should be consulted. For example, applications for controlled variables selection for reactor, separator and recycle processes are reported in [65]; as well as for optimization and selection of controlled variables for heat-integrated distillation columns [27], and for the Tennessee-Eastman process [66].

3.1 Self-Optimizing Control

A typical control system is organized in a hierarchical structure (Figure 3.1) with several layers each operating on a different time scale. In fact, it is known that weeks are usually needed for scheduling, days are needed for site-wide optimization, hours for local optimization, minutes and seconds for the control layer (often further divided into two levels, with primary controlled variables, supervisory control, and secondary controlled variables).
The layers are interconnected through the controlled variables: the setpoints of the controlled variables are the (internal) variables that link two layers in the control hierarchy. The upper layer computes the setpoint values to be implemented by the lower layer. Thus, the selection of the controlled outputs, for the control layer is usually related to the hierarchical structure of the control system. The importance of selecting the right set of output controlled outpus is clear: we want to find those variables that when kept at constant setpoint, indirectly lead to the near optimal operation with acceptable loss. In this Section, a brief explanation on self-optimizing control is given.

The term *Self-Optimizing Control* was proposed by Skogestad in 2000 [98] because of its close relation to *self-regulatory control* which is when acceptable dynamic performances can be achieved with no control (i.e., with constant manipulates variables). Correspondingly, self-optimizing control is when acceptable economic performance can be achieved without continuous optimization (i.e., with constant setpoints) when disturbances occur. In

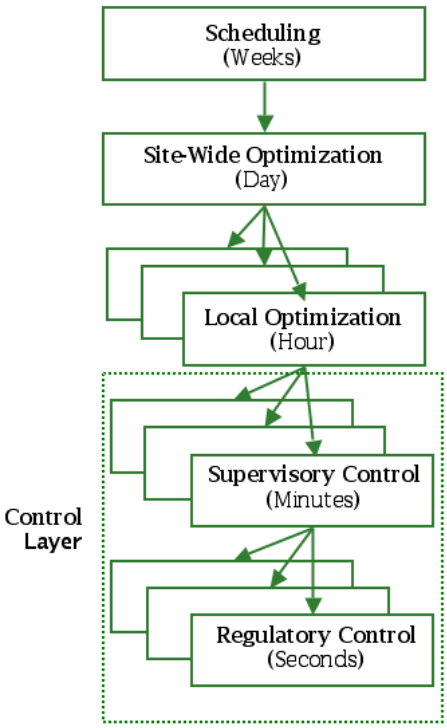


Figure 3.1: Typical control hierarchy in a chemical plant [101]

self-optimizing control, the structure resulting from an optimization problem is utilized to making structural decisions on the controlled variables.

Self-Optimizing Control *is when acceptable operation (small loss) can be achieved using constant setpoints for the controlled variables (without the need to reoptimize when disturbances occur).*

In [99] a simple and systematic procedure for control structure design is proposed. The procedure starts from the definition of a cost function J to be minimized and a loss expressed as the difference between the value of the objective function using constant setpoints and the true optimal value of the objective function. The procedure is twofold: firstly, we must define the controlled variables (through top-down considerations) and then the control design has to be defined. In the following of this Chapter is followed step by step.

3.2 Controlled Variables Selection

What should we control? As mentioned, the goal is to find the controlled variables with good self-optimizing properties. To answer the question we need to evaluate the loss imposed by keeping the selected controlled variables at constant setpoints. In order to determine the loss and the optimal operation for the process, the operational objectives, the process constraints, the degrees of freedom and the manipulated variables must be defined.

3.2.1 Operational objectives and constraints

The operational objectives must be clearly defined before attempting to design the control system; e.g., J may be selected as the operational cost. Other objectives, such constraints, should be formulated distinguishing between transient and steady state constraints. The first ones can be violated (during transient but not at steady-state or in average) and transient constraints must be violated neither in transient nor in steady state.

3.2.2 Degrees of freedom analysis and manipulated variables

The number of *dynamic degrees of freedom* N_m, is equal to the number of manipulated variables. N_m is usually easily obtained by process insight as the number of independent variables that can be manipulated by external means. Typically, the number of adjustable valves plus other adjustable electrical and mechanical variables is considered.
The *optimization degrees of freedom* N_{opt}, represents the degrees of freedom that affect the operational cost J. In most cases, the cost depends on steady-state only, and N_{opt} equals the number of steady-state degrees of freedom N_{ss}. It can be found by counting the manipulated variables and subtracting the number of variables that need to be controlled except those with no steady-state effect (N_{0y}) plus the number of manipulated variables with no steady-state effect (N_{0m}):

$$N_{ss} = N_{opt} = N_m - (N_{0m} + N_{0y})$$

The number of unconstrained steady-state degrees of freedom $N_{opt,free}$, is equal the number of steady-state degrees of freedom minus the number of active constraints at the optimum:

$$N_{opt,free} = N_{opt} - N_{active} \tag{3.1}$$

The manipulated variables are given by the process design and the degrees of freedom analysis should be used to check that there are enough degrees

of freedom to meet the operational objectives both at steady-state and dynamically.

3.2.3 Optimal operation

The definition of the optimal operation for the process is a critical and important task. In fact, the output resulting from the optimization problem also represents the setpoints for the selected controlled variables. In general, optimal operation for given disturbances $\mathbf{d}$ can be found by solving the following problem:

$$\min_{\mathbf{x},\mathbf{u}} \int_{t=0}^{T} J(t, \mathbf{x}, \mathbf{u}, \mathbf{d}) dt \tag{3.2}$$

subject to:

$$\begin{aligned} &\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}_0, \mathbf{u}_0, \mathbf{d}) \\ &\mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{d}) \leq 0 \\ &\mathbf{x}|_{t=0} = \mathbf{x}_0, \ \mathbf{u}|_{t=0} = \mathbf{u}_0, \\ &\mathbf{y} = \mathbf{f}_y(\mathbf{x}, \mathbf{u}, \mathbf{d}) \end{aligned} \tag{3.3}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the vector of internal dependent variables (states), $\mathbf{u} \in \mathbb{R}^{n_u}$ is the vector of manipulates variables, $\mathbf{d} \in \mathbb{R}^{n_d}$ is the vector of external disturbances, $\mathbf{y} \in \mathbb{R}^{n_y}$ is the vector of the measurements and J is the scalar objective function. The equality constraints vector $\mathbf{f}$ corresponds to the model equations, while the inequality constraint vector $\mathbf{g}$ corresponds to the process constraints. We also distinguish between active constraints and inactive constraints. For a given operating point i, an active constraint j satisfies $g_{ij} = 0$, whereas active constraints j satisfie $g_{ij} < 0$.

If we consider slowly varying disturbances, we can make a pseudo-steady state assumption such that the dynamic optimization problem is reduced to:

$$\min_{\mathbf{x},\mathbf{u}} J(\mathbf{x}, \mathbf{u}, \mathbf{d}) \tag{3.4}$$

subject to:

$$\begin{aligned} &f(\mathbf{x}, \mathbf{u}, \mathbf{d}) = 0 \\ &g(\mathbf{x}, \mathbf{u}, \mathbf{d}) \leq 0 \end{aligned} \tag{3.5}$$

such that $J_{opt}(\mathbf{d}) = J(\mathbf{x}_{opt}(\mathbf{d}), \mathbf{u}_{opt}(\mathbf{d}), \mathbf{d})$. For the nominal case when $\mathbf{d} = \mathbf{d}_0$, the corresponding optimal setpoints are expressed as:

$$\mathbf{c}_{opt}(\mathbf{d}_0) = \mathbf{c}(\mathbf{x}_{opt}(\mathbf{d}_0), \mathbf{u}_{opt}(\mathbf{d}_0), \mathbf{d}_0)$$

If we try to implement these setpoints and if there are implementation errors in the optimally active constraints infeasibility arises. The truly optimal constant setpoints are obtained by including all expected uncertainties (all expected disturbances and implementation errors) and evaluating the appropriate cost.

A first optimization is typically performed in order to find the value for the variables $\mathbf{u}$ that when manipulated minimize the objective function J fulfilling the set of constraints ($\mathbf{f}$, $\mathbf{g}$). The optimization problem can be regarded as a *stochastic optimization problem* with disturbances that are effected by random terms $\Delta\mathbf{d}$. In such a situation, the problem in expressions 3.4 and 3.5 becomes:

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{u}} J(\mathbf{x},\mathbf{u},\mathbf{d}) \\ &\mathbf{f}(\mathbf{x},\mathbf{u},\mathbf{d}) = 0 \\ &\mathbf{g}(\mathbf{x},\mathbf{u},\mathbf{d}) \leq 0 \\ &\mathbf{d} = \mathbf{d}_0 + \Delta\mathbf{d} \end{aligned} \tag{3.6}$$

with $\Delta\mathbf{d}$ the random vector varying over the set $\mathcal{D}_d \subset \mathbb{R}^{n_d}$. Conceptually, this is the same approach proposed by Glemmestad *et al.* [32], with the disturbances $\mathbf{d}$ partitioned into two contributors:

$$\mathbf{d} = \mathbf{d}_0 + \mathbf{d}_u$$

with $\mathbf{d}_0$ the nominal disturbance and $\mathbf{d}_u$ the deviation from $\mathbf{d}_0$ and the real disturbance until the new optimization is carried out.
Furthermore, in order to avoid an excessive computational time, some assumptions such as the magnitude of disturbances should be made. This leads to the definition of a pre-specified (discrete) set of unknown disturbances and target satisfaction is demanded for such discrete set of disturbances.

The solution to optimization problem results in the definition of an optimum value for the manipulated inputs $\mathbf{u}$. At this point, we must define the constant setpoint values. This can be done including the uncertainty related to implement the optimal solution. Following Glemmestad *et al.* [32] and also Govatsmark [35], we try to find setpoints $\mathbf{c}_s$ that when implemented, minimize an objective function J fulfilling a set of constraints ($\mathbf{f}$, $\mathbf{g}$, $\mathbf{c}$) affected by random parameters ($\Delta\mathbf{d}$ and $\Delta\mathbf{d}_c$). Also in such a case, only a

discrete set of disturbances is required:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} J(\mathbf{x},\mathbf{u},\mathbf{d}) \\ \mathbf{f}(\mathbf{x},\mathbf{u},\mathbf{d}) = 0 \\ \mathbf{g}(\mathbf{x},\mathbf{u},\mathbf{d}) \leq 0 \\ \mathbf{c}(\mathbf{x},\mathbf{u},d) = \mathbf{c}_s + d_c \\ \mathbf{d} = \mathbf{d}_0 + \Delta\mathbf{d} \\ \mathbf{d}_c = \mathbf{d}_{c,0} + \Delta\mathbf{d}_c \end{aligned}$$

with $\mathbf{c}_s \in \mathbb{R}^{n_u}$ and $\Delta\mathbf{d}$ and $\Delta\mathbf{d}_c$ random vectors varying over the set $\mathcal{D}_d \subset \mathbb{R}^{n_d}$ and $\mathcal{D}_c \subset \mathbb{R}^{n_u}$, respectively.

In order to consider all the different situations and possibilities we can examine a similar deterministic problem. Instead of minimizing the expected cost, we could minimize some mean weighted cost function:

$$J_w = \sum_{i=0}^{m} \omega_i J(\mathbf{x}_i, \mathbf{u}_i, \mathbf{d}_i) \tag{3.7}$$

where, the weights ω_i correspond to a reasonable probability distribution for all expected disturbances $\mathbf{d} \in \mathcal{D}$, with the set $\mathcal{D}$ representing the possible values of the unknown disturbances. The problem might be infinite dimensional, but we can simply consider a discrete number of operating points ($i = 0, \ldots, m$, where 0 denotes the nominal point and m is the number of "disturbed" operating points). The resulting *robust optimization* problem has an optimal value found as solution to the following problem:

$$(\mathbf{x}_{robust}, \mathbf{u}_{robust}, \mathbf{c}_{s,robust}) = arg[\min_{\mathbf{x}_i,\mathbf{u}_i,\mathbf{c}_s} \sum_{i=0}^{m} w_i J(\mathbf{x}_i, \mathbf{u}_i, d_i)] \tag{3.8}$$

subject to:

$$\begin{aligned} \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, d_i) = 0 \\ \mathbf{g}(\mathbf{x}_i, \mathbf{u}_i, d_i) \leq 0 \\ \mathbf{c}(\mathbf{x}_i, \mathbf{u}_i, d_i = \mathbf{c}_s + d_{c,i}) \\ d_i = d_0 + \Delta d_i \\ d_{c,i} = d_{c,0} + \Delta d_{c,i} \end{aligned} \tag{3.9}$$

Since the solution of the optimization problem for each set of candidate variables is needed, the robustly optimal setpoints ($\mathbf{c}_{s,\,robust}$) are found from solving such a problem. Eventually, the **loss** for a given disturbance d and implementation error d_c with constant robust setpoints is expressed as:

$$L(\mathbf{d}, \mathbf{d}_c, \mathbf{c}_{s,\,robust}) = J_c(\mathbf{c}_{s,\,robust} + \mathbf{d}_c, \mathbf{d}) - J_{opt}(\mathbf{d}) \tag{3.10}$$

3.2.4 Optimal Controlled Variables

In literature, a consistent number of methods and applications (*exact local method*, [40], *optimal linear combination of variables*, [39], *gradient function*, [13], *null space method* [1]) to select the optimal controlled variables can be found; in this section only the methods applied to the activated sludge process are reported and discussed.

Qualitative Rules

To approach the problem in a systematic manner, it is useful to consider the reasons why a constant setpoint policy may not be optimal. Generally, there are two reasons: namely, the presence of disturbances ($\mathbf{d}$) and implementation errors (ξ). This has some implications on the choice of the controlled variables $\mathbf{c}$. Following the self-optimizing philosophy, in order to minimize the effect of disturbances, we want the optimal value of $\mathbf{c}$ to remain constant. That is, the sensitivity of $\mathbf{c}_{opt}(J)$ to changes in $\mathbf{d}$ should be as small as possible. In addition, the sensitivity of $\mathbf{c}$ to changes in the independent variable $\mathbf{u}$ should be as large as possible to minimize the effect of implementation errors. For these reasons, it follows that the controlled variable $\mathbf{c}$ should have the following properties:

1. the optimal value of $\mathbf{c}$ should be insensitive to disturbances;
2. the controlled variable should be easy to measure and control;
3. the controlled variable should be sensitive to changes in the (steady-state) degrees of freedom;
4. for cases with more than one unconstrained degree of freedom, the selected controlled varables should be independent.

The first rule tends to minimize the effect of disturbances $\mathbf{d}$. The second rule reduces the magnitude of ξ. The last two minimize the effect of implementation error ξ.

Minimum singular value rule

The minimum singular value rule ([101], [40]) bases the selection of the controlled variables on a scaled steady-state gain from the inputs to the candidate outputs. It follows in a quite straightforward manner the qualitative rules reported in the previous section. In fact, for small disturbances the above condition may be combined into a single rule: the *minimum singular value rule*. The rule can be summarized into the following statement: select as candidates those sets of controlled outputs which correspond to

the large minimum singular value of the appropriately scaled steady state gain matrix $\mathbf{G}$ from $\mathbf{u}$ to $\mathbf{c}$.
The use of the minimum singular value rule is summarized in the Skogestad and Postlethwaite book [101]. For the sake of completeness it is reported in the following:

1. From a (nonlinear) model compute the optimal parameters (inputs and outputs) for various conditions (disturbances, operating points).

2. Obtain for each candidate output the variation in its optimal value, $v_i = (y_{opt,max} - y_{opt,min})/2$.

3. Scale the candidate outputs such that for each output of the sum of the magnitudes of v_i and the control error e_i is comparable.

4. Scale the inputs such that a unit deviation in each input its optimal value has the same effect on the cost function J.

5. Select as candidate those sets of controlled outputs corresponding to the large minimum singular value.

Brute Force Approach

This is not an optimal method to select the controlled variables, but it provides a useful heuristic. It requires to evaluate the loss for alternative sets of controlled variables. This is done by solving the nonlinear equations, and evaluating the cost function J for various disturbances $\mathbf{d}$ and control error $\mathbf{e}$, assuming $\mathbf{y} = \mathbf{c}_s + \mathbf{e}$ where the reference value $\mathbf{c}_s$ is kept constant [101]. Here, $\mathbf{c}_s$ is usually selected as the optimal value for the nominal disturbance, but this may not be the best choice and its value may also be found by optimization (see Section 3.2.3). The set of controlled outputs with smallest worst-case or average value is then preferred.

3.3 Controllability

The selected controlled variables can be analyzed to see if they are adequate with respect to other criteria that may be relevant. The feasibility region and the input-output controllability are such criteria. In fact, before starting any controller design one should first determine how *"easy"* is the plant to control.

Input-output controllability is the ability to achieve acceptable control performance. That is, to keep the output within specific bounds or

displacements from their references, in spite of unknown but bounded variations (such as disturbances and plant changes) using available inputs and variable measurements.

A plant is controllable if the existing controller yields to acceptable performances for all expected plant variations. Input-output controllability analysis is applied to a plant to find out what control performance can be expected. The methods available for controllability analysis are largely qualitative. In most cases the simulation approach is used: i.e., performances are assessed through exhaustive simulations.

3.4 Controller Design

Having established the optimal controlled variables, the next step is to find a suitable control structure: that is, to find the actual implementation of the optimum policy in the plant using such control scheme. If we need more than one controlled variable, the problem becomes a multivariable problem with several possible inputs and several possible outputs. One of the main challenges when defining a controller for a multivariable plant is the choice of a suitable control structure. A typical starting point is to use multiple independent single-loop controllers with each controller using one input variable to control a preassigned output variable. The selection requires to quantify the significance of the interactions in the plant model. Primarily because of the interactions among the process variables, multivariable systems cannot, in general, be treated like multiple independent single-loop systems.

In this section the basic theory for a decentralized control structure is reported.

3.4.1 Decentralized Control

Variables interaction is a common feature that generates difficulties to control a process variable without perturbing other variables of interest (one input signal affecting several outputs signals). An important question arises from the so called pairing problem of selecting which input signal to control and which output signal to get the most efficient control with a low degree of interaction. If interactions in the open loop are severe, a multivariable control structure may be preferable. Nevertheless, if a sparse control structure can be used instead of a full multivariable one, much could be gained in terms of reducing the controller complexity.

A key issue is the way inputs and outputs should be paired. The most significant result is the seminal work of Bristol [12], who developed the idea of the *Relative Gain Array*, RGA. The RGA is a measure used in order to decide a suitable input-output pairing when applying a decentralized control structure. After Bristol's work was published, several researchers have studied the properties and usage of the RGA: for instance, Skogestad and Morari [100], have shown that plants with large RGA elements are very sensitive to modelling errors. Other authors have proposed new measures of interaction and criteria to choose a sensible input-output pairing. Among them, the *Nierderlinski Index* [84], the *Relative Interaction Array* [120], the *Relative Dynamic Array* [116] are examples of later refinements (see also [75] and [94]).

The RGA often provides a limited knowledge about when to use multivariable controllers and gives no indication of how to choose the multivariable structure. A different approach for investigating interactions using so called *participation matrices* was introduced by Conley and Salgado [19]. In this approach, the controllability and observability gramians of a system are used in order to quantify the degree of interaction. This work was followed by the paper of Wittenmark and Salgado [117] where the Hanken-norm of the system is used to develop the so called *Hankel Interaction Index Array*.

3.4.1.1 Relative Gain Array

The RGA for a quadratic plant is given by:

$$\text{RGA} \equiv \Lambda(\mathbf{G}(s)) = \mathbf{G}(s) \times (\mathbf{G}(s)^{-1})^T \tag{3.11}$$

where $\mathbf{G}(s)$ is the transfer functions matrix and the operator $\times$ denotes the Schur product (i.e., elementwise multiplication). Each element in the RGA is defined as the open-loop gain divided by the gain between the same variables when the other loops are controlled.

$$\begin{aligned} \lambda_{ij} &= \frac{(\partial y_i/\partial u_i)_{u_{k\neq j}}}{(\partial y_i/\partial u_j)_{y_{k\neq i}}} = \\ &= \frac{\text{gain with all other loops open}}{\text{gain with all other loops closed}} \end{aligned} \tag{3.12}$$

For a 2×2 plant, a symmetric RGA matrix is obtained:

$$\text{RGA} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 1-\lambda_{11} \\ 1-\lambda_{11} & \lambda_{11} \end{pmatrix}$$

Although, definition 3.12 is limited to steady-state ($s = 0$), expression 3.11 may be used to compute RGA as a function of frequency ($s = j\omega$) to

obtain the dynamic RGA. In this work only the static steady-state RGA is used. This is motivated by the slow time constants in the activated sludge processes [91].
Depending on the value of λ, a number of different cases occur [85]:

- $\lambda_{ij} = 1$, indicates that the open-loop gain between y_i and u_j is identical to the closed-loop gain. This is the ideal case when no interaction between the loops is present. The pairing should be along the diagonal (u_i with y_i and u_j with y_j).

- $\lambda_{ij} = 0$, indicates that the open-loop gain between y_i and u_j is zero. This is the same situation as above, except that now the suggested pairing is along the anti-diagonal (pairing u_i with y_j and u_j with y_i should be a good solution).

- $0 < \lambda_{ij} < 1$, indicates that the gain between y_i and u_i is smaller than the closed-loop gain. This situation indicates that the gain increases when the loops are closed: i.e., there is interaction and if possible avoid pairing y_i with u_j, whenever $\lambda_{ij} = 0.5$.

- $\lambda_{ij} > 1$, indicates that the open-loop gain between y_i and u_j is larger than the closed loop gain. This situation is also undesiderable. The higher the value of λ_{ij}, the greater the opposition u_j experiences from the other control loops in trying to control y_i. Therefore, if possible, do not pair m_j with y_i if λ_{ij} takes a very large value.

- $\lambda_{ij} < 0$, indicates that the open-loop and closed-loop gains between y_i and u_j have opposite signs. This situation corresponds to the worst case scenario because the sign of the gain changes when the loops are closed (this is highly indesiderable).

To summarize we report the following rule: *pair input and output variables with positive RGA elements that are the closest to 1.0.*

As previously said, the RGA may reveal information regarding possible control difficulties and, thus, about the plant robustness. This kind of information is linked to the condition number of the system.

Letting $\mathbf{G}(s)$ the linear transfer function matrix of the plant, the condition number of the plant is the ratio between the maximum and minimum singular values of $\mathbf{G}$ at any given frequency:

$$\gamma(\mathbf{G}) = \frac{\sigma_{max}(\mathbf{G})}{\sigma_{min}(\mathbf{G})} \tag{3.13}$$

For MIMO linear systems, the gain at any frequencies is bounded by the smallest and the largest singular values for the process transfer function matrix **G**. Thus, if the condition number is large, this span will be large and the process model will show high directionality: i.e., the actual gain is highly dependent on the direction of the input vectors. The condition number is scaling dependent and, thus, first a proper scaling of the model is to be selected.

Niederlinski Index

Even through pairing rules are usually sufficient in most cases, it is often recommended (especially with 3×3 and higher dimensional systems) to use this rule in conjunction with the stability considerations provided by the following theorem originally due to Niederlinski [84]: *under closed-loop conditions in all n loops, the multiloop system will be unstable for all possible values of controller parameter (i.e. it will be structurally monotonic unstable), if the Niederlinski index N is negative.* Formally:

$$N \triangleq \frac{|\mathbf{G}(0)|}{\prod_{i=1}^{n} g_{ij}(0)} < 0 \tag{3.14}$$

This result is both necessary and sufficient only for 2×2 systems. For higher dimensional systems, it provides only sufficient conditions: i.e., if equation 3.14 holds then the system is definitely unstable. However, if equation 3.14 does not hold, the system may or may not be unstable: the stability will, in this case, depend on the value taken by the controller parameters. Another rule follows: *any loop pairing is unacceptable if it leads to a control system configuration whose Niederlinski index is negative.*

3.4.1.2 Hankel Interaction Index Array

A measure based on gramians is able to handle the disadvantages of the RGA matrix [19]. This measure is based upon the sum of the squared Hankel singular values for the elementary subsystems of the process. In 2002, a modified version of the interaction measure was suggested by Wittenmark and Salgado [117]. It only uses the Hankel norm of the subsystems and is called the *Hankel Interaction Index Array*, HIIA.

In order to illustrate the gramians-based interaction measure, we start this section the bare definition of gramians.
Consider a linear system expressed in a state space form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

where $\mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^n \times \mathbb{R}^m$, $\mathbf{C} \in \mathbb{R}^q \times \mathbb{R}^n$ and $\mathbf{D} \in \mathbb{R}^q \times \mathbb{R}^n$. The *controllability gramian* $\mathbf{W}_c$, and the *observability gramian* $\mathbf{W}_o$ associated to the system satisfy the Lyapunov equations:

$$\begin{aligned} \mathbf{W}_c - \mathbf{A}\mathbf{W}_c\mathbf{A}^T - \mathbf{B}\mathbf{B}^T = 0 \\ \mathbf{W}_o - \mathbf{A}^T\mathbf{W}_o\mathbf{A} - \mathbf{C}^T\mathbf{C} = 0 \end{aligned} \tag{3.15}$$

The controllability and observability gramians of a system quantify the difficulty to control and observe the system state. For instance, the ranks of the matrices are the dimensions of the controllable and the observable subspace, respectively.
As shown in [19] and in [117], it is possible to split the system given by ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$) into fundamentals subsystems ($\mathbf{A}$, $\mathbf{B}_j$, $\mathbf{C}_i$, D_{ij}), where $\mathbf{B_j}$ denotes the j-th column in $\mathbf{B}$, $\mathbf{C}_i$ is the i-th row in $\mathbf{C}$ and D_{ij} is the (i,j)-th element of $\mathbf{D}$. For each subsystem the controllability and observability gramians can be evaluated and the controllability and observability gramians for the whole system are represented by the sum for the all subsystems. Both gramians depend on the choice of the state-space realization.

Denoting $\lambda_1, \lambda_2, \ldots, \lambda_n$ $(i = 1, 2, \ldots, n)$ the eigenvalues of $\mathbf{W}_c\mathbf{W}_o$ the system Hankel singular values σ_i^H are expressed as:

$$\sigma_i^H(\mathbf{G}) = \sqrt{\lambda_i} \tag{3.16}$$

where the conventional inequality $\sigma_i^H \geq \sigma_{i+1}^H \geq 0$ holds. In [117], it is also shown that the Hankel norm of $\mathbf{G}$ given in 3.16 can also be interpreted as a gain between the past inputs and the future outputs. The Hankel norm of a system with transfer function matrix, $\mathbf{G}$, is defined as:

$$\|\mathbf{G}\|_H = \sqrt{\lambda_{max}(\mathbf{W}_c\mathbf{W}_o)} = \sigma_1^H \tag{3.17}$$

showing that the Hankel norm is thus the maximum Hankel singular value. The measure is invariant with respect to the state-space realization is, therefore, a well suited combination of the controllability and observability measures. If the Hankel norm is calculated for each fundamental subsystem and arranged in a matrix $\bar{\Sigma}_H$ given by:

$$[\bar{\Sigma}_H]_{ij} = \|\mathbf{G}_{ij}\|_H \tag{3.18}$$

the matrix can be used as an interaction measure. In [117] a normalized version of equation 3.18 is proposed:

$$[\Sigma_H]_{ij} = \frac{\|\mathbf{G}_{ij}\|_H}{\sum_{kl}\|\mathbf{G}_{kl}\|_H} \tag{3.19}$$

In the HIIA the larger is the ij-th element the larger is the impact on the corresponding input-output pairing. If the intention is to use a decentralized controller, the HIIA can be used and interpreted in the same way as the RGA.

Chapter 4

Controlled Variables Selection for an ASP

From a control engineering point of view, the activated sludge process is a complex topic for several reasons. First of all, the process is time varying in terms of steady state behavior: i.e., it is a biological process where temperature, composition of the influent wastewater and amount of biomass on flow vary with time. Furthermore, the WWT has different dynamics with constants ranging from seconds to days [5]. For example, the response to changes in air flowrate or chemical dosage is nearly instantaneous (seconds), while dissolved oxygen or pumping control affects the process in minutes. The distribution of sludge is effective in hours. The sludge age control to support adequate organisms has to be considered in terms of days.

In such a context, the main objectives of a global wastewater treatment plant control are:

- maintaining liquid and sludge inventory controller;
- maintaining required effluent quality;
- disturbances rejection;
- efficient operation and reduction costs.

The task for the plant operator and global plant strategy is to reduce operational costs maintain sludge and liquid inventory and reject disturbances. That is, in order to run a plant economically operation costs such as pumping energy and aeration energy should be minimized. At the same time, the discharge concentrations to recipients should be kept at low level. Minimizing the operational costs and at the same time treating the wastewater

properly may lead to conflicting goals. In other words, a proper operation in a wastewater treatment plant can be translated in a constrained optimization problem.

In literature several works about the optimization problem of an activated sludge processes can be found. For instance, Chachuat in his PhD thesis [16] and in [17] have investigated the optimal sequence of aeration and non-aeration times in a sequencing activated sludge bioreactor. Samuelsson *et al.* [92] showed how with an appropriate nitrate cost function, the optimum cost can be reached with respect to the effluent legislation requirements. In Gillot *et al.* [31] an objective cost function is defined in order to standardize a cost calculation procedure integrating both investment, fixed and variable operating costs. In addition, Vanrolleghem *et al.* [113] proposed a economic index including weighted investment and operating costs that in junction to a robustness index can be used to evaluate the transferability of control strategies to different situations. The economic index has also been used in [118] where the authors applied it to a Japanese study case using genetic algorithm optimization.

In this Chapter it is shown how an answer to the optimization problem can be given only by designing the control structure in such a way that the operational costs are minimized. In Chapter 3 we have showed how the controlled variables can be important to link the layers in the hierarchic control structure. For that reason, the aim here is to find the optimal controlled variables for the activated sludge process in a wastewater treatment plant. We follow the procedure illustrated in the previous Chapter:

- **Step 1**: Identify operational constraints, and preferably identify a scalar cost function J to be minimized;
- **Step 2**: Identify degrees of freedom and manipulated variables;
- **Step 3**: Optimization;
- **Step 4**: Identify primary controlled variables;
- **Step 5**: Design the controller.

4.1 Operational objective

In order to define the operational objective for the activated sludge process, we must firstly define the cost function to be minimized and the constraints to which it will be subjected to.

4.1.1 Cost Function

The overall cost in a wastewater treatment plant is highly dependent on the wastewater treatment system itself and it can be divided into manpower, energy, maintenance, chemicals, sludge treatment and disposal evaluated on a time basis. Therefore, an inventory has to be made of the different costs so that the individual importance of each different term is determined. The importance varies as a function of the wastewater treatment plant, the origin of the wastewater, the legislatory and other place-and-time-depending circumstances [15]. The aim is to analyze the steady state (stationary) operational cost of the ASP shown in Chapter 2. To express the cost for such a process, the following partial costs are considered:

- Pumping costs due to the required pumping energy;
- Aeration costs due to the required aeration energy;
- Sludge disposal costs.

For the sake of completeness, chemical dosage costs should be taken into account, even if this is not a charge truly relevant in the considered wastewater treatment plant. For this reason, the chemical dosage cost was not considered in defining the following overall costs.

To express the partial costs, we adopt the expressions suggested in the *COST Benchmark* [20]. The total average energy over a certain range of time T, depends directly on the recycled sludge flowrate (Q_r) and on the waste sludge flowrate (Q_w):

$$E_P = \frac{0.04}{T} \int_{t_0}^{t_0+T} (Q_r(t) + Q_w(t))dt \tag{4.1}$$

with E_P expressed in kWh/day.
The aeration energy (expressed in kWh/day) can be, in turn, expressed as:

$$E_A = \frac{24}{T} \int_{t_0}^{t_0+T} \sum_{i=1}^{n} (0.4032(K_{la,i})^2 + 7.8408K_{la,i})dt \tag{4.2}$$

where $K_{la,i}(t)$ is the oxygen transfer function in the aeration tank i ($i = 1, \ldots, n$ with n equal to the number of the reactors, see Chapter 2). Considering that we give oxygen to both anoxic and aerobic zones, $E_{A,tot} = E_A^p + E_A^n$.

Assuming a price k_D equals to 80 €/$tonn$ on the sludge disposal and TSS_w $[gSS/m^3]$ the total suspended solids concentration wasted with Q_w, the disposal cost per day (€/day) is expressed:

$$C_D = \frac{1}{T}\int_{t_0}^{t_0+T}(k_E TSS_w Q_w)dt \tag{4.3}$$

Assuming a constant energy price ($k_E = 0.09$ €/kWh), the total energy cost (in €/day) can be calculated during a representing time interval T from 4.1 to 4.3 as:

$$C_{tot} = k_E(E_P + E_{A,tot}) + C_D \tag{4.4}$$

The cost function $J = C_{tot}$ to be minimized depends on the influent characteristics, the reactor volume (especially for the aeration energy part), the hydraulic retention time, and the general process behavior.

4.1.2 Constraints

As for the constraints, those related to effluent discharge and, of course, those related to the operability of the process are considered.

Operational Constraints

As for the operational constraints, we identify those related to the flow in the tank, to the aeration compressor, to the valve in the plant, and so on. But most importantly, we identify the dissolved oxygen concentration in the aeration tank (in the denitrification and nitrification zone), the nitrate concentration in the denitrification zone and the constraints for good operational conditions of the secondary settler. In such a way, the cost function optimization can ensure a proper operation for the activated process.

The *Dissolved Oxygen* (DO) concentration is one of the principal parameters in an ASP. The amount of oxygen transferred to the aeration tanks should be equal to the amount of oxygen required by the microorganisms in the activated sludge process to oxidize the organic material and to maintain residual DO operating levels. When oxygen limits the growth of microorganisms, filamentous microorganisms may predominate, and the settleability and quality of activated sludge may be poor. On the other hand, an excessively high DO, meaning also a high flow rate, leads to high energy consumption and may also deteriorate the sludge quality. In practice, the DO concentration in the aeration tank should be maintained at about 1.5 to $4\,gO_2/m^3$ in the aerobic aeration tanks: $2\,gO_2/m^3$ is a commonly used value. In fact, $4\,gO_2/m^3$ does not improve operation significantly, but increases the

aeration cost considerably [46]. In the anoxic zone a lower aeration is needed in order to satisfy only the mixing requirements in the denitrification zone. Summarizing, the following DO constraints are defined:

- DO concentration in the aerobic zone: $1.5 \leq DO^n \leq 4\, gO_2/m^3$;
- DO concentration in the anoxic zone: $0.05 \leq DO^p \leq 0.5\, gO_2/m^3$.

Furthermore, if we assure that the nitrate consumption in the last predenitrification zone is not exceeding a certain level, we can guarantee a not excessive air consumption in the aeration zones. According to some works in literature ([47], [87], [92]), for most reasonable operating points the nitrate concentration in the anoxic compartment should be maintained in the interval $1-3\, gN/m^3$ when an internal recirculation is present. In the considered treatment plant there is not an internal recirculation, and we verified that for the nitrate/nitrite in the last anoxic zone (S_{NO}^{p3}) a lower constraint of $0.75\, gN/m^3$ can assure a good behavior in the predenitrification zone.

We know that the purpose of the recycled flowrate Q_r is to maintain a sufficient concentration of the activated sludge in the aeration tank so the required degree of treatment can be obtained. It is also important to prevent the loss of sludge solids in the effluent. For this reason, it is important to define some indexes to represent the sludge behavior both in the bioreactor and in the settler. For example, we know that the excess activated sludge produced each day must be discharged to maintain a given *Food to Microorganisms Ratio* (F/M) or a given *Sludge Retention Time* (SRT, also known as *sludge age* or *mean cell residence time*). So, we consider the F/M and the SRT as operational constraints for the secondary settler performance.

The common definition of the SRT is a steady-state definition that assumes all flowrates and concentration constant. The calculation is based on the total biomass present in the system (i.e., the bioreactor and the settler):

$$SRT = \frac{TX_a + TX_s}{\phi_e + \phi_w} \tag{4.5}$$

In equation 4.5, TX_a is the total biomass in the reactor and TX_s is the total biomass in the settler:

$$\begin{aligned} TX_a &= \sum_{i=1}^{n} (X_{BH,i} + X_{BA,i}) V_i \\ TX_s &= \sum_{j=1}^{m} (X_{BH,j} + X_{BA,j}) z_j A \end{aligned}$$

where $i = 1, \ldots, n$ are zones in the bioreactor of volume V_i and $j = 1, \ldots, m$ are layers of thickness z_j in the secondary settler model with a cross sectional area A (see Chapter 2 for more details on the secondary settler model). In expression 4.5, the term ϕ_e is the loss rate in the effluent, and ϕ_w is the loss rate of biomass in the wastage:

$$\begin{aligned} \phi_e &= (X_{BH,1} + X_{BA,1})Q_e \\ \phi_w &= (X_{BH,w} + X_{BA,w})Q_w \end{aligned}$$

On the basis of laboratory tests and actual operating data, it has been found that a SRT of about 3 to 15 days results in the production of a stable, high quality effluent sludge with excellent settling characteristics [46]. Indeed, the lower limit of the SRT is apparently imposed by nitrifiers kinetics and can be determined by multiplying a safety factor $(1.5-2.5)$ to the minimum SRT required by nitrifiers (determined from the maximum specific growth rate of nitrifiers and their decay rate) [87]. This secures a reasonable amount of nitrifiers in the system so that satisfactory nitrification is achieved during periods of high nitrogen loading.

The next considered parameter is the F/M ratio. This is a process control variables that helps to determine the load of the incoming food (the COD mass in the influent flow rate) divided by the amount of microorganisms in the system (the volatile suspended solids in the aeration tank, the *Mixed Liquor Suspended Solids*, $MLSS$):

$$F/M = \frac{Q_{in} COD_{in}}{V\, MLSS} \tag{4.6}$$

where

$$MLSS = \frac{\sum_{i=1}^{n} TSS_i V_i}{V} i_{vt}$$

i_{vt} is a stoichiometric coefficient that represent the VSS/TSS ratio and in this case is equal to 0.6 $[gVSS/gTSS]$. Typical values for the Food to Microorganisms ratio reported in literature vary from 0.05 to 1[1] $gCOD/gSS/d$.

A summary of operational constraints for our activated sludge process is reported in Table 4.1.

Effluent Constraints

Assuming that the equipments and the process are well functioning is important, but the main aim of a wastewater treatment plant is to satisfy the

[1]In literature, [46] reported values of F/M expressed in terms of BOD concentration in the influent instead of COD. Since, in our model the organic matter is always expressed in terms of COD, units have been change considering a COD/BOD_5 ratio of 0.66.

Operational Constraints						
Anoxic DO	0.05	$\leq$	DO^p	$\leq$	0.5	$[gO_2/m^3]$
Aerobic DO	1.5	$\leq$	DO^n	$\leq$	4	$[gO_2/m^3]$
Sludge Age	9.3	$\leq$	SRT	$\leq$	15	$[d]$
F/M ratio	0.05	$\leq$	F/M	$\leq$	1	$[gCOD/gSS/d]$
Anoxic S_{NO}	0.75	$\leq$	$S_{NO}^{p,3}$	$\leq$	1	$[gN/m^3]$

Table 4.1: Operational constraints

effluent requirements. For that reason, we identify as effluent constraints the concentrations of organics, nitrogen, and other pollutants in the output flow. The final constraint of the ASP is defined by the legislation requirements for effluents deriving from a wastewater treatment plant. In order to avoid both infeasibility and failure in the ASP, the effluent constraints have been reduced with respect to the values imposed by law.

Summarizing, the cost function J is also subjected to the constraints reported in Table 4.2.

Effluent Constraints				
COD	COD^{eff}	$\leq$	125	$[gCOD/m^3]$
Total SS	TSS^{eff}	$\leq$	35	$[gSS/m^3]$
Total N	TN^{eff}	$\leq$	18	$[gSS/m^3]$
Ammonia	S_{NH}^{eff}	$\leq$	0.6	$[gN/m^3]$
Nitrate	S_{NO}^{eff}	$\leq$	10	$[gN/m^3]$

Table 4.2: Effluent constraints

4.1.3 Disturbances

Disturbances are a major reason why control is required. Compared to most other process industries, a wastewater treatment plant is subjected to extremely large disturbances [87]. As a result, the plant is rarely in a steady state, being subjected to transient behavior all the time. Consistent performance must be maintained in the presence of these disturbances. It is not uncommon that a treatment suffers from sludge settleability problems due to an outbreak of filamentous bacteria induced by disturbances. Also operations imposed by on-line control systems may themselves induce a bacterial population shift. Further internal disturbances may be generated by inadequate or inappropriate operations including human errors, unsuitable

or malfunctioning actuators and/or sensor failures. Most importantly, discrete events such as rainstorms and peak of loads may occur from time to time.

In order to give a better representation of the true behavior of a wastewater treatment plant, typical variations, in dry weather conditions, are simulated using the weighting function depicted in Figure 4.1. Both inlet flow rate and load [49] are considered starting from the nominal conditions reported in Table 4.3.

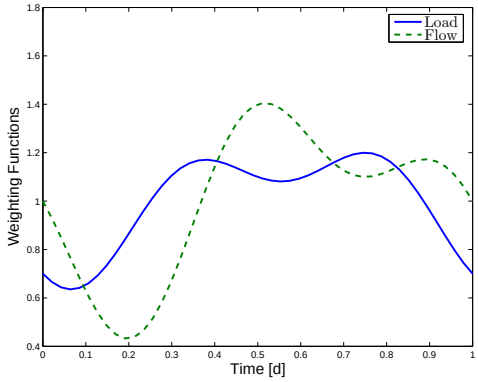


Figure 4.1: Typical weighting functions for dry weather conditions [49]

Influent				
Flowrate	Q_{in}	=	6152	$[m^3/d]$
COD	COD_{in}	=	221	$[gCOD/m^3]$
Total Suspended Solids	TSS_{in}	=	46	$[gSS/m^3]$
Nitrate	$S_{NO,in}$	=	0.22	$[gN/m^3]$
TKN	TKN_{in}	=	22	$[gN/m^3]$
Ammonia/TKN	fnh_{in}	=	0.36	$dimensionless$

Table 4.3: Nominal conditions for the compositions and flowrate

The nominal average conditions are augmneted by 20% in order to have the following disturbances:

- $d_1 = (Q_{in}$ and $COD_{in}) + 20\%$;
- $d_2 = (Q_{in}$ and $TKN_{in}) + 20\%$;
- $d_3 = (Q_{in}$ and COD_{in} and $TKN_{in}) + 20\%$.

As explained in the following section, we consider the influent flow rate constant or not depending on the presence of an equalization tank in the wastewater treatment plant. Of course, when the Q_{in} flow rate is constant, the disturbances considered are the only increment on COD and TKN concentrations.

4.2 Degrees of freedom and manipulated variables

As for the degrees of freedom for optimization, we must identify the number of degrees of freedom for control, N_m. If we look at the schematic representation of that plant in Figure 4.2, we note that there are few variables that we can manipulate (this is quite common in a biological wastewater treatment plant, [86]). However, there is potential to make a better use of the existing manipulated variables.

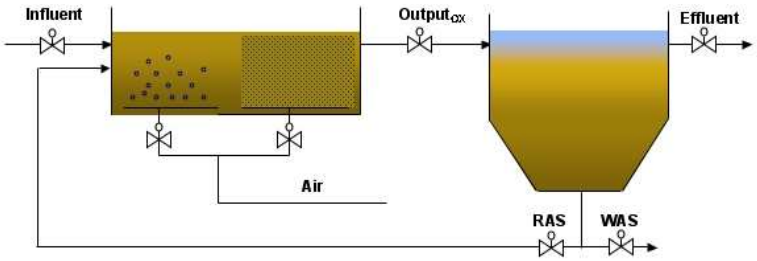


Figure 4.2: Schematic representation of an activated sludge plant, with spotlight on manipulated variables

From Figure 4.2, the number of manipulated variables N_m is equal to 7, including the influent flow rate. Namely, the output flow from the aeration tank should be not taken into account since it is actually self-regulating. The same can be said for the effluent flow from the secondary settler. It follows that $N_m = 5$.

The optimization is generally subjected to several constraints and, N_{opt} degrees of freedom should be used to satisfy the constraints and optimize the operation. In our case, we have $N_{opt} = 4$, because we are not considering the influent flow rate as manipulable variable. If we consider that the DO concentration in both anoxic and aerated zone is always fixed at the setpoint value by the airflow controller on the bioreactor, the number of *active constrains* is $N_{active} = 2$. Recalling equation 3.1, the number of *free* unconstrained degrees of freedom that are left to optimize the cost function J in our process is:

$$N_{opt,free} = 2 \tag{4.7}$$

In the following, the recirculated sludge flowrate (Q_r) and the wastage flowrate (Q_w) are considered as the needed manipulated variables to saturate the $N_{opt,free}$ degrees of freedom.

Some further considerations can be made. In fact, we are considering a wastewater treatment plant where an equalization tank is present at the top of the activated sludge process. This means that the flow rate can be assumed constant. However, the Q_r flow rate is proportional to the influent flow rate (with a proportionality of approximatively 1.3 in the our case plant). For that reason, if we consider Q_r at a constant value leads to:

$$N_{opt,free} = 1 \tag{4.8}$$

Generally speaking, the problem is that very few wastewater treatment plants have the luxury of an influent at a constant flow rate (because an equalization tank is not present): for that reason, we consider to restrict the previous degree of freedom in 4.8 and we consider both cases in expression 4.7 and 4.8. The manipulated variables considered are again the Q_w and Q_r. Summarizing, two different situations are investigated:

1. The influent flow rate is constant and we optimize the system with respect to the waste flow rate: $N_{opt,free} = 1$.

2. The influent flow rate in not constant and we optimize the system with respect to both the waste flowrate and the returned flowrate: $N_{opt,free} = 2$.

4.2.1 Waste Activated Sludge Flowrate

Manipulation of the waste sludge flowrate is used to control the total inventory of sludge mass (a slow process). This, in turn, influences the SRT, giving a significant impact on the formation of certain organisms such as filaments and autotrophs, and it is belived also to influence the sludge bulking [86].

4.2.2 Return Activated Sludge Flowrate

The return sludge is used to keep the sludge within the system. Basically there are two common practices for return sludge flowrate manipulation:

- constant flowrate;
- ratio control.

The second one, that is a fixed ratio between the return sludge flowrate and the influent flowrate, is a common practice. It aims at assuring an adequate sludge within the system at different influent flowrate. As showed in Section 4.2.1, the waste sludge is used to balance the growth of organisms so that the average sludge mass is constant. Actually this means that the recycle ratio has to be kept constant within quite narrow limits in order to keep the overall sludge mass balance. The lower limit of the instantaneous value of the sludge flowrate is determined by the sludge retention time in the settler. The upper limit is determined by the hydraulic load to the clarifier and the maximum dry flow rate in the return sludge. The upper and lower limits are calculated by the expression reported in [86]. For the considered activated sludge process, the ratio Q_r/Q_{in} should be comprises between 0.75 and 1.5 [46].

4.3 Optimization and Controlled Variables Selection

In this section the results and considerations on the optimization procedure along with the adopted methodology are given. First of all, we define the candidate controlled variables (Section 4.3.1), whose setpoints have to be estimated throughout the optimization approach (Section 4.3.2). In Section 4.3.3 and in Section 4.3.4, the results for the one degree of freedom and for two degrees of freedom cases are discussed.

4.3.1 Candidate controlled variables

The candidate controlled variables are defined, since we are only manipulating the recirculated sludge flowrate, Q_r, and the wastage flowrate, Q_w.

The measurement of the *Sludge Blanket Height* (SBH) is a potential candidate controlled variable. The solids form a sludge blanket at the bottom of the secondary settler. The SBH varies in thickness from time to time and might fill the entire depth of the settling tank, if the Q_r pumping capacity is inadequate. For that reason, an optimal blanket level should be maintained in the secondary settler. As suggested by Olsson *et al.* [87], a certain SBH should be maintained in the settler and the optimal level depends on the hydraulic and substrate loading to the plant as well as the sludge settleability and the design of the settler. Unfortunately, it is not straightforward to establish the dependency of the optimal SBH level on these factors. Usually it can be determined by experience and from a balance between efficient settling depth and sludge storage. Generally, $0.3 \leq SBH \leq 0.9\ m$ [46] is

a typically accepted range for the sludge blanket height. One particular difficulty is the lack of a suitable model to quantify the dependency of effluent suspended solids concentration on the SBH and the solids loading rate. In this work the Takács settling model [107] is used (see Chapter 2) to represent the secondary clarifier modelled with ten layers. The SBH is calculated on the basis of concentration of the TSS in the different layers. The difference in all adjacent layers is calculated and the higher of the two layers with the biggest difference is chosen as the present layer. The secondary clarifier has 4 m water depth and leading to very small TSS difference between the layers. More complex multi-dimensional hydrodynamic models produce more accurate prediction, but they were demonstrated to be not suitrable for on-line use due to their high complexity and high computational demand. Therefore, we do not consider the SBH parameter for the secondary settler performance check. Instead we consider the following candidates:

- Sludge retention time, SRT;
- Food to microorganisms ratio, F/M;
- Mixed liquid suspended solids, MLSS;
- Nitrate/nitrite concentration at the exit of the anoxic zones, S_{NO}^{p3};
- Effluent ammonia, S_{NH}^{eff}.

4.3.2 Optimization Procedure

As a starting point the nominal operation conditions used the TecnoCasic were considered and Table 4.4 energy consumptions in these conditions are reported. The Table shows that it is more beneficial to find savings by lowering the energy consumption for aeration than for anything else. Reducing pumping energy and sludge recirculation will hardly lead to significant savings.

	Total Energy Consumption		Percentage on total ASP
Aeration	2178.00	€	99%
Pumping	21.75	€	0.99%
Sludge Disposal	0.25	€	0.01%
Total ASP	2200.00	€	100%

Table 4.4: Relative energy consumption at nominal conditions

Therefore, the attention firstly focuses on aeration and some interesting considerations can be drawn on the *DO* control which is actually present in the considered wastewater treatment plant. The average of two oxygen sensor signals located in the anoxic and aerobic zones is compared with a constant averaged *DO* setpoint of 2.5 gO_2/m^3. The controller maintains the desired oxygen setpoint by manipulating the ratio between the two zones. This leads to a very high oxygen supply in the aerobic zone ($4 - 5\, gO_2/m^3$) and a very low oxygen supply in the anoxic zone ($0.005 - 0.09\, gO_2/m^3$). During the degree of freedom analysis the airflow rate has not been considered even though, a first optimization needs to be performed in order to find the oxygen optimal setpoint. The aim of this optimization is to reduce the oxygen consumption in the bioreactor without influencing the overall system behavior.

In the Sections 4.1 and 4.2, we have defined the objective functions and the degrees of freedom of our optimization problem. This means that we must find the constrained minimum of a scalar function of several variables starting from an initial estimate. This is generally referred to as constrained nonlinear optimization. The nonlinear problem is approached splitting it into smaller subproblems that can then be solved and used as the basis of an iterative process. The problem was solved using a *Sequential Quadratic Programming*, SQP, techniques which represents state-of-the art in nonlinear programming methods. The *Kuhn-Tucker* equations are necessary conditions for optimality in this problem and the solution of these equations forms the basis to SQP algorithms [11].

Tracking the activated sludge model with Simulink, we solved the optimization problem for the considered wastewater treatment plant, satisfying the constraints reported in Table 4.1 and in Table 4.2. Generally speaking, the optimization procedure is conceptually the same both in the one degree and two degrees of freedom case. In order to define the optimal setpoints for the candidate controlled variables, the optimum is found for the nominal conditions simulating different disturbances with weight drawn form a probability distribution function assumed normal.

As mentioned, a first optimization is performed in order to reduce the oxygen consumption in the bioreactor (which leads to an overall cost reduction) and then optimal values for the dissolved oxygen concentration in both controlled anoxic and aerated zone have been found. During, this first optimization also the considered manipulated variables for the self-optimizing approach were considered in order to reduce the computational load during the second optimization procedure. This is performed in order to find

robusts setpoints for our candidate controlled variables (Section 4.3.1). Results regarding the first optimization at nominal conditions are reported in Table 4.5. A remarkable cost reduction is obtained.

	Initial Condition	One DOF Optimization	Two DOF Optimization
DO^p in $[gO_2/m^3]$	0.09	0.21	0.25
DO^n in $[gO_2/m^3]$	4	2.5	2.5
Aeration Cost in €	2200	1470	1464

Table 4.5: Optimization of oxygen setpoints

The Simulink model has been run for 100 days simulation using weighting functions for both load and flow. Only the last 15 days corresponding for stable regimes were considered when evaluating the cost function and the constraints.

4.3.3 One degree of freedom

The one degree of freedom situation corresponds to the case where the influent flow rate is constant. This means that we optimize only with respect to one manipulated variable: Q_w. As previously stated, in order to define the DO concentrations (able to reduce the oxygen consumption in the reactor) a first optimization is performed. As an interesting note, we highlight that the cost function in our activated sludge process is monotonically increasing with disturbances: that is, a minimum cost will be always obtained at minimum disturbance. For this reason, we perform the first optimization procedure reducing the influent load of 10%, obtaing in this way an optimal reference cost $J_{opt} = 1022$ €. Furthermore, the values found for the dissolved oxygen concentration are reported:

- $DO^p = 0.21\, gO_2/m^3$;
- $DO^n = 2.5\, gO_2/m^3$.

Using oxygen concentration values as constant setpoints for the aeration control allows us to investigate the steady state process behavior at nominal conditions and different Q_w.

In Figure 4.3, the effluent concentrations (related to the effluent constraints) are plotted against the Q_w flowrate. We note that the effluent COD concentration (Figure 4.3a) stabilizes at a constant value, even if care should be taken if disturbances such as d_1 or d_3 occur (that is influent COD variations). The total suspended solids concentration (Figure 4.3b) in the effluent

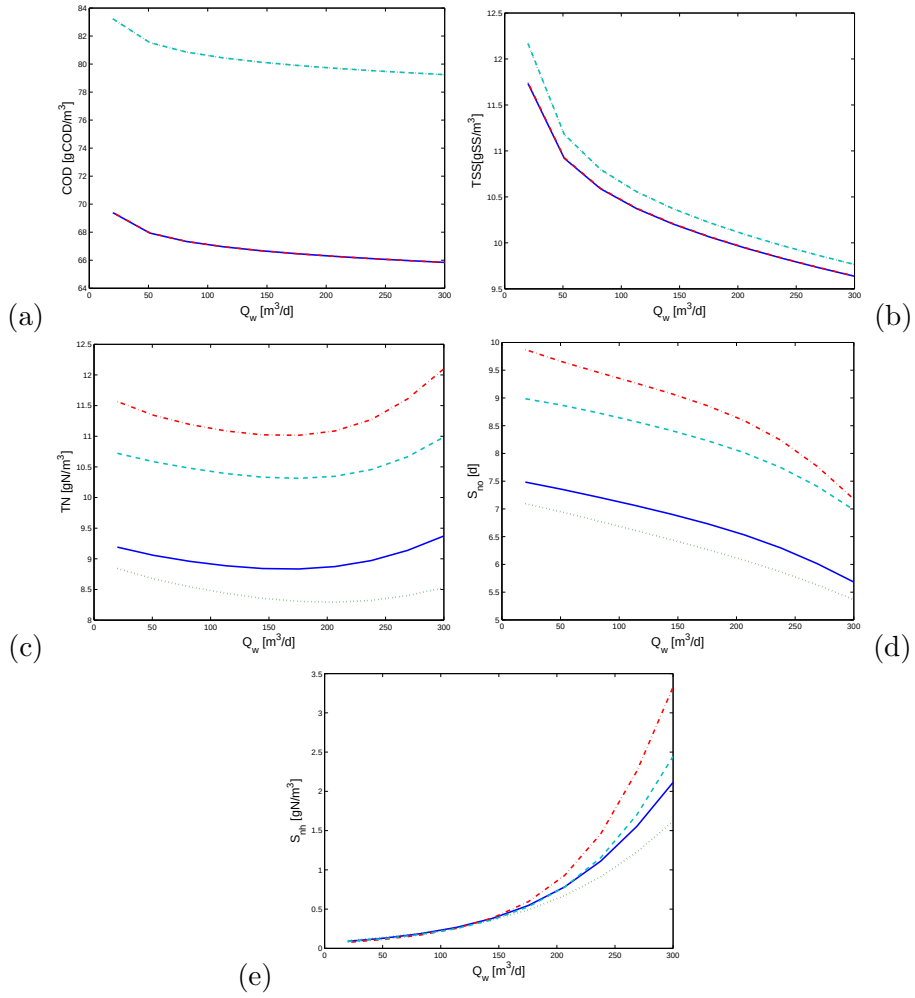


Figure 4.3: One degree of freedom: process composition at different Q_w [Nominal condition (solid); d_1 (dot); d_2 (dot-dashed); d_3 (dashed)].

seems not to be so sensitive to disturbances. The effluent TN concentration (Figure 4.3c) is not very sensitive to wastage flowrate variations, whereas it is very sensitive to disturbaces such as d_2 (that is, influent TKN variations). The same can be noted for S_{NO}^{p3} (Figure 4.3d), whereas the effluent ammonia (Figure 4.3e) is not very sensitive to disturbances but increases with the wastage flowrate. For all the effluent concentration in Figure 4.3, the constraints are respected within the same range as SRT and F/M.

In Figure 4.4, the air consumption for both denitrification (a) and nitrification (b) zones and also the total air consumption (c) are plotted against the

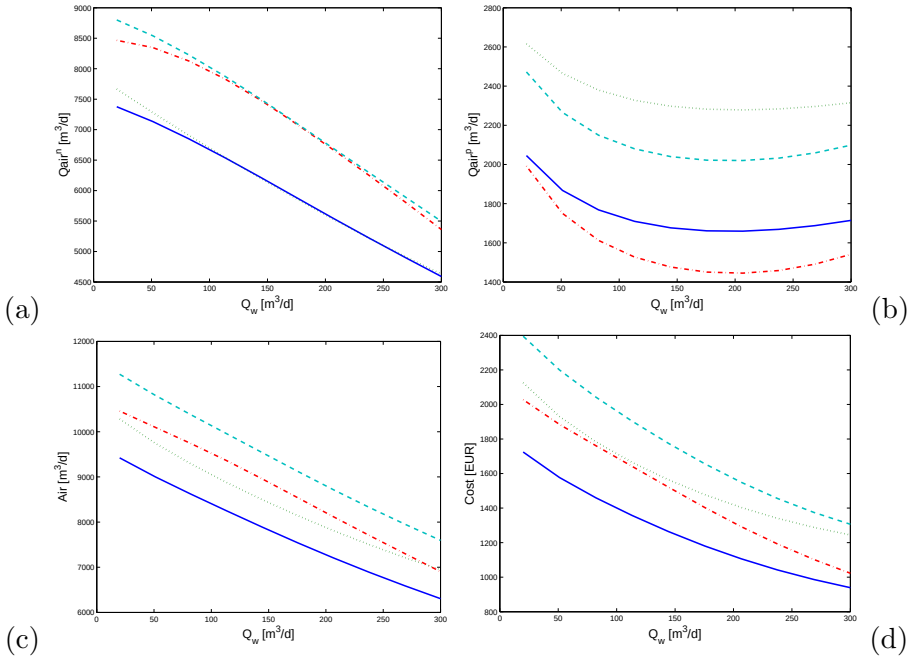


Figure 4.4: One degree of freedom: Costs at different Q_w [Nominal condition (solid); d_1 (dot); d_2 (dot-dashed); d_3 (dashed)].

Q_w flowrate. As expected, the air consumption decreases with the increase of the wastage flowrate. This is obvious, since increasing Q_w we are also increasing the total sludge mass inside the bioreactor. This is reflected also in the total cost, which decreases with the increase in Q_w: but, looking at the optimization contraints we can observe that an *optimal* cost of $1500 - 2000$ € for a wastage flowrate within the range $60 - 100\, m^3/d$ is expected.

The analysis is also useful to make some assumptions to select the controlled variables. From Figure 4.5, we notice that with a sludge flow rate between $60 - 80\, m^3/d$, the SRT and F/M constraints are respected. We also note that SRT and F/M are insensitive to disturbances, whereas S_{NO}^{p3} is not, and variations in TKN lead to variations in the nitrate and nitrite concentrations with different Q_w values. In the region of interest, the effluent ammonia concentration is insensitive to disturbances, whereas it becomes sensitive as the wastage flowrate increases. Moreover, the MLSS is only weakly sensitive to disturbances (of course, when d_1 and d_3 occur the suspended solids in the mixed liquor increase). Recalling the qualitative rules exposed in Chapter 3, a first screening in the candidate variable selection can be made. In fact, we note that S_{NO}^{p3} is not responding to any of those

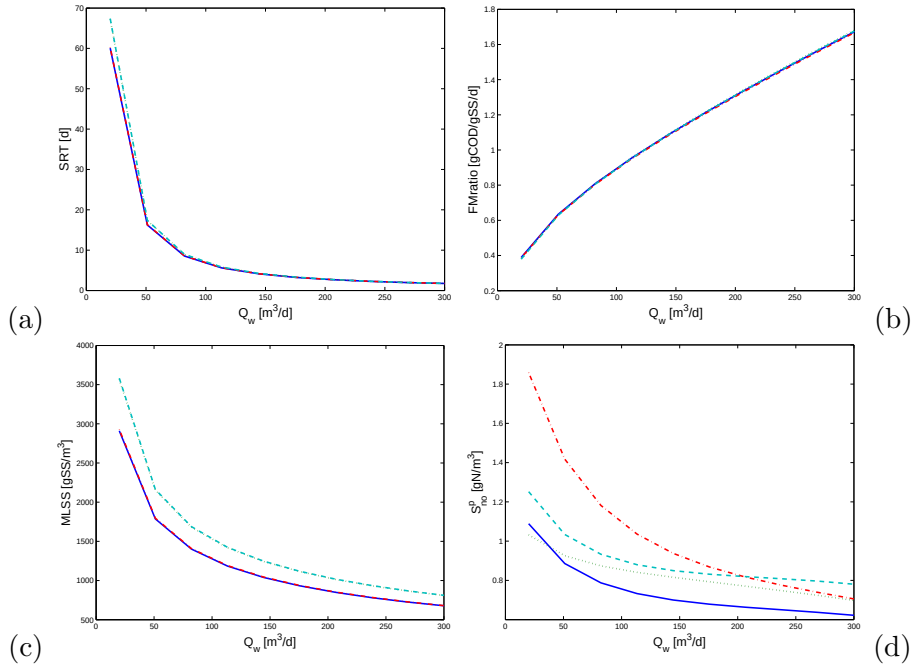


Figure 4.5: One degree of freedom: operating variables at different Q_w [Nominal condition (solid); d_1 (dot); d_2 (dot-dashed); d_3 (dashed)].

suggestions: it presents sensible variations to disturbance but negligible variations to the manipulated variable Q_w. For this reason, the system behavior when SRT (c_1), F/M (c_2), MLSS (c_3) and ammonia in the effluent (c_4) are the candidate controlled variables is investigated. In Table 4.6 the setpoints for the candidate controlled variables are given, and in Table 4.7 the investigation results are reported.

c_1^{sp} =SRT	c_2^{sp} =F/M	c_3^{sp} = MLSS	$c_4^{sp} = S_{NH}^{eff}$
9.5 $[d]$	0.74$[gCOD/gSS/d]$	1194$[gSS/m^3]$	0.17$[gN/m^3]$

Table 4.6: Setpoints for the candidate controlled variables at 1DOF

Recalling the loss definition given in Chapter 3, the process loss at different disturbances and different control configurations can be computed. We analyze also the *open loop* strategy. This is a poor policy to adopt, but is a good reference to understanding how the system could operate if there were no control applied. The results are reported in Table 4.7, where we note

		Loss [€]		
	Nominal	d_1	d_2	d_3
c_1	459	772	765	1037
c_2	463	788	768	1053
c_3	**337**	**502**	**662**	**701**
c_4	469	816	745	1040
Open Loop	455	779	762	1045

Table 4.7: Loss Investigation in the one degree of freedom case

that c_3 (the suspended solids in the mixed liquor) gives the minimum loss and SRT, F/M and also the nominal conditions give almost the same losses. This is not surprising if we think that the first optimization have been already performed and that the open loop results implements the (constant) manipulated variables value found with that optimization.

Once the MLSS setpoint is determined the realization of the feedback strategy is straightforward, assuming that the time scale of the dynamics is understood. In Figure 4.6 the main effluent concentrations and the air consumption cost for the controlled system for different disturbances are showed. Subsequently, we note that the system satisfies the effluent concentration constraints and, as expected, the air consumption increases as the disturbances occurs, but with minimum loss for the process.

In Figure 4.7a the controlled variable SRT is showed, and in Figure 4.7b Q_w is drawn as manipulated variable. Since we can not keep the controlled variable at a constant setpoint, large variations in the manipulated variable arise when disturbances appear. This behavior demonstrates that it is not a good idea to keep SRT at constant value. In fact, as previously mentioned, the sludge age represents in some sense the nitrification capacity of the sludge which, on the other hand should be allowed to vary. This is expecially true with nitrogen load variations.

4.3.4 Two degrees of freedom

In the two degrees of freedom case, we assume that there are some failures in the equalization tank (if this unit is actually present in the wastewater treatment plant), or that the equalization tank is not present at all. This is equivalent to assume that the influent flowrate can not be considered constant. Therefore, we must optimize with respect to two manipulated variables: the Q_w and the Q_r. In Section 4.2.2, we observed that usually the recycled flow rate is proportional to the influent flowrate. Here, instead of

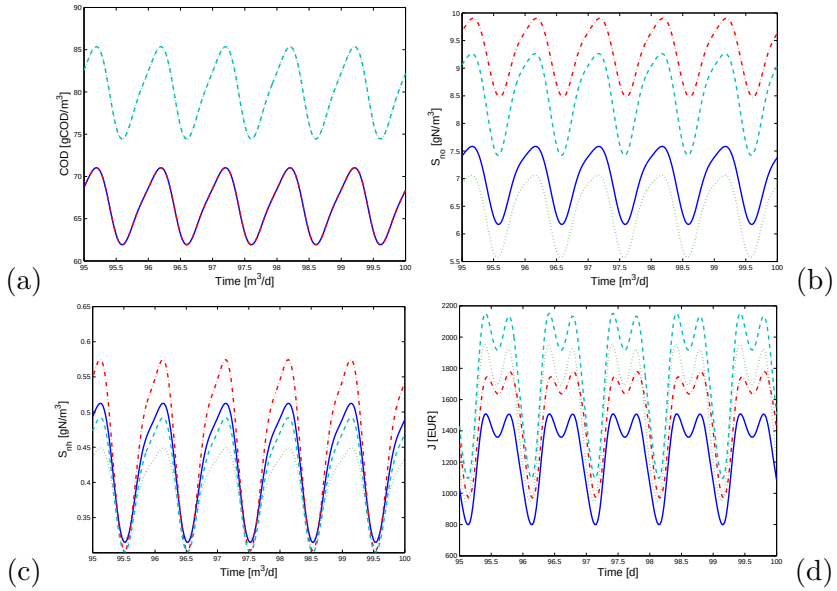


Figure 4.6: Effluent composition and Cost with c_3 controlled [Nominal condition (solid); d_2 (dot-dashed); d_3 (dashed)].

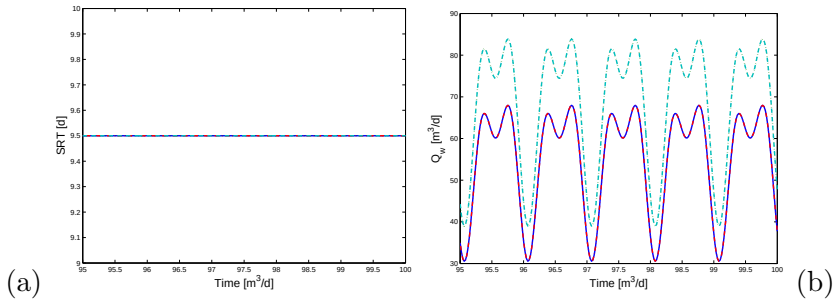


Figure 4.7: Controlled c_1 and Manipulate Q_w variables [Nominal condition (solid); d_1 (dot); d_2 (dot-dashed); d_3 (dashed)].

optimizing with respect to Q_r, we optimize with respect to the ratio Q_r/Q_{in}. As in the one degree of freedom case, a first optimization has been performed decreasing the influent flowrate and also the load concentration (leading to an the optimum reference cost value for the cost function and optimal oxygen setpoints for the anoxic and aerobic zones). The corresponding results are:

- $DO^p = 0.22\, gO_2/m^3$;
- $DO^n = 2.5\, gO_2/m^3$.

In such a case, we graphically investigate the process behavior using the *operating space diagrams*. An operational space diagram is a contour plot of an output variable against the manipulated variables. It is constructed by calculating the steady state of the process over a range of values for the manipulated variables. In the wastewater treatment field, they were introduced by Hopkins *et. al* [44] in order to define an efficient tool to design a plant and to compare a large number of operational strategies. In our study, it helped us to identify the behavior for the function J. The diagrams provide information of the optima and their relation to the constraint, therefore showing how close they are to the constraints. In addition also information about the sensitivities of the output to the manipulated variables are provided.

In Figures 4.8 and 4.9 the operational maps of the nominal conditions are reported. It was verified that the system behaves in the same way it does in the one degree of freedom case when disturbances occurs. As we can note, in the two degrees of freedom case the operating region is between $60-90\,m^3/d$ with respect to the wastage flowrate Q_w, whereas, Q_r is limited to vary within 7000 and 9000 m^3/d (corresponding to $1.14 < Q_r/Q_{in} < 1.5$ at the average influent flowrate).

Figure 4.8 reports the operational constraints maps: from there, we note that the sludge age constraint (Figure 4.8a) is satisfied only for $40\,m^3/d < Q_w < 80\,m^3/d$. The diagram shows that SRT increases too much as the wastage flowrate goes to zero, whereas the dependency with Q_r is not so high. The same can be said for the F/M ratio constraints (Figure 4.8b), which is respected throughout the operation region. The nitrate/nitrite concentration in the anoxic zone (Figure 4.8c) shows weak dependency with the two manipulated variables. However, its constraints are respected in the region of interest. In the same Figure, the main effluent concentrations are also reported. The effluent organic matter expressed in terms of COD is respected in the region of interest (Figure 4.8d). The same consideration applies to the ammonia (Figure 4.8d), and nitrate/nitrite concentrations (Figure 4.8e), in the effluent constraints.

As an interesting investigate, we look at Figure 4.9: here the air flowrate in both bioreactor zones and the total air consumption are reported. We note that the total cost is not actually dependent on variations in the recycled flowrate, whereas it varies with Q_w and it decreases as the wastage flowrate increases (as in the one degree of freedom case).

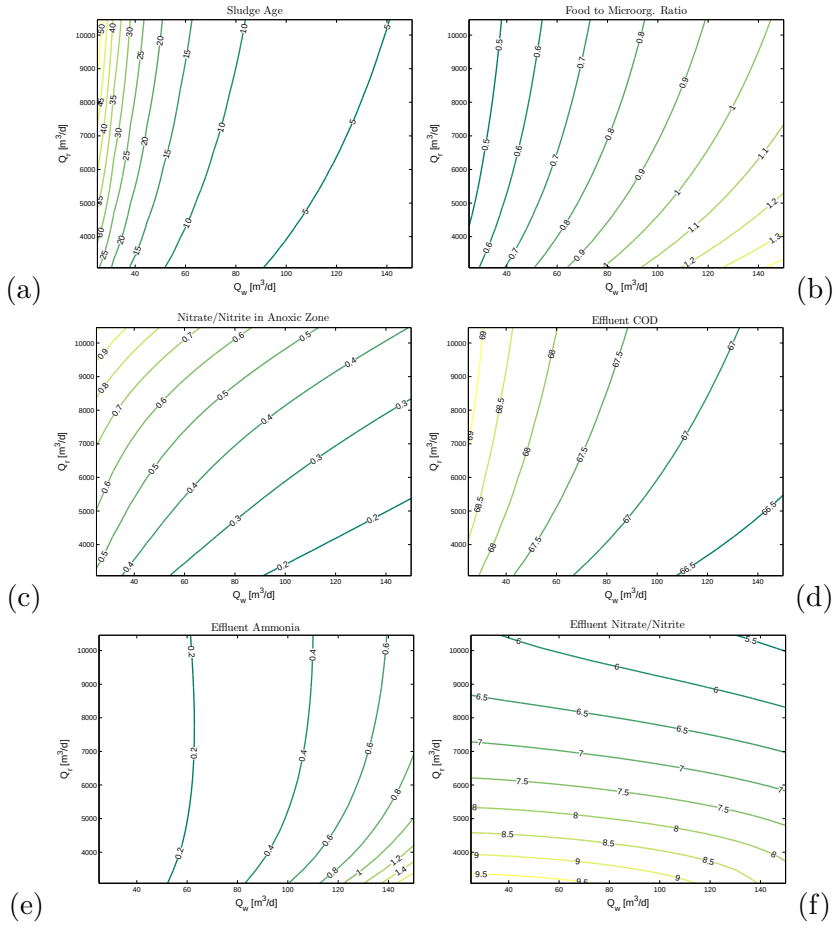


Figure 4.8: Operative maps for constraints

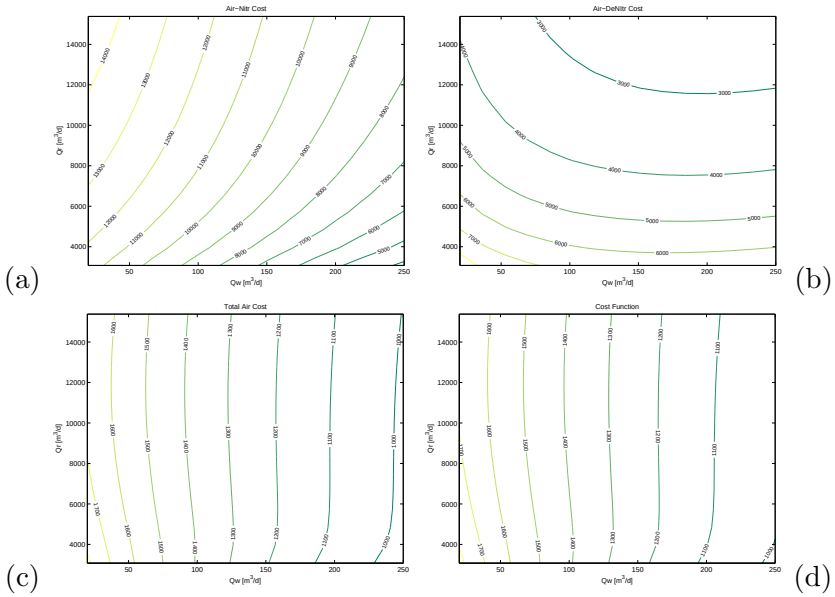


Figure 4.9: Costs operative maps

Having defined an appropriate operating region from the maps, the procedure goes further to defining the control structure for the two degrees of freedom case. As exposed in Chapter 3, we apply the minimum singular value rule. The controlled variables set corresponding to the larger value of $\underline{\sigma}$ (the minimum singular value) are selected. The procedure is summarized as follows:

1. Linearize the model and find the scaled gain matrix.

2. Compute the minimum singular value for the whole system (it will be the reference value).

3. Define the possible control configuration.

4. Choose the controlled variable configuration with larger minimum singular value.

Using the Simulink model and Matlab, the linearization task has been performed, around linearization points found from simulations. The obtained linear model has the usual state-space form:

$$\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x} + \mathbf{B}\,\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\,\mathbf{x}$$

where $\mathbf{x}$ is the state vector of 158 state variables[2]. The input vector is given as:

$$\mathbf{u} = \begin{bmatrix} Q_r \\ Q_w \end{bmatrix} \tag{4.9}$$

and the output state vector is given by:

$$\mathbf{y} = \left[SRT,\, F/M,\, S_{NH}^{eff},\, MLSS,\, S_{NO}^{p3} \right]^T \tag{4.10}$$

It follows that $\mathbf{A}$ is a 158×158 square matrix, whereas the dimension of $\mathbf{B}$ is 158×2 and $dim(\mathbf{C}) = 5 \times 158$.

To obtain a model that is well conditioned and numerically easier to solve, scaling procedure is performed. Scaling the inputs and the candidate measurements is also a very important task before comparisons are made. This has been done by introducing scaled variables: $\mathbf{u} = \mathbf{D}_u^{-1}\mathbf{u}^0$ and $\mathbf{y} = \mathbf{D}_y^0\mathbf{y}^0$. The original model can be written as: $\mathbf{y}^0 = \mathbf{G}^0\mathbf{u}$. The superscript "0" denotes the nominal state, $\mathbf{G}^0$ is the original transfer matrix between $\mathbf{y}^0$ and $\mathbf{u}^0$, $\mathbf{D}_u$ and $\mathbf{D}_u$ are diagonal scaling matrices estimated using some heuristics reported in [101]:

$$\mathbf{D}_y = \begin{bmatrix} 1.72 & 0 & 0 & 0 & 0 \\ 0 & 0.44 & 0 & 0 & 0 \\ 0 & 0 & 0.08 & 0 & 0 \\ 0 & 0 & 0 & 0.60 & 0 \\ 0 & 0 & 0 & 0 & 25.9 \end{bmatrix} \quad \mathbf{D}_u = \begin{bmatrix} 12966 & 0 \\ 0 & 110 \end{bmatrix}$$

Thus, the scaled model is given as:

$$\mathbf{G} = \mathbf{D}_y^{-1}\mathbf{G}^0\mathbf{D}_u$$

allowing to reduce the condition number by one order of magnitude.

After the scaling, the model gain matrix and the RGA-matrix are expressed as:

$$\mathbf{G} = \begin{bmatrix} 6.35 & -12.24 \\ -0.92 & 1.4 \\ -0.27 & 2.55 \\ 1326 & 2035 \\ 0.04 & -0.015 \end{bmatrix} \quad \Lambda(\mathbf{G}) = \begin{bmatrix} -2.25 & 2.82 \\ 1.6 \times 10^{-6} & 1.1 \times 10^{-6} \\ -0.08 & 0.51 \\ 3.33 & -2.33 \\ 3.2 \times 10^{-4} & 9.5 \times 10^{-4} \end{bmatrix}$$

[2]It must be recalled that we are considering a model composed by 3 anoxic zones (3×13 state variables), 3 aerobic zones (3×13 states variables) and a settler with 70 soluble + 10 particulate state variables.

The singular value decomposition of $\mathbf{G}$ gives two non-zeros singular values, the minimum being $\underline{\sigma}^{ref} = 1.79$.

The RGA-matrix of $\mathbf{G}$, the overall non-square matrix, is a useful screening tool when there are numerous candidates [101]. The five row-sums of the RGA-matrix are:

$$\Lambda_{\Sigma} = [0.57,\ 5 \times 10^{-7},\ 0.43,\ 1,\ 4.2 \times 10^{-4}]^T \tag{4.11}$$

indicating that one should select MLSS and SRT (corresponding to the two largest elements), to maximize the projection of the selected outputs onto the space corresponding to the two non zero singular values. We verify in the following that is assumption is not the right one. In fact, using the minimum singular value rule we have a more reasonable criterion for selecting the controlled outputs.

Recalling the rule exposed in Chapter 3, we select a set of outputs such that the elements in $\mathbf{G}^{-1}(0)$ are small, or alternatively such that $\underline{\sigma}(\mathbf{G}(0))$ is as large as possible. In Table 4.8, the possible configurations are reported with the associated minimum singular value and condition number. From there we note that the configuration giving the larger minimum singular value is the forth, c_4. The sludge retention time and the nitrate/nitrite composition in the anoxic zone are indicated as controlled variables. However, also the configuration c_7 and c_{10} present a minimum singular value very close to the reference value, so we should take into account also those configurations. On the other hand, c_9 presents a too high condition number to justify its inclusion.

Next step is to determine whether and how those variables can be coupled with the manipulated variables: that is $u_1 = Q_r$ and $u_2 = Q_w$, in order to define the decentralized control configuration. For this case the RGA of the selected square transfer function matrices $\mathbf{G}_4$, with the two inputs is considered:

$$\Lambda(\mathbf{G}_4) = \mathbf{G} \times \mathbf{G}^{-T} \tag{4.12}$$

Specifically, inputs and outputs variables should be paired so that the diagonal element of the RGA are as close as possible to unity:

$$\Lambda(\mathbf{G}_4) = \begin{bmatrix} 0.1897 & 0.8103 \\ 0.8103 & 0.1897 \end{bmatrix} \quad \Lambda(\mathbf{G}_7) = \begin{bmatrix} 0.227 & 0.773 \\ 0.773 & 0.227 \end{bmatrix}$$

$$\Lambda(\mathbf{G}_{10}) = \begin{bmatrix} 0.227 & 0.773 \\ 0.773 & 0.227 \end{bmatrix}$$

	Configuration	$\underline{\sigma}$	$Con.Number$
c_1	SRT-F/M	0.164	85.00
c_2	SRT-S_{NH}^{eff}	0.925	15.14
c_3	SRT-MLSS	1.315	43.27
c_4	SRT-S_{NO}^{p3}	**1.632**	8.45
c_5	F/M-S_{NH}^{eff}	0.654	4.58
c_6	F/M-MLSS	0.000	$5.4 \times 10^{+15}$
c_7	F/M-S_{NO}^{p3}	**1.522**	1.17
c_8	S_{NH}^{eff}-MLSS	0.063	133.62
c_9	S_{NH}^{eff}-S_{NO}^{p3}	3.970	973.54
c_{10}	MLSS-S_{NO}^{p}	**1.615**	34.18

Table 4.8: Minimum singular value for the proposed configurations

Furthermore, the Niederlinski Index is always greater then zero, as we can note from Table 4.9. Hence, the loop pairing is acceptable and leads to a stable control configuration.

	c_4	c_7	c_{10}
N	5.47	3.61	3.61

Table 4.9: Niederlinski Index for the candidate controller sets

From the RGA analysis, follows the pairing:

- control the sludge age (or the FM ratio or the MLSS) by manipulating the wastage flowrate;
- control the nitrate/nitrite concentration by manipulating the recycled flowrate.

Before approaching the controller design step, the setpoints for the controlled variables must be defined. As we have done for the one degree of freedom case, the setpoints are calculated by optimization and they are reported in Table 4.10.

Once the optimal value is found (in this case, the optimization procedure returnes $J_{opt} = 1015$ €) we define the loss at steady-state conditions. In Table 4.11, the loss values for each control configuration are reported. We note that even if c_4 is the best configuration, in terms of loss c_{10} gives more apropriate results. Controlling the SRT implies keeping the nitrification capacity of the sludge (measure of the maximum nitrification rate)

SRT^{sp}	F/M^{sp}	$MLSS^{sp}$	$S_{NO}^{p3\_sp}$
10.8 $[d]$	0.72 $[gCOD/gSS/d]$	1268 $[gSS/m^3]$	0.78 $[gN/m^2]$

Table 4.10: Setpoints for the controlled variables in the 2DOF

		Loss [€]		
	Nominal	d_1	d_2	d_3
c_4	482	1499	1473	1822
c_7	431	1479	1462	1805
c_{10}	**410**	**1374**	**1400**	**1741**
Open Loop	450	1420	1444	1870

Table 4.11: Loss Investigation in the two degree of freedom case

at a constant level, and especially when the flow rate and load are not constant this should be allowed to develop in the system as a result of an increase influent. The results is analogous in the one degree of freedom case.

To verify the system behavior, dynamic simulations with periodic influent flow and load are performed. The effluent conditions are investigated with this configuration and in Figures 4.10 the main effluent concentrations and also the corresponding cost results are reported. From there we can note that the effluent concentrations are kept at the constraint levels and that the cost function increases when disturbances occur (Figure 4.10d). This is not surprising because, as stated previously, the cost function for our system is a monotonically increasing function, and this is more evident especially when periodic disturbances take place.

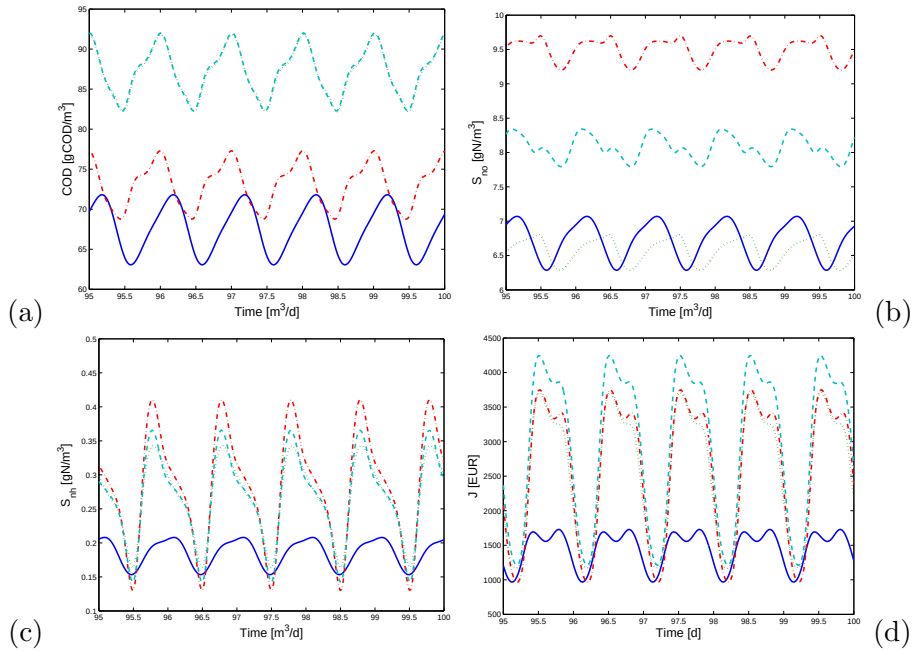


Figure 4.10: Two degrees of freedom: effluent concentrations and cost for the proposed controller [Nominal condition (solid); d_1 (dot); d_2 (dot-dashed); d_3 (dashed)].

Part II

Model Order Reduction and Software Sensors

So far, in the controlled variables selection we assumed that every measure would be available at every time. However, in the wastewater treatment, variables such as concentrations, SRT or F/M, are determined by off-line laboratory analysis, limiting the on-line monitoring and control purposes. In a successful manner, concentration can be estimated on-line using a *soft-sensor*, which represents a combination of robust hard-sensors and a mathematical model defined to reconstruct the time evolution of the unmeasured states. An important advantage is that soft sensors can be constructed on the basis of a simple model with uncertain inputs and parameters.

With regard to the activated sludge process the state of the art model has been considered. This is a rigorous model that represents in a fairly complete sense the biological reactions. It is useful for several reasons: for instance in system design to check the system behavior under extreme dynamic loading conditions. However, there are many practical applications where a simplified model of reduced order is sufficient from an engineering point of view: e.g., for a soft-sensor design. In fact, we have seen that the full model suffers the following disadvantages: first of all, the calibration of all the kinetic and stoichimetric parameters is a cumbersome task and secondly, the model is nonlinear due to the presence of Monod like kinetics in the mass balance equations. Moreover in the wastewater treatment, variables such as concentrations are determined by off-line laboratory analysis, limitating for on-line monitoring and control purposes.

The goal in the following is to select a suitable reduced order model for such purposes. The following properties should be satisfied:

- The stiffness of activated sludge process argues to develop models that are suited for different timescales. Therefore, model reductions based on timescale separation are the natural approach.
- It is desired that the reduction method is systematic and straightforward to avoid the time consuming trial-and-error and iteration methodology.
- The states of the model must retain their physical interpretation after reduction, so that the interpretability is preserved.

Therefore, the following Chapters give first of all an overview on model order reduction techniques (Chapter 5). This overview is then useful as background to introduce the reduced models already present in literature and to deduct a new reduced model based on numerical techniques and

physical considerations (Chapter 6). Eventually, in Chapter 7 applications on an activated sludge process using soft sensors, using reduced models are shown.

Chapter 5

Overview on model reduction techniques

This Chapter describes methods for reducing the order of the process model. Keeping in mind that the final aim is to find a simpler model for the activated sludge process, the goal here is to describe these methods that represent the starting point for the reduction procedure. Generally speaking, there are two principlal model reduction techniques:

- intuitive model reduction;
- mathematical techniques.

The first one implies that the modeller uses his/her knowledge and experience of the true process dynamics, in combination with the defined purpose of the new model, to define a simplified model. Assumptions can be made with respect to:

- components, e.g., aggregation of variables, such as COD as a measure of pollutant concentrations;
- processes, such as aggregation of reaction: e.g., modelling nitrification as one-step process whilst it is a two-step process;
- kinetics.

Even if these simplifications may affect both the model structure and its functional relationships in the model, this is the most common approach used to simplify models [52]. However, a more rigorous mathematical approach is preferred here, this is also the reduction approach used to develop our reduced model in Chapter 6. Indeed, a wide range of mathematical methods for model reduction are available in literature. We place particular

emphasis on reducing model techniques that preserve stability, controllability and observability of the full model.

The Chapter is organized as follows: after a problem statement, the basic distinction between linear and nonlinear model reduction approach is made. Even if the ASM1 is a nonlinear model, we focus on linear models (assuming that the ASM1 has to be linearized) and for the sake of completeness the nonlinear model approaches are also reported.

5.1 Problem Statement

Model reduction has a long history in system and control literature. In fact, the general topic of dimension reduction in dynamical systems is pervasive in applied mathematics literature. The system is modelled by means of a set of first order differential equations coupled with a set of algebraic equations:

$$\begin{aligned} \dot{\mathbf{x}} &= f(\mathbf{x}(t), \mathbf{u}(t)) \\ \mathbf{y} &= h(\mathbf{x}(t), \mathbf{u}(t)) \end{aligned} \tag{5.1}$$

In this setting, $\mathbf{u} \in \mathbb{R}^m$ are the inputs, $\mathbf{x} \in \mathbb{R}^n$ are the state variables and the function f describes the system dynamics. The outputs are denoted as $\mathbf{y} \in \mathbb{R}^q$ are the outputs or set of observations and h describes the way that the observations are deduced from the states and the inputs. The system complexity is defined by the number of states n.
The first requirement is that the number of states (i.e., the number of first order differential equations) of the approximant model is less than that of the original system: $k < r$. Obviously, in absence of additional requirements, this condition is easy to satisfy by means of mere elimination of equations and state variables. The difficulty arises when imposing additional limitations, such as:

- small approximation errors;
- preservation of stability;
- computationally efficient procedure.

The majority of the model reduction techniques have been developed for linear models. Despite this, most model reduction methods are inappropriate because they are invariably loosing the physical meaning. For example, reduction techniques such as *principal component analysis* frequently make use of linear transformations to reduce the model dimensions. These transformations invariably lead to an alternative coordinate system where state

variables have little physical significance. A different approach is the use of *structural dominance* concepts. Measures indicative of the strength if the coupling between model components are developed. By neglecting weak couplings, a reduced order model can be formed.

5.2 Linear systems

Model reduction techniques for linear models are based on first principle criteria, empirical criteria or a combination of both. In practice, we consider a special case of the model (5.1):

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \end{aligned} \tag{5.2}$$

In equation 5.2 a continuous-time, time-invariant, linear, dynamical system is presented: $\mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^n \times \mathbb{R}^m$, $\mathbf{C} \in \mathbb{R}^q \times \mathbb{R}^n$ and $\mathbf{D} \in \mathbb{R}^q \times \mathbb{R}^m$. Together with the initial conditions $\mathbf{x}(t_0) = \mathbf{x}_0$, the realization in equation 5.2 is uniquely described by the matrix 4-tuple $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. The equivalent notation:

$$\mathbf{G} = \left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right) \in \mathbb{R}^{(n+q)} \times \mathbb{R}^{(n+m)} \tag{5.3}$$

is more common in control theory.
Assuming that we are given a realization ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$) of order n. The purpose of model reduction is to find a reduced realization $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}})$ with $\hat{\mathbf{A}} \in \mathbb{R}^k \times \mathbb{R}^k$, $\hat{\mathbf{B}} \in \mathbb{R}^k \times \mathbb{R}^m$, $\hat{\mathbf{C}} \in \mathbb{R}^q \times \mathbb{R}^k$, such that the input-output behaviour of the reduced system approximates the original system in some sense. In a different way but with the same scope, the aim is to show how the state vector $\mathbf{x}$ can be partitioned into more important states $\mathbf{x}_1$ and less important states to be eliminated.

Modal truncation [21], *singular perturbation* [29], *balanced truncation* [77], [108], *optimal Hankel norm approximation* [33], *frequency weighted balanced truncation* [28] are well known model reduction methods for stable systems (see [4] for an overview of the different model reduction approaches). In the following sections, we describe only the methods that we applied to the activated sludge model, making a fundamental distinction between modal reduction (Section 5.2.1) and balanced reduction (Section 5.2.2) approaches.

5.2.1 Modal Reduction

Reduction methods such as truncation or singular perturbation require the solution of eigenvalues problem of order n. This makes their standard imple-

mentation computationally expensive for large systems ($n > 1000$). However, they are very useful for system of moderate size ($n \in [5, \ldots, 1000]$) like the ASM1 case. Moreover, they provide the mathematical basis for reduction by time scale separation, and represent systematic techniques. The reduction methods rely on two mathematical techniques that exhibit, the eigenvalues to states association and the reduction method itself. The reduction techniques depend on the successful application of an eigenvalue to state association algorithm. In the following Sections 5.2.1.1 we see how the eigenvalue can be linked to the corresponding state. In Sections 5.2.1.2 and 5.2.1.3 the information given by the eigenvalues are used in model reduction.

5.2.1.1 Eigenvalues to states association

Eigenvalues are mathematical properties of a model that provide valuable information on the transient response of the whole system. If they can be linked to the various model states, then the dynamics of each state can be classified and the states dynamics at either extremes (fast or slow) can be removed from the reduced model.
To determine the eigenvalues, the corresponding algebraic equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{X} = 0$ is solved, where $\mathbf{I}$ is an identity matrix of the same dimension of $\mathbf{A}$ and $\mathbf{X}$ is a matrix of vectors called *eigenvectors*. For this equality to be true the determinant of the term in brackets must be zero. This represents a simple technique for calculating the eigenvalues by hand. Other numerical methods, not discussed here, are used in practice.
Several important properties related also to the process stability can be deduced from the eigenvalues analysis. Remember that the process is stable if after a period of time, the variable return to the initial steady-state value. Numerically, this can be translated into:

- positive real component imply instability;
- large imaginary components indicate oscillatory dynamics;
- large negative components imply a fast decay to steady-state.

Now, the question that may arise is *how do the eigeinvalues relate to the state variables?* For this system an unknown correspondence exists between states and eigenvalues.

The mathematical *homotopy method* is used to aswer this question. It was proposed by DeCarlo and Saeks [22], and later defined by Robertson [89]. Furthermore, in literature several examples of its applications can be found: for example, Monge and Georgakins [76] applied it to catalytic cracking processes, Robertson and Cameron [90] studied the startup and shutdown

modeling of an evaporator system, Steffens *et al.* [105] and Chachuat [16] utilized it for the ASM1.
The homotopy method involves starting with a system with an obvious relationship between eigenvalues and states: a diagonalized **A** matrix. In such a situation the correspondence between states and eigenvalues is known: each diagonal element corresponds to an eigenvalue which is, in turn, directly related to the states. This system is then transformed using homotopy into the actual system while tracing the eigenvalues. A suitable homotopy function is:

$$\mathbf{H}(r) = (1-r)\mathbf{A}_D + r\mathbf{A} \tag{5.4}$$

where **H** is the homotopy matrix, **A** is the linearized model matrix, $\mathbf{A}_D$ is the diagonal matrix of **A** and $r \in [0,1]$ is a parameter to achieve the linear progression. In the homotopy procedure, all transformed systems must exhibit a strong degree of association between states and system modes, i.e. the eigenvalues. If an intermediate system is encountered (where this direct association does not exist), then no definitive association can be concluded for the final system **A**. If the eigenvalue does not change with r, then the eigenvalue to state association is relative obvious.

When the eigenvalue to state association is defined it is possible to start reducing the model.

5.2.1.2 Truncation

Truncation is a common form of model reduction. In fact, it is the most common form since every finite dimensional linear model is a truncated model in the sense that there is always some dynamic aspect of the physical system that is neglected. We say that truncation deletes some of the modes (*modal truncation*) or states (*states truncation*) from the full order model. If the equivalence between modes and states is established with the homotopy analysis, the two methods are equivalent.
Considering the truncated variables at their steady-state values, the $k-$ order truncation of the system in (5.3) is given as:

$$\mathbf{G}_{trunc} = \left(\begin{array}{c|c} \mathbf{A_{11}} & \mathbf{B_1} \\ \hline \mathbf{C_1} & \mathbf{D} \end{array}\right) \tag{5.5}$$

The truncated model $\mathbf{G}_{trunc}$ is equal to **G** at infinite frequency: $\mathbf{G}(\infty) = \mathbf{D}$. For simplicity, a matrix **A** in Jordan form is considered for easy ordering of

states so that the eigenvalues are also ordered as $|\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|$

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{5.6}$$

The fastest modes are removed from the model after truncation. The difference between $\mathbf{G}$ and $\mathbf{G}_{trunc}$ following the $k-$order model truncation is given by [101]:

$$\mathbf{G} - \mathbf{G}_{trunc} = \sum_{i=k+1}^{n} \frac{c_i b_i^T}{s - \lambda_i} \tag{5.7}$$

where c_i and b_i are the elments of $\mathbf{C}$ and $\mathbf{B}$ matrices, respectively. Furthermore, it is interesting to note that the error (5.7) depends on the residues $c_i b_i^T$ as well as on λ_i. The distance of λ_i from the imaginary axis is therefore not a reliable indicator of whether the associate mode should be included in the reduced order model or not.

5.2.1.3 Singular Perturbation

Singular perturbation is an established tool for the reduction of models, that displays multiple time-scale behaviors. The aim in applying this technique is to separate distinct type of responses: in fact, a system may have some dynamics that are faster than the dynamics of interest [61]. However fast dynamics can interact with slower dynamics so that truncation of the fast dynamics might not be appropriate. Singular perturbation allows us to take into account the interactions without including only the dynamic effects of the fast dynamics.

The terminology *singular perturbation* comes from the fact that a small parameter ϵ (the so called *perturbation parameter*) scales the derivatives of fast dymanics:

$$\begin{aligned} \dot{\mathbf{x}}_1 &= f_1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}) \\ \epsilon\, \dot{\mathbf{x}}_2 &= f_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}, \epsilon) \end{aligned} \tag{5.8}$$

It is then possible to generate a reduced order model applicable to the time-scale of interest. To apply the method, the model must satisfy the following criteria [90]:

1. Sufficiently differentiable over the range of considered model states.

2. Existance of small physical parameters comparable in magnitude to of ϵ.

3. The time-scale of interest approximatively of $1/\epsilon$ (corresponding to the *time-horizont* of the simulation).

In a linear first-order system, the negative inverse of the eigenvalue λ represents the system time constant, $\tau = 1/\lambda$. Even though the system of interest possess neither of these characteristics, the eigenvalue still provide a good indication of the "*time constant*" of the process.

Singular perturbation approximation is equivalent to model reduction by *residualization* where the derivatives of the fastest states are *residualized*: that is, we simply set $\dot{\mathbf{x}}_2 = 0$ (or equivalently $\epsilon = 0$). This allows to transform the set of $n = k + k_a$ differential equations in 5.2, into a set of k differential equations plus a set of k_a algebraic equations. This way, we obtain the partition of the state equations into two sets of dynamical equations characterized by the variables $\mathbf{x}_1$ and $\mathbf{x}_2$ of dimension $k - k_a$, and k_a respectively, formally:

$$\begin{aligned} \dot{\mathbf{x}}_1 &= f_1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}) \\ 0 &= f_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}, \epsilon) \end{aligned} \tag{5.9}$$

An important property of residualization is that it preserves the steady-state gain of the system $\mathbf{G}_a(0) = \mathbf{G}(0)$. This is not surprising if we think that residualization sets some derivatives to zero, which are zero anyway at steady-state. But this is in contrast with truncation which retains the system behavior at infinite frequency. Truncation is therefore to be preferred when accuracy is required at high frequency, whereas residualization (or generally speaking, singular perturbation) is better for low frequency modelling.

While the validity of performing model simplification via truncation or singular perturbation can be evaluated by the degree to which the decoupling conditions are satisfied and the degree of frequency separation between the deleted dynamics and the desired dynamics, there is no guarantee on the accuracy the resulting simplified model. An advantage of these approaches is that the form of the model (resulting from simplification) is the same as the corresponding portion of the original. Therefore, if the model had a special structure before simplification then that structure is retained in the simplified form. This can be important in allowing the modeller to use his knowledge to interpreter the accuracy of the resulting simplified model as well as the effect of various physical parameters on the system response.

5.2.2 Balanced Reduction

Balanced model reduction for linear system was first introduced by Moore [77] in order to eliminate states that contribute little to the input-output behavior of a system. Whereas modal reduction techniques are based on eigenvalues analysis (and in some sense on stability considerations), balanced reduction methods take into account some other important system properties such as controllability and observability. In other words, modal reduction keeps most of the system behavior intact (i.e., modal truncation preserves the poles and zeros of a system), the reduction methods based upon balancing retain most of the input-output behavior of the system. Balanced reduction methods are explicitly based on asymptotically stable minimal realization at which the controllability ($\mathbf{W}_c$) and observability ($\mathbf{W}_o$) gramians are diagonal and equal. The computation of the gramians matrices (Chapter 3) involves the solutions of a pair of *algebraic Lyapunov equations*[1]:

$$\begin{aligned} \mathbf{W}_c - \mathbf{A}\mathbf{W}_c\mathbf{A}^T - \mathbf{B}\mathbf{B}^T = 0 \\ \mathbf{W}_o - \mathbf{A}^T\mathbf{W}_o\mathbf{A} - \mathbf{C}^T\mathbf{C} = 0 \end{aligned} \tag{5.10}$$

If the solution of the above equations is:

$$\mathbf{W}_c = \mathbf{W}_o = diag(\sigma_1^H, \sigma_2^H, \ldots, \sigma_n^H) \triangleq \Sigma \tag{5.11}$$

where σ_i are the ordered Hankel singular values of $\mathbf{G}(s)$, defined as $\sigma_i^H \triangleq \sqrt{\lambda_i(\mathbf{W}_c\mathbf{W}_o)}$, for $i = 1, \ldots, n$, then the system is called *balanced*.
If the system is stable and controllable then the controllability gramian (5.12) will also have full rank:

$$\mathbf{W}_c \triangleq \int_0^\infty e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T t} dt \tag{5.12}$$

For stable and observable systems the observability gramian (5.13) will have full rank:

$$\mathbf{W}_o \triangleq \int_0^\infty e^{\mathbf{A}^T t}\mathbf{C}^T\mathbf{C} e^{\mathbf{A}t} dt \tag{5.13}$$

If the open loop system is unstable or marginally stable then gramians can not be computed because $\mathbf{W}_c \to \infty$ and $\mathbf{W}_o \to \infty$ [101]. Gramians play an important role in input-output energy considerations and they provide a motivation for some of the aforementioned model reduction methods. In a balanced realization the value of each σ_i^H is a relative measure of the

[1] In Chapter 3 we used those equations in order to define the Hankel Interaction Index Array (HIIA) matrix; of course, they have the same meaning changing only the utilization scope.

contribution that x_i takes into the input-output behavior of the system. Therefore if $\sigma_1^H \gg \sigma_2^H$, then the state x_1 affects the input-output behavior much more than x_2, or any other state because of the ordering of the σ_i. The main idea is that the singular values of the controllability gramian correspond to the amount of energy that has to be put into the system in order to move the corresponding states. For the observability gramian, its singular values refer to the energy that is generated by the corresponding states.
Now, if a nonsingular transformation matrix $\mathbf{T} \in \mathbb{R}^n \times \mathbb{R}^n$ exists, such that we can transform the gramians according to:

$$\begin{aligned} \bar{\mathbf{W}}_c &= \mathbf{T}\mathbf{W}_c\mathbf{T} \\ \bar{\mathbf{W}}_o &= \mathbf{T}^{-1}\mathbf{W}_o\mathbf{T}^{-1} \end{aligned}$$

it can be shown that also a state-space transformation exists:

$$\bar{\mathbf{x}} = \mathbf{T}\mathbf{x} \tag{5.14}$$

The transformed system is given by [37]:

$$\begin{aligned} \dot{\bar{\mathbf{x}}} &= \mathbf{T}\mathbf{A}\mathbf{T}^{-1} + \mathbf{T}\mathbf{B}\mathbf{u} = \bar{\mathbf{A}}\bar{\mathbf{x}} + \bar{\mathbf{B}}\mathbf{u} \\ \mathbf{y} &= \mathbf{C}\mathbf{T}^{-1}\bar{\mathbf{x}} = \bar{\mathbf{C}}\bar{\mathbf{x}} \end{aligned} \tag{5.15}$$

and can be expressed in balanced form using the equivalent notation:

$$\left(\begin{array}{c|c} \mathbf{T}^{-1}\mathbf{A}\mathbf{T} & \mathbf{T}^{-1}\mathbf{B} \\ \hline \mathbf{C}\mathbf{T} & \mathbf{D} \end{array}\right) \tag{5.16}$$

The new system given in equations 5.15, or in 5.16, is then called *balanced realization*. We can say that if the system is in balanced form, the Hankel singular values provide a measure for the importance of the states, because the state with the large singular value is one which is affected the most by control moves and the output is mostly affected by a change of this state. Once the system is in balanced form the state vector can be partitioned into the more important state ($\bar{\mathbf{x}}_1$) and the less important component $\bar{\mathbf{x}}_2$, ($\mathbf{x} = \left[\begin{smallmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \end{smallmatrix}\right]$), in order to eliminate (truncate or residulize) the less important states. The reduced system retains the best possible approximation to the full order system [77] according to such criteria.

To summarize, the balanced model reduction can be achieved following the steps in Figure 5.1 and, depending on the model need, we can define a truncated balanced model (see section 5.2.2.1) or a residualized balanced model (see 5.2.2.2).

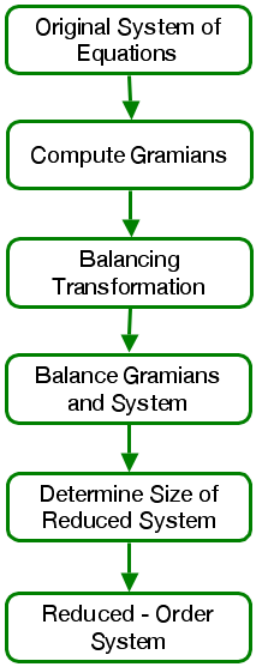


Figure 5.1: Balanced Reduction Scheme

5.2.2.1 Balanced Truncation

As for the modal reduction methods we truncate the states with small eigenvalues, in balanced truncation we eliminate the states that corresponds to small singular values. This reduction method results in a good approximation to the original system over the whole frequency range. The disadvantage of this approach is that it does not preserve the steady-state behavior of the original system and therefore will result in offset [101].
Reduction by truncation leads to the balanced reduced system given as:

$$\begin{aligned} \dot{\bar{\mathbf{x}}}_1 &= \bar{\mathbf{A}}_{11}\bar{\mathbf{x}}_1 + \bar{\mathbf{B}}_1\mathbf{u} \\ \mathbf{x}_2 &= \mathbf{x}_2^{ss} \\ \mathbf{y} &= \bar{\mathbf{C}}_1\bar{\mathbf{x}}_1 \end{aligned} \tag{5.17}$$

The result is a system of ordinary differential equations that contains fewer states than the original one. The number of states that can be truncated depends on the system itself and on the accuracy that is required to system behavior.

5.2.2.2 Balanced Residualization

If it is important to maintain the same steady state behavior for the reduced system as for the full order system, residualization can provide better results than truncation. As for the modal form, residualization is based on the idea that the derivatives of the states corresponding to the small Hankel singular values can be approximated to zero while the rest of the system is retained. That is, if in the modal form we discard the fastest states, we discard the less controllable and observable states.

Both balanced truncation and residualization methods are related to the corresponding modal methods, but whereas modal reduction keeps most of the system behavior intact, the reduction methods based upon balancing retain most of the input-output behavior of the system. Therefore, modal reduction is optimal for uncontrolled systems because it is based upon dynamic behavior that the states of the system exhibit. On the other hand modal reduction does not consider the influence of the input-to-state (controllability) or the state-to-output (observability) behavior. Balanced reduction methods are more suitable than modal reduction for system that will be used for controller design because they preserve the input-output behavior.

5.3 Nonlinear Models

So far we have shown how a linear model can be reduced and we have also seen that balancing for linear systems is a powerful technique that is simple to implement. But, especially when model-based control methods such nonlinear model predictive control are needed, a nonlinear model can provide a more accurate description of the process dynamics. However, nonlinear controllers have some drawbacks when compared to linear controllers due to the increased complexity introduced by the nonlinearity of the model. For that reason it is interesting to report in this section some systematic techniques to achieve a nonlinear model reduction. Anyway, the implementation of nonlinear controller is not part of this thesis.

To extend the model reduction approaches discussed in the previous section to nonlinear systems several new challenges arise, such as numerical problems and a lack of a suitable theory [37].

5.3.1 Balancing for nonlinear systems

In the nonlinear case a general balancing scheme is not available. One possibility is to linearize the system and apply the methods described in Section

5.2. However nonlinear behavior will be lost and some results for a specific class of nonlinear systems do not exist. Lall *et al.* [64], and also Hahn *et al.* ([37], [38]) proposed a systematic procedure for the nonlinear model reduction: the proposed approach led to an efficient algorithm based on the covariance matrices for the input-to-state (controllability) and state-to output (observability) behavior of the nonlinear system. In order to go into more details a few terms need to be defined.

First of all, it should be noted that, since no theory for general nonlinear balancing exists, control-affine systems are the only type of systems that can be balanced[2]. For such a system, the *controllability energy function*:

$$\mathbf{L}_c = \min_{(\mathbf{u}\in L_2(\infty,0),\, \mathbf{x}(-\infty)=0,\, \mathbf{x}(0)=x_0)} \frac{1}{2}\int_{-\infty}^{0} \|\mathbf{u}(t)\|^2 dt$$

and also the *observability energy function*:

$$\mathbf{L}_O = \frac{1}{2}\int_{0}^{\infty} \|\mathbf{y}(t)\|^2 dt$$

where $x(0) = x_o$, $u(t) \equiv 0$, $0 \leq t < \infty$, are defined. The observability and controllability energy functions are related to the gramians and their relation is given by the following equations:

$$\mathbf{L}_c(\mathbf{x}_0) = \frac{1}{2}\mathbf{x}_0^T \mathbf{W}_c^{-1} \mathbf{x}_0$$
$$\mathbf{L}_o(\mathbf{x}_0) = \frac{1}{2}\mathbf{x}_0^T \mathbf{W}_o \mathbf{x}_0$$

Unfortunately, it is not easy to calculate the energy functions in the nonlinear case. Whether in the linear case the Gramians can be defined by solving the Lyapunov equations in (5.10), for the non linear case the resolution is more complicated [93]. For this reason we report the method presented in [37] where **empirical gramians** are calculated from process data. The gramians are then balanced by the same procedure used for linear systems. The balancing transformation is used within a Galerkin projection in order to transform the nonlinear system into the balanced form. The resulting nonlinear equations can be reduced using different truncation or residualization methods.

[2] A system that is not control affine, i.e., which is linear in the actions but nonlinear with respect to the states, must be linearized with regard to the control input, in order to balance it.

5.3.2 Empirical Gramians and Model Reduction

Empirical Gramians were proposed by Lall *et al.* [63]. They are restricted to stable (in the sense of Lyapunov) nonlinear control-affine systems. For the nonlinear system defined in (5.1) the following sets need to be defined:

$$\begin{aligned} T^m &= \{T_1, \ldots, T_r;\ T_i \in \mathbb{R}^m \times \mathbb{R}^m,\ T_i^T T_i = I,\ i = 1, \ldots, r\} \\ M &= \{c_1, \ldots, c_s;\ c_i \in \mathbb{R},\ c_i > 0,\ i = 1, \ldots, s\} \\ E^m &= \{e_1, \ldots, e_m;\ \text{standard unit vectors in } \mathbb{R}^m\} \end{aligned}$$

where:

- r, is the number of matrices for excitation directions
- s, is the number of different excitation sizes for each direction, and
- m, is the number of inputs to the system

The *controllability covariance matrix* is defined as:

$$\mathbf{W}_{c,emp} = \sum_{i=1}^{m} \sum_{j=1}^{r} \sum_{k=1}^{s} \frac{1}{rsc_k^2} \int_0^\infty \Phi^{ijk}(t)dt \tag{5.18}$$

where $\Phi^{ijk}(t) \in \mathbb{R}^m \times \mathbb{R}^m$ is given by $\Phi^{ijk}(t) = [x^{ijk}(t) - x_0^{ijk}][x^{ijk}(t) - x_0^{ijk}]^T$ and $x^{ijk}(t)$ is the state of the nonlinear system corresponding to the impulse input $u(t) = c_k T_j e_i v t + u_0$ (c_k describes the inputs size, $T_j e_i$ determines the input direction, $v(t)$ denotes the nature of the input and u_0 refers to the input at the original steady state). The quantities x_0^{ijk} refer to the steady state of the system. If $v(t)$ is piecewise constant as in the case of impulse and step inputs then $x_0^{ijk}(t)$ will also be piecewise constant. The nature of the input should be chosen in such a way that it is consistent with the typical input behaviour of the plant.

The *empirical observability gramian* is defined as:

$$\mathbf{W}_{o,emp} = \sum_{j=1}^{r} \sum_{k=1}^{s} \frac{1}{rsc_k^2} \int_0^\infty T_j \Psi^{jk}(t) T_j^T dt \tag{5.19}$$

where $\Psi^{jk}(t) \in \mathbb{R}^m \times \mathbb{R}^m$ is given by $\Psi_{pq}^{jk} = [y^{ijk}(t) - y_0^{ijk}][y^{ijk}(t) - y_0^{ijk}]^T$, and $y^{ijk}(t)$ is the output of the nonlinear system corresponding to the initial condition $x(0) = c_k T_j e_i + x_0$. The y_o^{ijk} refers to the output measurement corresponding to the steady state of the system.

In [63], it is shown that both of these gramians reduce to linear gramians for linear models. Furthermore, the empirical gramians will reduce to gramians of the linearized system for small perturbations around operating point. The empirical gramians have to be determined from experimental or simulation data, collected within the region where the process is to be controlled. The empirical gramians matrices capture part of the nonlinear behaviour within the region of operation.
At this point the empirical gramians can be balanced by the methods for linear model reduction, and from the balanced gramians it can be concluded which states contribute the most to the input-output behavior of the system. States that are either unobservable or uncontrollable do not influence the input-output behavior and can be eliminated. Once the reduced number of states has been decided a Galerkin projection is performed to produce the reduced-order nonlinear system of the form:

$$\begin{aligned} \dot{\bar{\mathbf{x}}}_1 &= \mathbf{P}\mathbf{T} f(\mathbf{T}^{-1}\bar{\mathbf{x}}(t), \mathbf{u}(t)) = \mathbf{P}\bar{f}(\bar{\mathbf{x}}(\mathbf{t}), \mathbf{u}(t)) \\ \bar{\mathbf{x}}_{\mathbf{2}} &= \bar{\mathbf{x}}_2^{ss} \\ \mathbf{y} &= h(\mathbf{T}^{-1}, \bar{\mathbf{x}}) = \bar{h}(\bar{\mathbf{x}}(t)) \end{aligned} \tag{5.20}$$

where $\mathbf{P} = [\mathbf{I}\ \mathbf{0}]$ is a square matrix of full rank and its dimension is equal to the number of states of the system.

Chapter 6

ASP Reduced Order Models

Dynamic simulations based on rigorous and detailed modelling have become a standard tool in many engineering fields. Rigorous models are applied for a variety of tasks: they allow the exploration of the impact of changing some design configurations, can be used to provide tool to actively explore new ideas and improve the learning process as well as allowing the operators training facilities and thereby increasing their ability to handle unforeseen situations. In other words, they are useful in understanding the system behavior.

As seen earlier, the ASM1 is a widely used model for simulation of the activated sludge process. The model includes 13 different components and the dynamic behavior of each component is described by a non linear differential equation (see the model description in Chapter 2 and [43]). Although the ASM1 model comprises much of the knowledge of the biological reactions (when nitrogen and carbonaceous pollutants are considered) a number of drawbacks exist [52]: lack of identifiability, awkward nonlinearities, difficult estimation and updating of time varying parameters. So, despite the usefulness of a rigorous model, there are some situations where simpler models are better suited. In process control, for example, the high dimensionality of large models results in enormous computational requirements and ill-conditioned problems due to interaction of slow and fast dynamics. In model identification, rigorous models typically require high investments in model tuning and validation, that can exhibit lack of parameter identifiability. In addition, detailed models can contain internal states whose behavior is difficult to verify so, much understanding can be acquired from a reduced model describing only the relevant phenomena. A simpler model can be useful to construct a *soft-sensor* which is a powerful tool to on-line estimate of unmeasured states.

The goal of this Chapter is to construct a simple reduced model for the activated sludge process, utilizing the systematic procedures exposed in Chapter 5. In Section 6.1, a brief overview of available reduced models is given and simulation results of two of them are developed and discussed along with our proposed model. The models represent the basis upon which we design the soft-sensors (Chapter 7).

6.1 Brief Overview of ASP Reduced Models

Different applications involves different requirement in the reduced model. For this reason many reduced models for the activated sludge process are present in literature. In this section a brief overview of some of the available models is presented, and for each case we discuss the treatment goal for model reduction. The motivation for the selected approach, together with the reduced models is also indicated. In most of the cases presented, the purpose of model reduction is an application to identification and control.

As showed in Chapter 5, two main approaches to reduce complex models exist: one is based on the modeller physical knowledge about the process and the other one is based on using systematic procedures. In Section 6.1.1 some reduced model are reported and in Section 6.1.2 some models deduced by means of systematic techniques are also presented.

6.1.1 Physical Knowledge Reduction Approach

The model reported in this section is essentially a simplification of the biological process either having as a starting point the ASM1 model or generally deduced from the authors knowledge about the process. The main aim of the brief review is to show that different models for the biological reaction can be achieved and the simplification level depends basically on how the model will be used.

For example, *Isaacs* [48] tested three model-based control strategies on a BioDeniPho[1] system. All controllers employ a relational model and a predictive model and different models were applied for different control strategies. They are based on mass balances over ammonia and nitrite, and volumetric rate for nitrification are adaptively estimated directly from process measurements of ammonia and nitrate concentration and linked to the control

[1]The BioDeniPho is an alternating sequential semi-batch system with nitrogen and phosphorus removal which is obtained by switching the flow and aeration pattern according to a cycling strategy.

variables through experimentally determined empirical correlations, whose parameters values are also adaptively estimated.

In 1989, *Marsili-Libelli* [73] developed a low order model for conventional activated sludge systems with BOD removal and nitrification. His motivation was that literature models are not suited for control, due to their complexity and poor identifiability. The model was developed to describe biodegradation of carbonaceous COD, nitrification, DO utilization, and sludge sedimentation. As for the bacterial growth, a predator/prey modified Volterra approach was used instead of the usually applied Modod model.

For *Cartersen* et al. [14] the problem to be solved was an identification problem. They formulated simplified models capable to give on-line information on the present state of the wastewater treatment plant. This was obtained by a built-in adaptivity of the models such that the model is updated for each new time step (when new information of the available on-line measurements are available). The resulting model is a grey box model, where the Monod-kinetic parameters of the nitrification and denitrification process can actually be identified and estimated by means of prediction error decomposition and maximum likelihood estimation. They showed for the BioDeniPho system that it is possible to find a unique solution for each of the unknown parameters of the model, starting from data collected in a real process.

The most straightforward model complexity reduction is obtained by assuming only oxic conditions and thus neglecting the denitrification processes. *Kabouris* and *Georgakakos* [58] investigated the application of an optimal control method to a reduced form of the ASM1 model obtained in such a way. The dissolved oxygen concentration is assumed to be fairly high enabling the decoupling of the oxygen dynamics from the rest of the process dynamics. Only five ASM1 state variables are retained: X_S, S_S, X_P, X_I and X_{BH} (see Chapter 2 for notation). Furthermore, only aerobic growth, decay and hydrolisis of particulate to soluble substrate are taken into account.

In 1995, *Zhao et al.* [119] introduced a simplified ASM1 model to describe the nitrogen dynamics in an alternating activated sludge process. The proposed model is represented by a set of dynamics equations in terms of ammonia and nitrate concentrations (on-line measurable).

Along the same line *Julien et al.* (in [57] and in [56]) developed a reduced-order model for identification and control of a single activated-sludge reactor operating nitrification and denitrification. The reduced model is split into

two submodels, one three-dimensional state submodel in aerobic conditions (considering as state variables nitrate, ammonia and dissolved oxygen) and one two-dimensional submodel in anoxic conditions, where only nitrate and ammonia state variables are considered. The identifiability is based on online oxygen and nitrate concentrations data, showing that the reduced model is structurally identifiable.

Starting from the assumption made in [57], **Gomez-Quintero *et al.*** [34] developed further modifications to the ASM1 model, hereafter referred as to the GQ *et al.* model. **Jeppsson and Olsson** (JOM) [55] reduced the complexity of the ASM1 on physical grounds. Since we use both models in soft-sensors design, the former reduced model has been modified in order to asses the observability. Anyway, we will discuss in finer detail these two models in the following Sections 6.2 and 6.3, respectively.

6.1.2 Systematic Reduction Approach

In Chapter 5, a different approach to achieve model order reduction has been analyzed. A frequently encountered strategy in the ASM1 model reduction is the *singular perturbation theory* (Section 5.2.1.3).

We start this overview from a general biological models with *Bastin and Dochain*. In 1990, they proposed a general rule for order reduction in their book on estimation and control of bioreactors [9]. Given the balance for a generic component ξ_i:

$$\dot{\xi}_i = \sum_{j \sim i} (\pm) k_{ij} \psi_j - D\xi_i - Q_i + F_i(^2)$$

the simplification is then achieved by setting ξ_i and $\dot{\xi}_i$ to zero:

$$\sum_{j \sim i} (\pm) k_{ij} \psi_j = Q_i - F_i$$

However, the rule is not general for several reasons [115]. Actually, it is not indicated in a general sense in which cases the dynamics of a component can be neglected. The general rule is only motivated in two specific situations: neglegible product dynamics for volatile products with low solubility and neglegible substrate dynamics in a model with biomass and substrate.

[2]The notation $i \sim j$ means that the summation is taken on the reactions with index j involving the component with index i. The terms k_{ij} are stoichiometric coefficients, ψ_i the reaction rates, D the dilution rate, F_i is the mass feed rate of the component ξ_i in the reactor and Q_i is the rate of mass outflow of such component.

Indeed, these example are not sufficient to explain all the possible situations with multiple time-scales.

A more general approach is taken in [111], where *Van Breusegem and Bastin* obtained the reduction of a reaction network with the assumption that some reactions are much faster that others. The reaction network of a reaction system describes the relations between components and reactions:

$$\sum_{i=1}^{n} \gamma_{ij} X_i \rightarrow \sum_{i=1}^{n} \delta_{ij} X_i$$

where the nonnegative real numbers γ_{ij} and δ_{ij} are the stoichiometric coefficients. They express the nominal quantity of the component X_i ($i = 1, \ldots, n$), which is consumed (δ_{ij}) or produced (γ_{ij}) by the j^{th} ($j = 1, \ldots, m$) reaction.

A systematic approach for reducing complex biological processes was proposed by *Steffens* et al. [105]. They applied the *eigenvalue to state association* proposed by Kokotovic *et al.* [61] to a carbon removal and nitrifying ASP model. As explained in Chapter 5, it is possible this way to quantify the "speed" of a state. This information is then used to reduce the model via singular perturbation analysis.

Keesman et al. [60] analyzed the endogenous phase in an aerobic biodegradation process with no external substate addition. The ASM1 is then reduced to the mass balance equations for S_S, X_S, X_{BH}, X_P where the former two are classified as fast states and the latter two as slow states via singular perturbation analysis. The separation improves the efficiency in computer simulations. Furthermore, it is shown that exclusion of S_S from the state vector significately increases practical observability.

Smeths et al. [103] proposed to reduce the complexity of the ASM1 (through linearization) in order to reduce the computational time. Their aim was also to provide a valuable tool in a risk assessment environment as well as in online MPC control strategies. The complexity reduction consisted of four steps:

1. Construction of representative input/output datasets generated by simulating the full ASM1 model;

2. The ASM1 model is rewritten in a state-space format with linear approximations of the (nonlinear) kinetic terms;

3. Identification of the unknown parameters in the linear terms on the basis of the generated input/output data;

4. Introduction of a multi-model identification procedure.

Linearizing the model around a time-varying reference trajectory (to reflect variations in load) results in a linear but time-varying model. The Authors applied the above reduction procedure to an activated sludge process model consisting of a denitrification tank followed by a nitrification tank and a settler.

In *Lee et al.* [67] it is noted that although design of controllers for biological processes is one of the main objectives in model-order reduction, the proposed models have not always addressed the closed-loop robustness issue. For that reason, they proposed a reduced-order linear state-space model for the nonlinear differential equation model of an activated sludge process.The method relies on an investigation of five different reduction techniques (direct truncating, frequency-weighted balanced realization, Schur balanced truncation, singular perturbation approximation and optimal Hankel norm approximation). They concluded that, on the analysis of the resulting frequency errors plots, the reduced model generated from singular perturbation approach gives the lowest errors in low frequency ranges and hence is deemed most suitable for controller design.

6.2 *GQ et al.* Model

This model was presented in 2000 by Gomez-Quintero *et al.* [34]: the strategy for reducing this model involves biochemical considerations. The model is based on some of the assumptions made by Julien *et al.* in [57]: the alkalinity concentration is omitted, the X_s state variable is included in the definition of X_I and a new parameter K_{NH_4H} is introduced to take into account the possible ammonia limitation for the aerobic and anoxic growth of heterotrophic biomass. Furthermore, some other assumptions and considerations were made:

- the term $\frac{S_s}{S_s+K_s}$ is replaced by a linear expression γS_s (where γ is scalar). This assumes that $S_s \ll K_s$ is low with respect to K_s.

- both variables X_s and X_{BH} evolve widely when operating conditions vary over a long time period so the rate value may change significantly and be fairly far from unity. This means that the term $\frac{X_s/X_{BH}}{X_s/X_{BH}+K_X}$ is substituted by a new parameter k_1 specific to the reduced-order model.

- the term $\frac{S_{NH}}{S_{NH}+K_{NH_4H}}$ varies between 0.95 and 0.99, when S_{NH} evolves from $0.95\,g/m^3$ to $10\,g/m^3$, with $K_{NH_4H} = 0.05\,g/m^3$. Hence, it is replaced by a constant K_2 that preserves he physical meaning of the heterotrophs growth rate μ_H as in the ASM1 model.

- the inflow nitrate concentration S_{NO}^{in} is neglected.

- during one nitrification-denitrification cycle, X_{BH}, X_{BA}, S_{ND} do not vary so much. They are assumed to be constant terms corresponding to their mean values.

The final reduced model is described by 4 state variables:

$$\mathbf{x} = [x_1, \ldots, x_4]^T = [S_{NO},\, S_O,\, S_{NH},\, S_S]^T \tag{6.1}$$

where S_{NO}, S_O, S_{NH}, S_S have the same meaning as in the ASM1 model. Also the number of parameter is reduced with respect to the ASM1 model. The following supplementary approximations were done before grouping the parameters:

- η_g has been approximated to the constant value 1;

- the sum $i_{BH} + \frac{1}{Y_A}$ has been approximate to $\frac{1}{Y_A}$ because $i_{BH} \ll \frac{1}{Y_A}$;

- the term $\frac{4.57-Y_A}{Y_A}$ has been replaced by $\frac{4.57}{Y_A}$ since $\frac{4.57}{Y_A} \gg 1$.

Furthermore, the reduced model involves 12 parameters:

$$\begin{aligned}\mathbf{p} &= [p_1, \ldots, p_{12}]^T \\ &= [Y_H,\, i_{XB},\, K_{OH},\, K_{OA},\, K_{NH},\, \eta_g,\, \eta_h,\, \alpha_1,\, \alpha_2,\, \alpha_3,\, \alpha_4]^T\end{aligned} \tag{6.2}$$

where α_1, α_2, α_3, α_4 are some "new" model parameters, which do not have a real physical meaning but are obtained grouping the ASM1 parameters according to the assumptions made above.
Considering the following *switching functions*, the reaction rates of the reduced model were given by equations in 6.3:

$$\begin{aligned} s_1 &= \frac{S_{NO}}{S_{NO}+K_{NO}}, & s_2 &= \frac{K_{OH}}{S_O+K_{OH}}, \\ s_3 &= \frac{S_{NH}}{S_{NH}+K_{NH}}, & s_4 &= \frac{S_O}{S_O+K_{OA}}, \\ s_5 &= \frac{S_O}{S_O+K_{NO}}, \end{aligned}$$

$$
\begin{aligned}
r_1 &= -Ax_4 s_1 s_2 + \alpha_2 s_3 s_4 \\
r_2 &= K_{La}(S_{OST} - x_2) - Ex_4 s_3 \\
r_3 &= -Bx_4(s_3 + s_1 s_2) - \alpha_2 s_3 s_4 + \alpha_3 \\
r_4 &= (\alpha_4 - Dx_4)s_5 + (C - Dx_4)s_1 s_2
\end{aligned}
\tag{6.3}
$$

where,

$$
\begin{aligned}
A &= \frac{\alpha_1(1 - Y_H)}{2.86 Y_H}, \quad & B &= \alpha_1 i_{XB}, \\
C &= \alpha_4 \eta_h & D &= \frac{\alpha_1}{Y_H}, \\
E &= \frac{\alpha_1(1 - Y_H)}{Y_H}, \quad & F &= 4.57\alpha_2
\end{aligned}
$$

6.2.1 Test Motion

Since we want to represent the whole bioreactor, with denitrification and nitrification processes, the GQ *et al.* model has been used for both of them. Furthermore, since the state variables in the reduced model are only the soluble ones, there is no need to represent the secondary settler for this system. The model has been tested upon the data provided from the considered wastewater treatment plant: flow rate, dissolved oxygen concentration in the basis, daily COD and nitrate and ammonia concentration every two or three days. Those data correspond to the exogenous inputs to the model:

$$
\mathbf{d} = [S_{NO}^{in},\, S_{NH}^{in}, S_S^{in},\, Q_{in}/V,\, Q_r/V,\, Q_w/V]^T \tag{6.4}
$$

Furthermore, from the full complex model simulations[3] the reduced model parameters were obtained. In Table 6.1 and in Table 6.2 the parameters for both anoxic and aerated zone are reported.

Anoxic α	Aerobic α
$\alpha_1^p = 4.845\, d^{-1}$	$\alpha_1^n = 163.9\, d^{-1}$
$\alpha_2^p = 152.098\, gm^{-3}d^{-1}$	$\alpha_2^n = 224.630\, gm^{-3}d^{-1}$
$\alpha_3^p = 42.909\, gm^{-3}d^{-1}$	$\alpha_3^n = 92.120\, gm^{-3}d^{-1}$
$\alpha_4^p = 340.241\, gm^{-3}d^{-1}$	$\alpha_4^n = 739.740\, gm^{-3}d^{-1}$

Table 6.1: Composite parameters in the GQ *et al.* reduced model

In Figure 6.1, the test motion for the reduced model (dashed line) is compared with the full model (solid line). As we can note, the reduced model

[3]The reference model in this case is the *twostepmantis* model in the GPS-X environment that is exposed in Chapter 2.

$\eta_h = 0.4$	$\eta_g = 0.5$
$K_{OA} = 0.23\, gO_2m^{-3}$	$K_{OH} = 0.2\, gO_2m^{-3}$
$K_{NH} = 0.2\, gNm^{-3}$	$K_{NO} = 0.1\, gNm^{-3}$
$i_{XB} = 0.086\, gNgCOD^{-1}$	$Y_H = 0.7$

Table 6.2: Kinetic and stoichiometric parameters in the GQ *et al.* reduced model

gives the motion tendency but with significant offsets due to the errors in the reduced model assumption and parameter identification. From these results follows that, using the reduced model, the soft-sensor gives a better inference of the modelling errors and reaches the actual concentration motion.

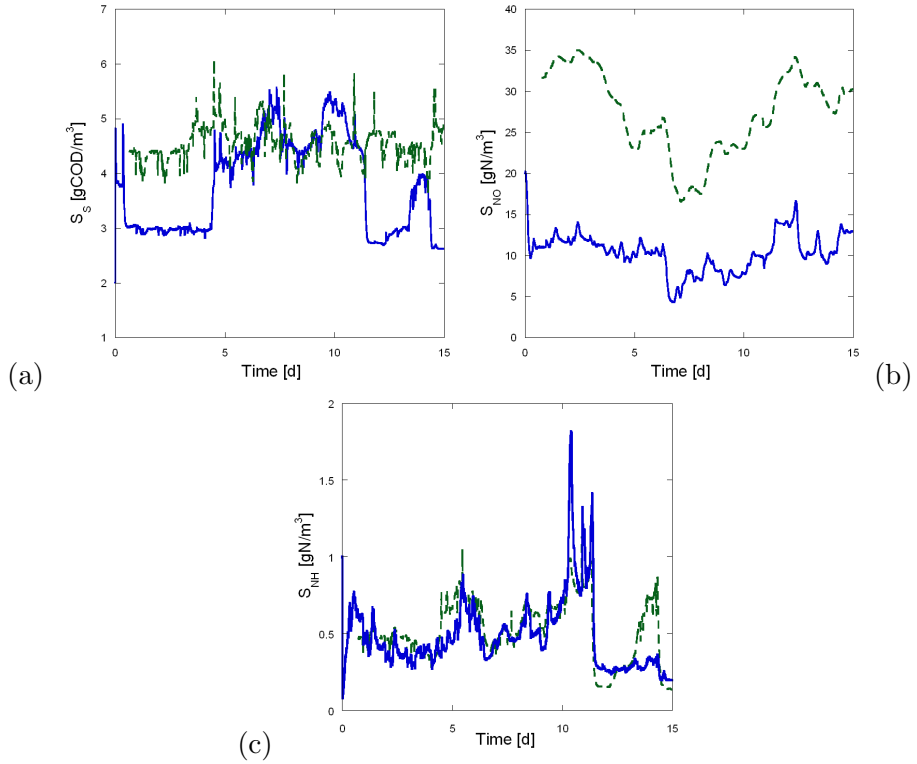


Figure 6.1: Test motion of the Gomez-Quintero *et al.* reduced model [Full model (solid); GQ *et al.* (dashed)]

6.3 JOM Model

One of the major limit encountered using the GQ *et al.* model was that the organic matter is represented only in terms of S_S (the readily biodegradable substrate) which is actually not directly measurable on-line. So, there was the need to find a model to represent the organic matter with some "real" measurements. We chose the reduced model proposed by Jeppsson and Olsson ([51], [55]), which includes the COD concentration as state variable. This model represents one of the first attempt to reduce the ASM1 on the basis of qualitative assumptions about the biological processes.

The four fraction of organic matter (soluble inert organic matter S_I, readily biodegradable substrate S_S, particulate and inert organic matter X_P and X_I, and slowly biodegradable substrate X_S) are replaced by a single variable X_{COD} which is considered directly measurable and comparable with the experimental data. The two types of heterotrophic and autotrophic microorganisms (X_{BH}, X_{BA}) described in the ASM1 are maintained. Heterotrophs are considered to growth in both anoxic and aerobic environments, whereas autotrophs grow only in an aerobic environment. The only two nitrogen fractions included in the reduced order model are the nitrate nitrogen and the ammonia nitrogen (S_{NH}, S_{NO}). Summarizing, the JOM model includes 5 state variables:

$$\mathbf{x} = [X_{COD},\, X_{BH},\, X_{BA},\, S_{NO},\, S_{NH}]^T$$

consisting of 8 parameters:

$$\mathbf{p} = [r_H,\, r_A,\, Y_H,\, Y_A,\, b_H,\, b_A,\, i_{XB}]^T$$

where $r_H\,[m^3\,(gd)^{-1}]$ and $r_A\,[m^3\,(gd)^{-1}]$ are the reaction rate factors for heterotrophs and autotrophs, respectively. The other parameters have the same meaning than in the ASM1 model.

Some other assumptions have been made about the reaction rates. In fact, in the ASM1 model the hydrolysis of entrapped organic nitrogen is described. This reaction continues via the ammonification process to form ammonia nitrogen. In the reduced model, ammonia nitrogen is assumed measured and its formation mechanism is not considered. The hydrolysis of entrapped organics in the ASM1 is simplified. Therefore, the decay material (heterotrophs and autotrophs) is formed into organic substrate and ammonia directly. Another reason for this simplification is that the true nature of the hydrolysis mechanism is not well known. Hence, from the eight processes represented in the ASM1 model only four remain in the reduced model.

6.3.1 The modified JOM

In the JOM, the oxygen concentration is excluded from the state variables set since it is assumed that the dissolved oxygen (DO) is controlled and the corresponding growth expressions become independent of DO variation.
Bearing in mind that our goal is to design a soft-sensor based on this reduced model with on-line measurements of dissolved oxygen, in order not to loose the system observability we can not assume a constant DO. For that reason, we modified the JOM including the oxygen mass balance (hereafter referred as to the $JOMO_2$ model). This means that we re-include also the ASM1 switching functions (in order to discriminate between the anoxic and aerobic environments) and we also consider the same model for both reactor zones instead of the two different models in the JOM. Including the ASM1 switching function we include also some nonlinearities in the JOM model. The modified JOM model now consists of six state variables:

$$\mathbf{x} = [X_{COD},\, X_{BH},\, X_{BA},\, S_O,\, S_{NO},\, S_{NH}]^T \tag{6.5}$$

Furthermore, in order to calculate the oxygen mass transfer coefficient, K_{La}, another equation was introduced in each of the models of the bioreactor. Adopting a Petersen matrix notation, the $JOMO_2$ model is summarized in Table 6.3.

For the sake of completness, we report that the JOM model was also modified in another work by Ayesa *et al.* [6] to evaluate its observability for two separate WWT plant configurations. Incorporating to the JOM model the particulate inert organic matter X_I allowed to model the suspended solids in the plant. Their objective was again to develop and apply advanced control strategies to the plants.

6.3.2 Test Motion

The reduced $JOMO_2$ model simulations was validated with respect to the full complex model[4]. The simulations were carried out reproducing a real situation for a horizon time of fifteen days, as for the Gomez-Quintero *et al.* in Section 6.2.1. Also in this case, the reduced model parameters were obtained on the basis of the full complex model simulation. In Table 6.4 the model parameters are reported. It should be noted that since they have, in some way, lost the original ASM1 meaning they are abeled with the "$*$" superscript.
In Figure 6.2 the bioreactor effluent concentrations for the reduced model are compared with the complex model. We note that a relatively good fitting

[4] As for the GO *et al.* the reference model is the *twostepmantis* model

Component → ↓ Process	1 X_{COD}	2 X_{BH}	3 X_{BA}	4 S_O	5 S_{NO}	6 S_{NH}	Reaction
Aerobic growth of het.	$-\frac{1}{Y_H^*}$	1		$-\frac{1-Y_H^*}{Y_H^*}$		$-i_{XB}$	$\mu_H^* X_{COD} X_{BH} \frac{S_O}{S_O+K_{OH}^*}$
Anoxic growth of het.	$-\frac{1}{Y_H^*}$	1			$-\frac{1-Y_H^*}{2.86Y_H}$	$-i_{XB}$	$\alpha\mu_H^* X_{COD} X_{BH} \frac{K_{OH}^*}{S_O+K_{OH}^*}$
Aerobic growth of aut.			1	$-\frac{4.57-Y_A^*}{Y_A^*}$	$\frac{1}{Y_A^*}$	$-i_{XB}-\frac{1}{Y_A^*}$	$\mu_A^* S_{NH} X_{BA} \frac{S_O}{S_O+K_{OA}^*}$
Decay of het.	1	−1					$b_H^* X_{BH}$
Decay of aut.	1		−1				$b_A^* X_{BA}$

Table 6.3: JOM with Dissolved Oxygen included

μ_H^*	=	0.005	d^{-1}	μ_A^*	=	0.177	d^{-1}
b_H^*	=	2	d^{-1}	b_A^*	=	0.06	d^{-1}
K_{OA}^*	=	0.05	gO_2m^{-3}	K_{OH}^*	=	0.4	gO_2m^{-3}
i_{XB}	=	0.086	$gNgCOD^{-1}$	α	=	0.8	m^3gN^{-1}
Y_H^*	=	0.2		Y_A^*	=	0.1	

Table 6.4: Kinetic and stoichiometric parameters in the $JOMO_2$ reduced model

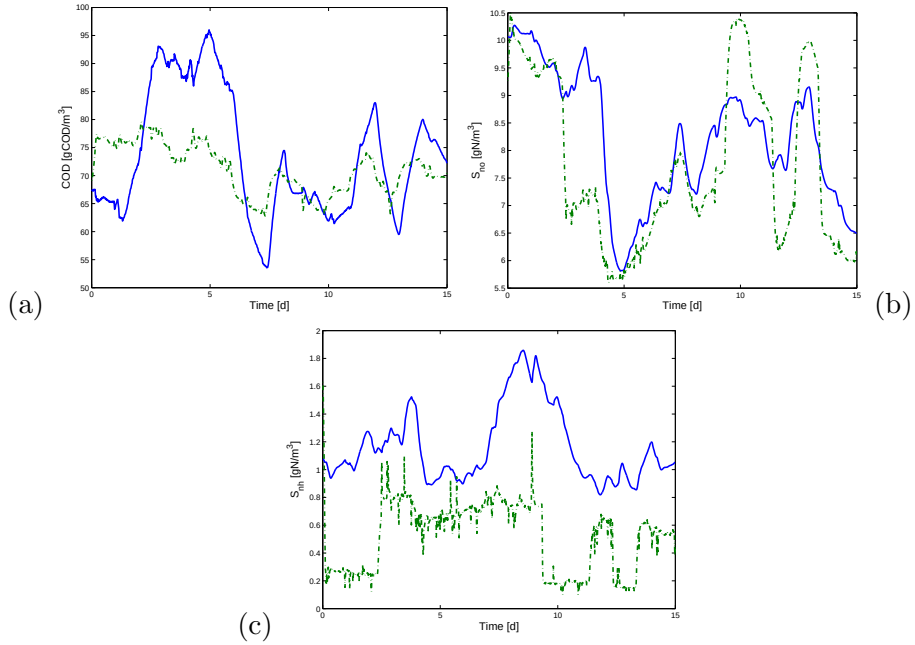


Figure 6.2: JOM_2 model test motion [Full model (solid); $JOMO_2$ (dashed)]

exists for nitrate concentration (Figure 6.2b), whereas the ammonia gives a good motion tendency (Figure 6.2c) but with a significant offset. Both S_{NO} and S_{NH} are soluble components: their behavior is not influenced by the presence of the secondary settler. It should be noted that the aim of the JOM_2 model reduction focuses on the bioreactor, whereas the secondary settler was actually not considered. In fact, the settling process was taken into account only by introducing a percentage of solids removal equal to 80%. This explains the COD behavior for the reduced model in Figure 6.2a that does not fit the effluent concentration of the full complex model (i.e. bioreactor + settler).

6.4 Proposed Low Order Model

Up to now, we have investigated the system behavior with low order models obtained with physical knowledge reduction. The GQ *et al.* model showed that the process identifiability might be difficult if the parameters lose their physical meaning: that is, the reduced model needs a recalibration procedure. Also, we noted that a model based upon "measurable" states for the organic pollutants (instead of S_S) can gain importance in practical applications. Mainly for this reason, we developed the $JOMO_2$ focusing only in the bioreactor model. Now we need a model which is able to represent the whole activated sludge process, i.e., a bioreactor and a secondary settling tank. This model has to be compared directly with the ASM1 model and has to be based upon measurable variables. For all these reasons, the purpose of this section is to develop simple dynamical models for the activated sludge system suitable for operation and control: they have to be easy to handle and/or take into consideration important properties of observability and controllability. The resulting model can be sufficient to describe major phenomena but still limited to be handled in a quite simple way. The systematic techniques exposed in Chapter 5 represent the theoretical framework for our reduction method.

We saw that the majority of model reduction techniques have been developed for linear models but, we know also that the ASM1 model is a nonlinear one. For that reason, the first step consists of linearizing the full model. Considering the ASM1 model implemented in Simulink (Chapter 2), the linearization task has been performed using Matlab. The stationary operating points were found firstly with a trim analysis and then from simulations. The obtained linear model is represented in the standard state space format as reported in 5.2. It is completely described by the 4-tuple (**A**, **B**, **C**, **D**).

6.4.1 Modal Model

The modal reduction approaches rely on the eigenvalue to state association. From the linearized model, the eigenvalue to state association was performed (Section 6.4.1.1) in order to truncate or residualize the model. Figure 6.3 and in Section 6.4.1.2 illustrate the adopted methodology.

6.4.1.1 Eigenvalue to State Association

The eigeinvalue to state association is defined with the homotopy method: this means that a system with a known correspondance between state and eigenvalues must be constructed. We start considering the diagonalized $\mathbf{A}^p$

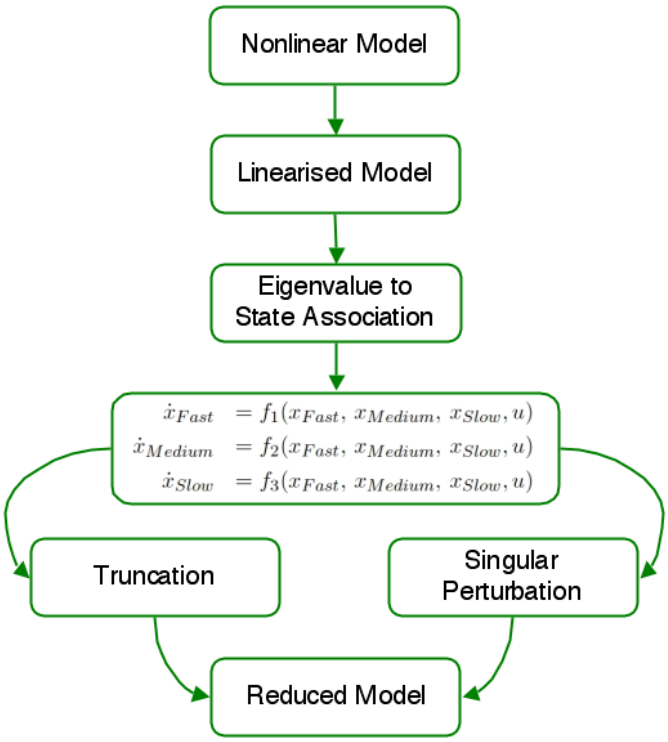


Figure 6.3: Systematic approach for modal reduction

and $\mathbf{A}^n$ matrices for the anoxic and aerated zone, respectively. In such a way each diagonal element (i.e., each eigenvalue) is directly related to the state. Then the system is transformed using the homotopy parameter $r \in [0, 1]$ into the actual system while tracing the eigenvalues. We considered different steady state points and we observed similar system behaviors for all of them. The results for one of them are reported; in particular, in Figure 6.4 the homotopy function for the anoxic zone components is shown and in Figure 6.5 the results for the aerated zone ones are illustrated (for a easier visualization the eigenvalues are reported in log scale). Although only the real part of the eigenvalue is plotted, in these figures every variable has a zero imaginary component. This aspect denotes a smooth (non-oscillatory) response to transient. In both situations, the state variables are classified on a time basis in three groups:

- *fast states*, with time constants of seconds;
- *medium states*, with time constants of minutes;
- *slow states*, with time constants of hours.

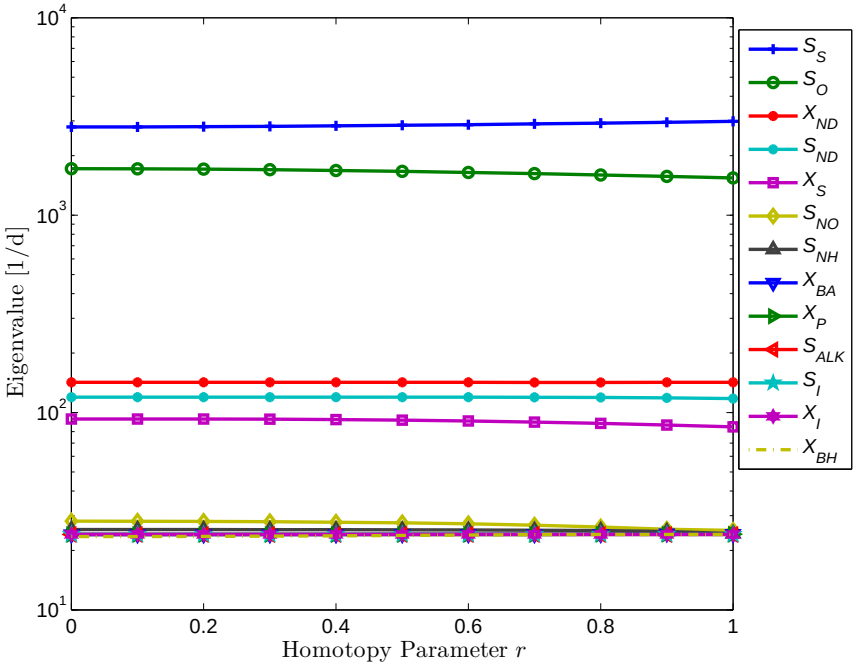


Figure 6.4: Eigenvalue traces for the ASM1 model variables in anoxic zone

From Figure 6.4 and 6.5, we notice that the system does not behave the same way in both zones. In order to decide which states are fast, medium or slow, a critical region in the eigenvalue space is used to define the region of "reducible" or "irreducible" states. Any state whose associated eigenvalues is outside the critical region is considered reducible. The region of interest is chosen as the one with eigenvalues comprised between 250 and 90 d^{-1} corresponding to the medium state variables. The remaining reduced states are the following:

$$\begin{aligned} \mathbf{x}^p &= [X_{ND},\, S_{ND},\, X_S]^T \\ \mathbf{x}^n &= [X_{ND},\, X_S,\, S_O,\, S_{NH},\, S_{ND}]^T \end{aligned} \tag{6.6}$$

6.4.1.2 Proposed Modal Model

If we had strictly considered the homotopy method, we would have obtained different models for the biological reactor zones for the state variables reported in 6.6. The problem is that in such a situation the system would have lost a lot of information:

- All the organic compound are represented only with X_S, the slowly biodegradable matter (representing only a small portion of the effluent

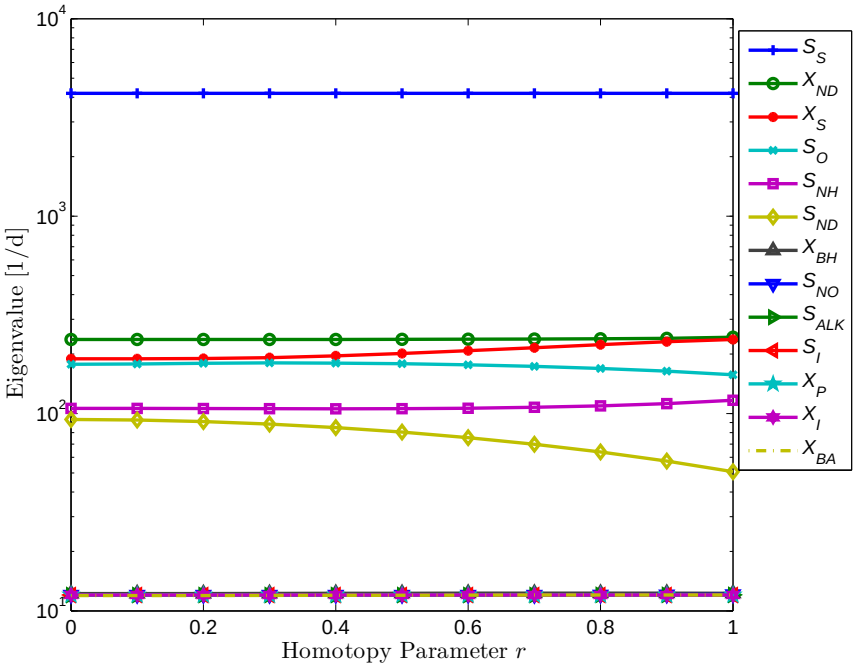


Figure 6.5: Eigenvalue traces for the ASM1 model variables in aerobic zone

COD). Being a particulate component, most of it will be recirculated back with the recycle flow.

- The nitrogen compounds are represented in the anoxic zone only as soluble and particulate biodegradable organic nitrogen, (S_{ND} and X_{ND}). They are essentially representative of the hydrolysis reaction: i.e., the nitrogen removal is not completely represented.

- In the aerobic zone only the ammonia concentration is considered and there is no information about the nitrate behavior.

As for the organic compound, we included the soluble inert organic matter S_I as representative for the COD concentration in the system. We know that inert compounds are not involved in any conversion process, this means that their mass balance contains no stoichiometric and kinetic coefficient. Since it is a soluble component, it leaves the system at the same concentration as it enters (almost 90% of the effluent COD, according to [96]). As a first attempt, we tried to implement the model with the state variables in 6.6 plus the S_I compound. The results were not so good since the nitrogen component is not taken into considerations (its value was constant). For that reason, we choose to represent both zones in the bioreactor with the

same model and also consider the nitrate/nitrite S_{NO}. The state variables for the proposed reduced model are the following:

$$\mathbf{x} = [S_I,\, X_S,\, S_O,\, S_{NO},\, S_{NH},\, S_{ND},\, X_{ND}]^T \tag{6.7}$$

In order to preserve the steady state behavior of the original system and not to destroy the physical interaction between the state variables, we considered the truncation approach with the truncated states as very fast (i.e., $\dot{\mathbf{x}}^{Trunc} \to \infty$). The other adopted approach is the *singular perturbation* method that is a *residualization* method. In this case, the derivatives of the not discarded state variables are approximated to zero (i.e., $\dot{\mathbf{x}}^{Res} \to 0$), leading to a system consisting of differential equations and algebraic equations. Inevitably, longer computational times are required. From the 13 state variables in the ASM1 model we obtain a reduced model with 7 state variables and we refer to this model as *Model1*.

Recalling some of the considerations made by Jeppsson [51] about the hydrolysis processes (not included in the JOM model) a further improvement is introduced. Jeppsson assumptions were motivated essentially because the hydrolysis is a process not well understood due to the uncertainty and to the complex description of the mechanisms involved. Not including the hydrolysis to describe the transformation of organic matter, the hydrolysis of the entrapped nitrogen should not been included as well (in fact, according to the ASM1 model they are the same process). This means that we can not consider also S_{ND} and X_{ND} as state variables. We noted that not considering the hydrolysis process, we should not consider also the slowly biodegradable organic matter, X_s. Neverthenless, this further assumption leads to a system with incomplete information about particulate compound. Therefore, a reduced model involving only 5 variables:

$$\mathbf{x} = [S_I,\, X_S,\, S_O,\, S_{NO},\, S_{NH}]^T \tag{6.8}$$

is developed and we refer to this as *Model2*.
In both approaches, the nonlinearities are maintained, meaning that the system keeps its properties not only in the region around the equilibrium value.

The models are tested considering the experimental data and in Figures 6.6 some results are reported. On the left column the test motions with the truncation approach are depicted, whereas on the right column presents the residualization approach results.

We can note that the effluent COD behaves the same way with the truncation Figure 6.6a approach and with the residualize one Figure 6.6b. Of

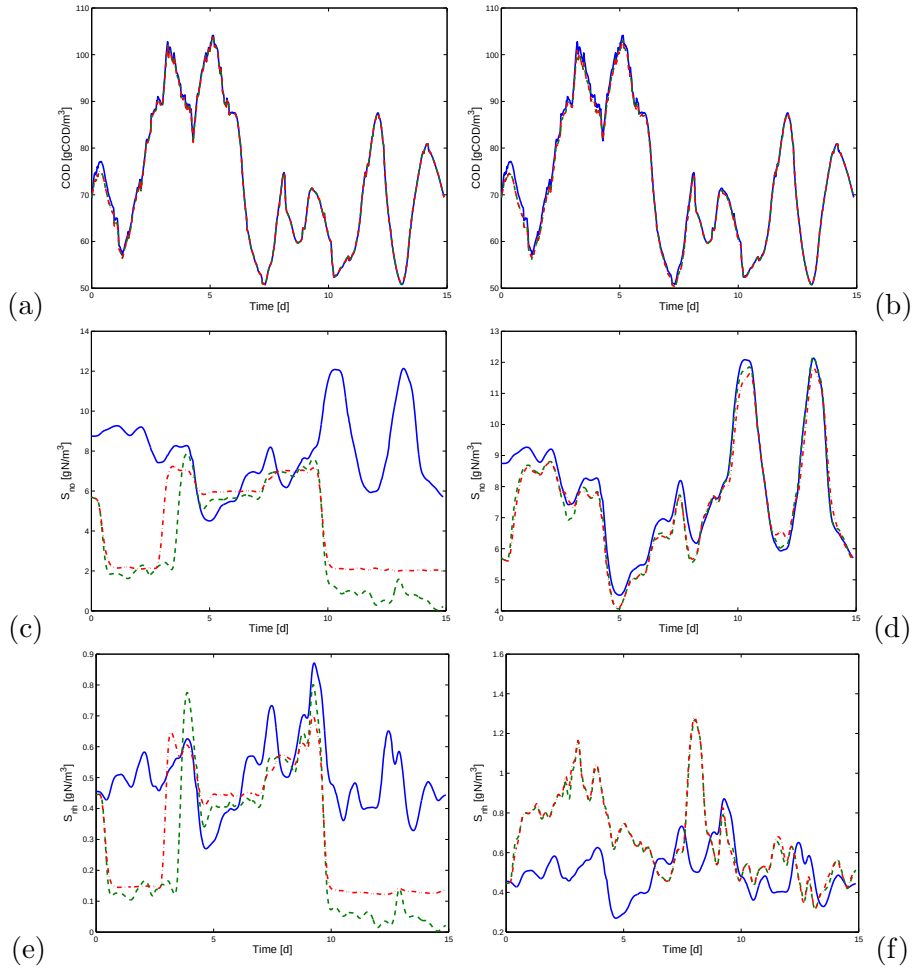


Figure 6.6: Effluent concentrations for the porposed modal models with experimental input data [ASM1 (solid); *Model1* (dashed); *Model2* (dot-dashed)]

course, the COD concentration is independent from the S_{ND} and X_{ND} presence, so its behavior is the same for both *Model1* and *Model2*.

In Figure 6.6c and 6.6d, the test motion for the effluent nitrate compound is shown. As we can note, the S_{NO} gives a good motion tendency when the residualizated approach is taken, but its behavior it is not so good when the model is expressed in the truncated form. Finally, the effluent ammonia concentration is shown in Figures 6.6e and 6.6f: in this case, we can note that the reduced models give better result with the residualization approach even if there is a significant offset (evident in the truncation case). This is

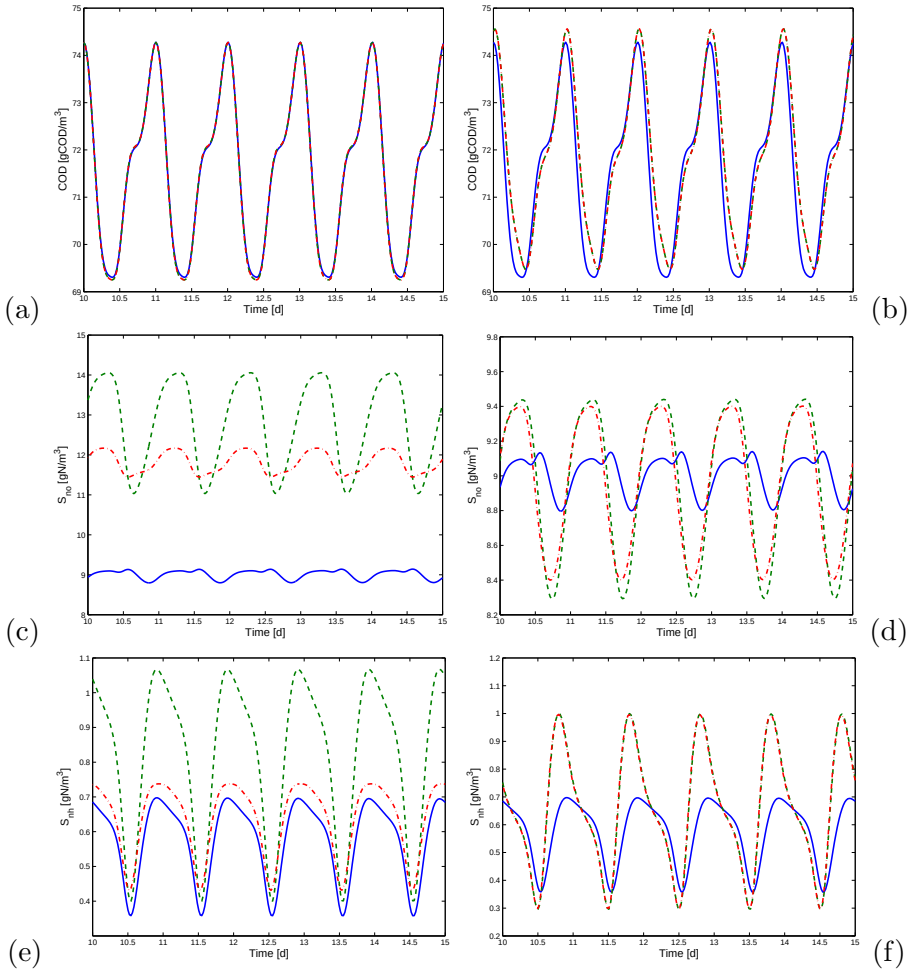


Figure 6.7: Proposed model test motion - Effluent concentrations with periodic input data [ASM1 (solid); *Model1* (dashed); *Model2* (dot-dashed)]

probably due to the fact that the residualization method is more suitable for low frequency modelling: this is not the case for the ammonia concentration. For this reason, since we are aware of the limits in the TecnoCasic data we have alternatively considered the typical variations of dry weather conditions using the weighting functions depicted in Figure 4.1, for both inlet flow rate and load [49]. In Figure 6.7, the output concentrations with the periodic input flow and concentrations are reported. From Figure 6.7a and 6.7b, we can note that effluent COD behaves the same way with both *Model1* and *Model2*.

From Figure 6.6 and 6.7, we also note that the residulialized model is more suitable than the truncated one, even if with truncated model the integration time is significatively less than the residualization time. In Table 6.5 the time elapsed, using a Pentium4 2.4 GHz with 1 Gb RAM machine, is reported for each model simulated for 15 days with periodic inputs. As we can note, the residualized model takes longer CPU time than the trucated one (this is obvious, if we think that by letting some derivatives approximately to zero in the residualization model we are introducing additional stiffness to the system). Even if we have a longer simulation period for the truncated model, the residualized is always faster than the full one.

	Elapsed Time	
Full ASM1	105	s
Model1 Truncated	24	s
Model2 Truncated	25	s
Model1 Residualized	47	s
Model2 Residualized	45	s

Table 6.5: Elapsed time comparison

As for the $Model2$ in residualization form, it is our proposed reduced model. It consists of 5 state variables (6.8) and 10 stoichiometric and kinetic parameters:

$$\mathbf{p} = [Y_H,\, Y_A,\, i_{XB},\, K_{OH},\, K_{OA},\, K_{NO},\, K_{NH},\, K_X,\, \eta_g,\, \eta_h]^T \tag{6.9}$$

along with the five theta coefficients:

$$\begin{aligned}
\theta_1 &= \mu_H X^r_{BH} \frac{S^r_S}{S^r_S + K_{OH}} \\
\theta_2 &= \mu_A X^r_{BA} \\
\theta_3 &= (1 - f_P)(b_H X^r_{BH} + b_A X^r_{BA}) \\
\theta_4 &= k_a S^r_{ND} X^r_{BH} \\
\theta_5 &= k_h X^r_{BH}
\end{aligned} \tag{6.10}$$

The superscript "r" is to highlight that we are considering the residualized state variables. In order to validate the model the assumptions are verified: that is, we essentially confirmed that the coefficients 6.10 can be considered constant without loosing in feasibility. In fact, we noticed that except θ_1, all of them are dependent only on slow dynamics, such as X^r_{BH} and X^r_{BA}, meaning that actually they vary very slowly. The term θ_1 also depends on S^r_S (which has shown a very fast dynamic either in anoxic or in aerobic

zone), however, it has been verified that S_S^r keeps a mean value relatively constant during the time of interest. Therefore, we consider θ_1 constant with time. Finally, the proposed model is summarized in a kind of Petersen matrix form as shown in Table 6.6.

1 S_I	2 X_S	3 S_O	4 S_{NO}	5 S_{NH}	Reaction
		$-\frac{1-Y_H}{Y_H}$		$-i_{XB}$	$\theta_1 \frac{S_O}{S_O+K_{OH}}$
			$-\frac{1-Y_H}{2.86Y_H}$	$-i_{XB}$	$\eta_g \theta_1 \frac{K_{OH}}{S_O+K_{OH}} \frac{S_{NO}}{S_{NO}+K_{NO}}$
		$-\frac{4.57-Y_A}{Y_A}$	$\frac{1}{Y_A}$	$-i_{XB}-\frac{1}{Y_A}$	$\theta_2 \frac{S_{NH}}{S_{NH}+K_{NH}} \frac{S_O}{S_O+K_{OA}}$
	1				θ_3
				1	θ_4
	-1				$\theta_5 \frac{X_S/X_{BH}^r}{K_X+X_S/X_{BH}^r}(\frac{S_O}{S_O+K_{OH}}$ $+\eta_h \frac{K_{OH}}{S_O+K_{OH}} \frac{S_{NO}}{S_{NO}+K_{NO}})$

Table 6.6: Proposed Model

6.4.2 Balanced Model

As stated at the beginning of this Chapter, different reduced model aims at different model reduction approaches in order to obtain the better model for each situation. In this Section, a model that preserves the controllability and observability of the system is developed. We start considering that a balanced realization is an asimptotically stable minimal realization where the controllability and observability gramians are equal and diagonal. Any minimal realization of a stable transfer function can be balanced by a simple state similarity transformation. Following the procedure summarized in

Figure 6.8, we can obtain the reduced balanced model.

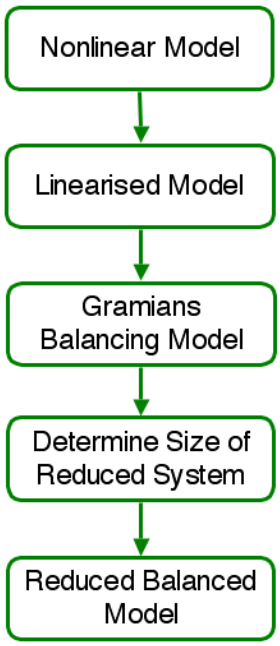


Figure 6.8: Systematic approach for balanced reduction

In our case, we first linearized the model for the whole activated sludge system (108 state variables: 13+1 for the controlled anoxic zone and 13+1 for the controlled aerated one in the bioreactor, 10 particulate variables and 70 soluble in the secondary settler). From the linearized system, the corresponding balanced formulation is derived using the `balreal` function[5]. The entries of the joint gramian (forming the vector of the Hankel singular values σ^H) are then used to reduce the model order. Because σ^H reflects the combined controllability and observability of individual states of the balanced model, we deleted those states associated to a small Hankel singular value (in our case $\sigma^H < 1e^{-5}$) while retaining the most important input-output characteristics of the original system. This is done using `balred` function, which computes the reduced-order approximation for the balanced system. This way, we reduced the model order from 108 to 28. Taking this approach it seems that the system becomes a kind of "black-box" system, where the real meaning of the internal state variable is lost, whereas the input-output

[5] `balreal` gives an equivalent realization with controllability and observability gramians equal and diagonal

relationship is still very well represented. It is clear now why this approach is very useful when the final goal for model reduction is to preserve the controllability and observability properties of the system to design for example a model based controller.

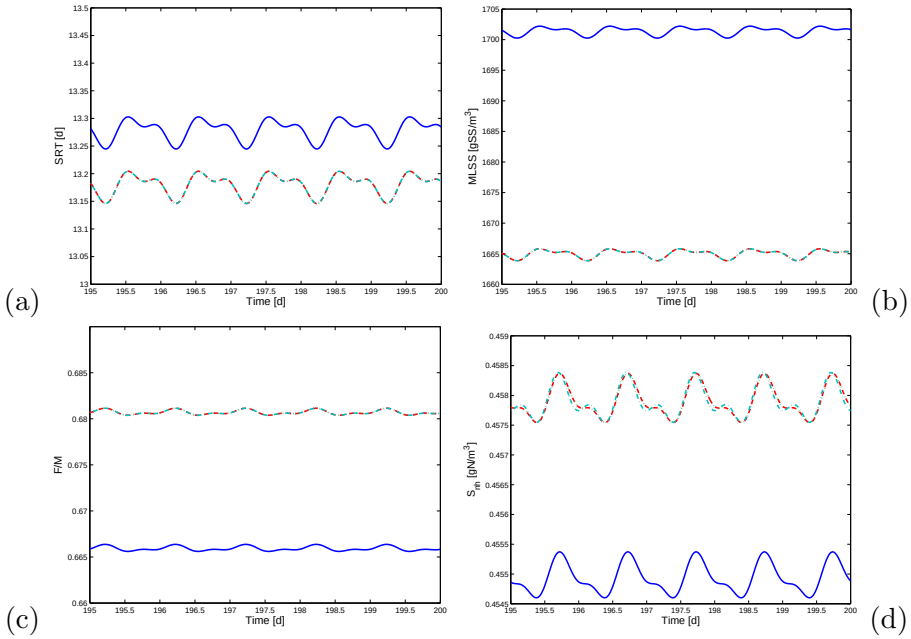


Figure 6.9: Balanced model test motion [ASM1 (solid); linearized (dot); balanced (dashed); residualized (dot-dashed)].

It should be noted that the state variables in the Matlab balanced models are in terms of perturbative variables. That means that an adequate transformation must be done. We investigate the reduced system behavior only with respect to the inputs (i.e., the manipulated variables) and the output (i.e., the controlled variables). In Figure 6.9 some results are reported. In such a case, the input to the system defined a periodic variation on the waste flowrate (as a manipulated variable), whereas the reported outputs are relative to the candidate controlled variables in Chapter 4: the sludge retention time (Figure 6.9a), the mixed liquor suspended solids (Figure 6.9b), the food-to-microorganisms ratio (Figure 6.9c), and of effluent ammonia (Figure 6.9d). As we can note all the models give a good motion tendency when compared to the full model (even if, in some case there is a significative offset). This is quite normal if we think that the model states are reduced to only 28. The constructed variables do not have a real physical meaning, a coordinate transformation must be realized to come back to their original

meaning.

In addition, also the **empirical gramians** could have been used for nonlinear model reduction. Instead, we have tested the approach to check the controllability and observability of the system [97] and to compare to the results obtained with linear gramians and nonlinear observability matrices. The obtained results are omitted because they were not so different that in the linear case, showing that nonlinearities are not so important in the considered region of interest.

Chapter 7

ASP Software Sensors

In Chapter 4 we assumed that all required measurements to build up an optimal control structure were available. Unfortunately, one of the major complications to control and automate a wastewater treatment plant is the lack of sensor for on-line measures. Until very recently, measurements in a treatment system were typically limited to flows, pH and dissolved oxygen. However in a successful manner we can use the simply models reduced (Chapter 6) for monitoring the process, i.e., to predict the time evolution of the process variables on-line. In this way, we can build very powerful monitoring tools which can be used to follow the time evolution of variables that are not accessible on-line. Such tools can also be used for diagnosis about the operation of the plant and help the operator or a supervision system to take the appropriate actions to maintain the process in a good operating conditions, diagnose possible process failure or prevent accidents. In the following, we call these monitoring tools *soft-sensors*.

This Chapter is organized as follows: firstly we briefly explain the soft-sensor meaning and how it might be constructed (Section 7.1). Then we concentrate on the activated sludge process, developing software sensors for the reduced models implemented in Chapter 6: in Section 7.2.1 the GQ *et al.* are used for an extended Kalman filter and a nonlinear geometric observer, the same observer is used also for the $JOMO_2$ model in Section 7.2.2 whereas a simple Luenberger-like observer is implemented with the proposed model *Model2* in section 7.2.3.

7.1 Software Sensors

A software sensor can be defined as an algorithm built from a dynamical model of a process to estimate on-line unmeasured variables and/or un-

known (or poorly known) parameters (e.g., specific reaction rates, or some other kinetic or yield coefficients) from few measurements available on-line (typically, flow rate, nutrient concentrations, turbidity, pH, etc.). In that sense, these tools can be viewed as "*sensors*" based on an algorithm (*software*): for this reason they are called "software sensors" or "soft-sensors". They essentially refer to the state estimation problem of determining the values of the states variables.

In order to describe the basic concepts of a state estimator, we consider the simple case of a process model (without any measurement errors or process noise) as expressed in the state space time invariant form:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} \end{aligned} \tag{7.1}$$

with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. The matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ contain the characteristics parameters of the system, $\mathbf{u}$ is the vector of the inputs or forcing functions, $\mathbf{x}$ are the state variables and $\mathbf{y}$ are the output observations.

If the system in 7.1 is observable, on-line estimates of the states $\hat{\mathbf{x}}$ are obtained from the following observer equation 7.2 in which a driving term is included to minimizing the *observation error* ($\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$) between measured values $\mathbf{y}$ and model prediction $\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}}$:

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}) \tag{7.2}$$

Estimates of the states are therefore obtained by simply integrating equation 7.2 and the design of the observer reduces to the adequate choice of the matrix $\mathbf{K}$, the *gain matrix*. The standard approach starts from the desire to minimize the observation error. The dynamics of the observation error are obtained by subtracting the observer equation (7.2) from the process model (7.1):

$$\begin{aligned} \dot{\mathbf{e}} &= \mathbf{A}(\mathbf{x} - \hat{\mathbf{x}}) - \mathbf{K}\mathbf{C}(\mathbf{x} - \hat{\mathbf{x}}) \\ \dot{\mathbf{e}} &= [\mathbf{A} - \mathbf{K}\mathbf{C}]\mathbf{e} \end{aligned} \tag{7.3}$$

The aim is reduced to the problem of designing the gain matrix in such a way that the observation error decreases in a desirable way.

These concepts bear the same meaning when a more general case is considered: that is, when a nonlinear process is corrupted (as showed in Figure 7.1) by process noise ($\xi(t)$, due to either unknown disturbances or model errors). In addition, a corrupted estimate of the initial conditions and output data (which are some combination of the state variable $\mathbf{h}(t)$), can be

considered. The magnitude of the estimate correction depends also in this case on the gain matrix $\mathbf{K}$ which in turns depends on the error statistics of the model and the output data.

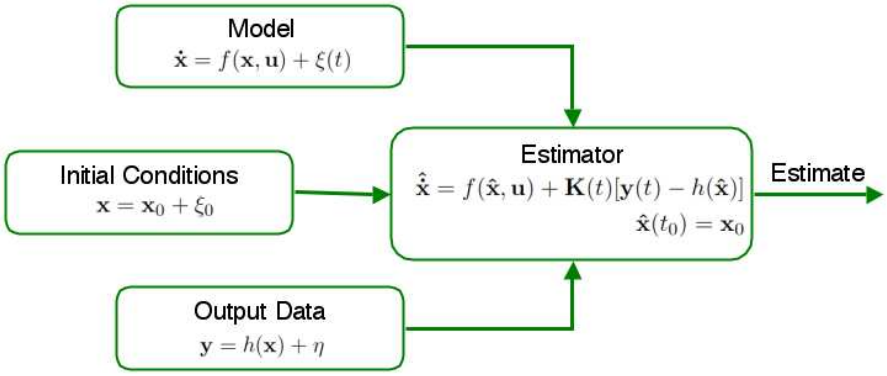


Figure 7.1: Sequential estimator structure

From this model generalization, a sort of classification has been made between *observers* and *estimators*; being the former estimators for state variables of a deterministic system [88], i.e., a system without any significant process noise or measurement error. Bastin and Dochain [9] introduced another distinction: they refer to state *observers* for reconstructing on-line the time evolution of unmeasured process component concentrations, and to on-line state *estimators* for unknown or badly known parameters.

In any case, the state estimation sense remains the same: the problem of determining the values of the state variables from only the knowledge of the outputs and the inputs. Depending on the way to choose the gain matrix $\mathbf{K}$, several examples of software sensors can be found in literature. In particular, there exist four fundamental approaches to observer design [2] for nonlinear system:

- The extended Kalman filter [50]: the most widely used state estimation technique in chemical engineering. Its design is simple but lacks in both stability criteria and systematic tuning procedures.
- Geometric observer [62]: which guarantees robust stability with linear input-output errors dynamics.
- High-gain approach [30]: which guarantees stability, but has a complex tuning procedure.

- Sliding mode [102]: which guarantees robust stability, but has an elaborate design.

In the following we overview only the applied approaches considering an activated sludge process, and even if the estimation technique adopted will be different, the basic procedure in the soft-sensor design is essentially the same. It consists the two following main steps:

- **First Stage**: observability or detectatibility assessment. The evaluation of some measures to determine the degree of observability, the definition of observability indices and unobservable states.

- **Second Stage**: estimator construction and tuning. A systematically with physical interpretation must be performed.

7.1.1 Software sensor design

We said that the first step in software sensor design is the definition of the observability and detectability of the system. So far in the previous Chapter, we have discussed the observability properties assuming that the concept was known, in this section we give a observability definition in order to better understand the first stage in the software sensor design.

Simply speaking, the notion of observability can be defined as the possibility to connect the state variables of a dynamical system to the measured variables via the dynamical model of the system. Essentially, a system is observable if every state variable of the system affects some of the process variables [24]. An important consequence of the observability of a system is the ability to reconstruct the time evolution of the state variables from measured variables in a arbitrary finite time evolution from any initial conditions. To be more precise:

If every initial system state $\mathbf{x}(t_o)$ *can be determined through knowledge of the system inputs* $\mathbf{u}(t)$ *and the system output* $\mathbf{y}(t)$ *over some finite time interval* $t_o \leq t \leq t_1$*, then the system is completely observable* [88].

Conditions for observability have been derived for a wide number of classes of systems. For simplicity, we consider a linear system in the state space form in 7.1. It can be shown [88] that the system is completely observable

if and only if the following matrix $\mathcal{O}$ has full rank.

$$\mathcal{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} \tag{7.4}$$

A weaker property than observabilty is *detectability*. Detectability is the property that all unstable modes of the process are observable:

- $m \leq \text{rank}(\mathcal{O}) \leq n$;
- eigenvalues of the matrix $\mathbf{A}$ must be strictly negative.

Clearly, any observable system is also detectable. The property of detectability is important because in such a way we can partition the state vector $\mathbf{x}$ in observable $\mathbf{x}_I$ and unobservable state $\mathbf{x}_{II}$. Moreover, we may successfully design an observer/estimator for an unobservable but detectable system so as to estimate the unstable modes.

7.1.2 Luenberger Observer

The Luenberger Observer is the simplest approach to observer design. The objective is to select $\mathbf{K}$ such that the error dynamics in 7.3 are asymptotically stable (i.e., the error converges to zero). This is achieved by choosing $\mathbf{K}$ such that [9]:

- the matrix $[\mathbf{A} - \mathbf{KC}]$ and its derivative are bounded, so that convergence is guaranteed:

$$\|\mathbf{A} - \mathbf{KC}\| \leq C_1$$
$$\|\frac{d}{dt}[\mathbf{A} - \mathbf{KC}]\| \leq C_2$$

- the eigenvalues of $[\mathbf{A} - \mathbf{KC}]$ have strictly negative parts, so that stability is assured

The importance of the state observabilty becomes clearer: if the system is not observable, it is then not possible to freely assign the dynamics of the observation errors (we can not freely choose the elements of $\mathbf{K}$).

7.1.3 Extended Kalman Filter

In 1960, Kalman published his famous paper describing a recursive solution to the discrete data linear filtering problem [59]. Since that time, due in large part to advances in digital computing, the Kalman filter has been the subject of extensive research and application. In this section, we give only a brief explanation of the Kalman filter and of the extended Kalman filter, which is applied to the activated sludge process.
For a more detailed and theoretical derivation of the Kalman filter and its variant some extensive references should be consulted. For example a very "friendly" introduction to the general idea of the Kalman filter can be found in the first chapter of [74], while a more complete discussion can be found in [69].

Simply speaking, the Kalman filter estimates a process by using a form of feedback control: the filter estimates the process rate at some time and then obtains feedback in the form of noisy measurements. As such the equations involved in the Kalman filter fall into two groups:

1. *Time update equations*, which are responsible for projecting forward (in time) the current state and error covariance estimates to obtain *a priori* estimates for the next time step.

2. *Time measurements equations*, which are responsible for the feedback, i.e. for incorporating a new measurements into the *a priori* estimate to obtain an improved *a posteriori* estimate.

The time update equations can also be thought of as *predictor* equations, while the measurement update are the *corrector* equations. Indeed the final estimation algorithm resembles that of a *predictor-corrector* algorithm. The basic operation of the EKF is showed in Figure 7.2.

The Kalman filter addresses the general problem of trying to estimate the states $\mathbf{x} \in \mathbb{R}^n$ of a discrete-time controlled process that is governed by a linear stochastic difference equation. However, some of the most interesting and successful applications of Kalman filter have been with the estimation of nonlinear processes (see for example [7] and [8]). A Kalman filter that linearizes around the current mean and covariance is called *Extended Kalman Filter* (EKF). The process is assumed to be governed by the nonlinear stochastic differential equation:

$$\dot{\mathbf{x}}_k = f(\mathbf{x}_k, \mathbf{u}_k, w_k) \tag{7.5}$$

with a measurement $\mathbf{z} \in \mathbb{R}^m$, such that:

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k) \tag{7.6}$$

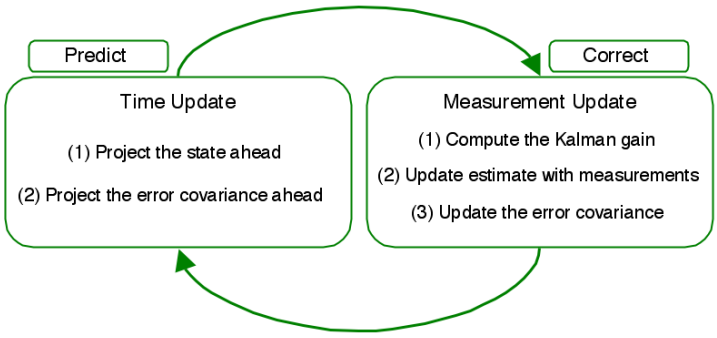


Figure 7.2: Scheme of the Extended Kalman Filter operation

The random variables $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the process and measurement noise. The nonlinear function f relates the state at the previous time step $k-1$ to the state at the current time step k and it includes as parameters any driving function $\mathbf{u}_k$. The nonlinear function h relates the state $\mathbf{x}_k$ to the measurements $\mathbf{z}_k$.

We define $\hat{\mathbf{x}}_k^- \in \mathbb{R}^n$ to be the *a priori*[1] state estimate at step k given knowledge of the process prior to step k, and $\hat{\mathbf{x}}_k \in \mathbb{R}^n$ to be the *a posteriori* state estimate k given the measurement $\mathbf{z}_k$. In mathematical notation, the EKF calibration procedure is:

- For the time update equations (i.e., the prediction equations):
  1. Projection of the state from t_k to t_{k+1}: $\dot{\mathbf{x}}_k = f(\mathbf{x}_k + \mathbf{u}_k, 0)$
  2. Projection of the error covariance from t_k to t_{k+1}:
     $\dot{\mathbf{P}} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k)\mathbf{P} + \mathbf{P}f^T(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) + \mathbf{Q}$.
     This represents the *Riccati equation*, where $\mathbf{Q}$ is the model error covariance, $\mathbf{P}$ is the estimate error covariance.
- Fo the measurement update equations (i.e., the correction part):
  1. Compute the Kalman gain matrix:
     $\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$
     where $\mathbf{R}$ represents the measurement error variance
  2. Update estimate with the measurements:
     $\mathbf{z}_k$: $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{h}_k)$
  3. Update the error covariance: $\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_k^-$

[1] Note that with the "minus sign" we will indicate the a priori variables

In the actual implementation of the filter, the measurement noise covariance $\mathbf{R}$ is usually measured prior to operating the filter. Measuring $\mathbf{R}$ is generally practical (possible) because we need to be able to measure the process anyway (while operating the filter) so we should generally be able to take some off-line sample measurements in order to determine the variance of the measurement noise. In other words, $\mathbf{R}$ depends on the measure precision. The determination of $\mathbf{Q}$ is generally more difficult as we typically do not have the ability to directly observe the process we are estimated; so, $\mathbf{Q}$ can be considered as a tuning parameter.

7.1.4 Nonlinear Geometric Observer

The nonlinear geometric observer (NGO) has been recently proposed by Alvarez and Lopez [3]. The NGO is a nonlinear estimator based on the differential geometry theory which is a method principally characterized by its applicability to either observable or detectable systems, encompassing a broad class of plants in process system engineering. The proposed estimator design includes a robust local convergence and a systematic construction-tuning procedure. It was successfully applied to solve the local nonlinear estimation problem of a free-radical homopolymer reactor [3], [2], to infer the concentration in a catalytic reactor [72], and to estimate the product composition profiles for a distillation column [109]. A detailed discussion on the construction of the estimator can be found in [3] and in [70]. Only a brief summary of the procedure is given here.

Let us consider a nonlinear dynamical system in the general form:

$$\begin{aligned}\dot{\mathbf{x}} &= f(\mathbf{x},\, \mathbf{u},\, \mathbf{p})\\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{p})\end{aligned}$$

where $\mathbf{p}$ are the model parameters, and $\mathbf{x}$, $\mathbf{u}$, $\mathbf{y}$ have the usual meaning with $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\mathbf{y} \in \mathbb{R}^m$.

According to Alvarez and Lopez [3], the motion $\mathbf{x}(t)$ is *robustly exponentially estimable* if there are m integers (*observability indexes*) $\kappa_1, \kappa_2, \ldots, \kappa_m$ ($\kappa_1 + \kappa_2 + \cdots + \kappa_m \leq n, \kappa_i > 0$) and a map $\Phi_{II}(\mathbf{x},\, \mathbf{u},\, \mathbf{r}) = [\Phi_{k+1}, \ldots, \Phi_n]$ such that, in some neighborhood about $[\mathbf{x},\, \mathbf{u},\, \mathbf{r}]$ the following conditions are verified:

1. the map $\Phi(\mathbf{x},\, \mathbf{u},\, \mathbf{p}) = [\Phi_I^T,\, \Phi_I^T]^T$ is robustly invertible with respect to $\mathbf{x}$;

2. the map $\psi = [L_f^{\kappa_1} h_1,\, \ldots,\, L_f^{\kappa_m} h_m]^T$ is Lipschitz continuous;

3. the motion of the unobservable state $\mathbf{x}_{II}$ is robustly exponentially stable.

The map Φ_I is defined as:

$$\Phi_I(\mathbf{x},\, \mathbf{u},\, \mathbf{p}) = [h_1,\, L_f h_1, \ldots,\, L_f^{\kappa_1 - 1}, \ldots, h_m,\, L_f h_m, \ldots, L_f^{\kappa_m - 1} h_m]^T \quad (7.7)$$

where, the expression $L_f^i h_j$ indicates the recursive directional derivative of the time varying scalar field $h_j(x,\, t)$, with $j = 1, \ldots, m$, along the time-varying vector-field $\mathbf{f}(x, t)$. When $k = n$ (i.e., when there are not unobservable dynamics and condition (3) is trivially met) the motion $\mathbf{x}(t)$ is said to be *robustly exponentially* (RE) *observable*. Otherwise, if $\kappa < n$ the motion is said to be *robustly exponentially* (RE) *detectable*.

When the above conditions are verified the following observer can be constructed:

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_I &= f_I(\hat{\mathbf{x}},\, \mathbf{u},\, \mathbf{p}) + \Phi_x^{-1} \mathbf{K}_0 [\mathbf{y} - \hat{\mathbf{y}}] \\ \dot{\hat{\mathbf{x}}}_{II} &= f_{II}(\hat{\mathbf{x}},\, \mathbf{u},\, \mathbf{p}) \\ \hat{\mathbf{y}} &= \mathbf{h}(\hat{\mathbf{x}}_I,\, \mathbf{p}) \end{aligned} \quad (7.8)$$

where $\mathbf{x} = [\mathbf{x}_I,\, \mathbf{x}_{II}]$, $\mathbf{x}_I$ and$\mathbf{x}_{II}$ being respectively the observable and unobservable states with $\mathbf{x}_I \in \mathbb{R}^k$ and $\mathbf{x}_{II} \in \mathbb{R}^{n-k}$.
In the particular case $k = n$, there are not unobservable dynamics and the proper form of the observer is:

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= f_I(\hat{\mathbf{x}},\, \mathbf{u},\, \mathbf{p}) + \Phi_x^{-1} \mathbf{K}_0 [\mathbf{y} - \hat{\mathbf{y}}] \\ \hat{\mathbf{y}} &= \mathbf{h}(\hat{\mathbf{x}},\, \mathbf{p}) \end{aligned} \quad (7.9)$$

In the equations (7.8) and (7.9), $\mathbf{K}_0$ represents a constant matrix whose entries are the gains of the observer. The value of the gains are tuning parameters of the observer, and they should be set such that the reference linear, non interactive and pole-assignable error dynamics is stable [3].

7.2 ASP Soft-Sensor Applications

Many soft-sensor applications in wastewater treatment can be found in literature. For example Lindberg [68] proposed on-line methods for estimating the time-varying respiration rate and the nonlinear oxygen transfer function from measurements of the dissolved oxygen concentration. Also a work presented by Sotomayor *et al.* [104] deals with oxygen estimation: they considered the design of a soft-sensor for on-line estimation of the biological activity of microorganisms in an ASP, which are intimately related to the

dissolved oxygen concentration. Benazzi *et al.* [10] proposed a soft-sensor implementation based on an EKF for the on-line tracking of the total suspended solids.

In our study, the on-line monitoring problem consists on designing a soft-sensor for estimating mainly the effluent ammonia and organic compound, from the available measures of dissolved oxygen in both bioreactor zones and the nitrate/nitrite in the aerobic zone:

$$\mathbf{y} = [y_1,\, y_2,\, y_3]^T = [S_O^p,\, S_O^n,\, S_{NO}^n]^T \tag{7.10}$$

The software sensors is based on the reduced models presented in Chapter 6 and the three measurements in equation 7.10.

7.2.1 Case 1: Application with GQ *et al.* Model

In this particular case, the soft-sensors (EKF and NGO) are based on the Gomez-Quintero *et al.* model [34] to infer the unmeasured readily biodegradable substrate and ammonia concentrations before the settler in an ASP.

Nonlinear Geometric Observer

We saw in Section (7.1.4) that the soft-sensor methodology is based on the methodology developed in [3] and [70], with a systematic construction, robust convergence rate, and with a simple tuning procedure.

According to Alvarez and Lopez [3], the bioreactor motion $\mathbf{x}(t)$ must be robustly-exponentially detectable with the observability vector:

$$\mathbf{k} = (\kappa_1,\, \kappa_2,\, \kappa_3)^T = (2,\, 2,\, 2)^T \tag{7.11}$$

Recalling the model equations in Section 6.2, the state partition can be written as:

$$\begin{aligned} \mathbf{x}_I &= [x_2,\, x_3,\, x_5,\, x_6,\, x_7,\, x_8]^T = [S_O^p,\, S_{NH}^p,\, S_{NO}^n,\, S_O^n,\, S_{NH}^n,\, S_S^n] \\ \mathbf{x}_{II} &= [x_1,\, x_4]^T = [S_{NO}^p,\, S_S^p] \end{aligned} \tag{7.12}$$

and recalling also the exogenous inputs $\mathbf{d}$ in (6.4), the map Φ is given by the measured outputs and some of their time-derivatives:

$$\Phi(\mathbf{x},\, \mathbf{d},\, \mathbf{p}) = [y_1,\, \dot{y}_1,\, y_2,\, \dot{y}_2,\, y_3, \dot{y}_3] \tag{7.13}$$

The conditions (1) and (2) in Section 7.1.4 are met along the biological reactor, and, therefore, the system is robustly-exponentially detectable. Furthermore, the observability and stability conditions have been verified.

The construction of the geometric estimator as in equations 7.8, follows from a straightforward consequence of the detectability properties. In [3] and [70] some strategies for the estimator tuning are given. According to this, the gains can be calculated as follows:

$$\mathbf{K}_0 = \begin{bmatrix} k_{11} & 0 & 0 \\ k_{12} & 0 & 0 \\ 0 & k_{21} & 0 \\ 0 & k_{22} & 0 \\ 0 & 0 & k_{31} \\ 0 & 0 & k_{32} \end{bmatrix} \tag{7.14}$$

where $k_{i1} = 2\zeta\omega_i$ and $k_{i2} = (\omega_i)^2$. The parameter ζ is the dumping factor, which iset according to literature [106] as $\zeta = 0.71$ in order to have a response with moderate oscillations. The characteristic frequency ω_i is selected such that the estimator response is faster than the reactor response. In this case, we selected the estimator parameters $\omega_1 = \omega_2 = \omega_3 = 150d^{-1}$.

Extended Kalman Filter

Two different EKFs are implemented: one to infer the state vector $\mathbf{x}^p$ in the anoxic zone using the measured dissolved oxygen concentration in that zone and the second to infer $\mathbf{x}^n$ in the aerobic state using the measured dissolved oxygen and the nitrate/nitrite concentration in that zone. Follows the state vector partition:

$$\mathbf{x}^p = [x_1,\, x_2,\, x_3,\, x_4,\, x_9]^T = [S^p_{NO},\, S^p_O,\, S^p_{NH},\, S^p_S,\, K^p_{La}]$$
$$\mathbf{x}^n = [x_5,\, x_6,\, x_7,\, x_8,\, x_{10}]^T = [S^n_{NO},\, S^n_O,\, S^n_{NH},\, S^n_S,\, K^n_{La}]$$

The state variables x_9, x_{10} correspond to the oxygen mass transfer in the anoxic and aerated zone, respectively. They were added to the estimator states because they are complex and not well known function of the system states.

The general form of the EKF is reported in 7.1.3, the tuning parameters in the Riccati equation are conveniently chosen in order to obtain good performance in this particular situation.

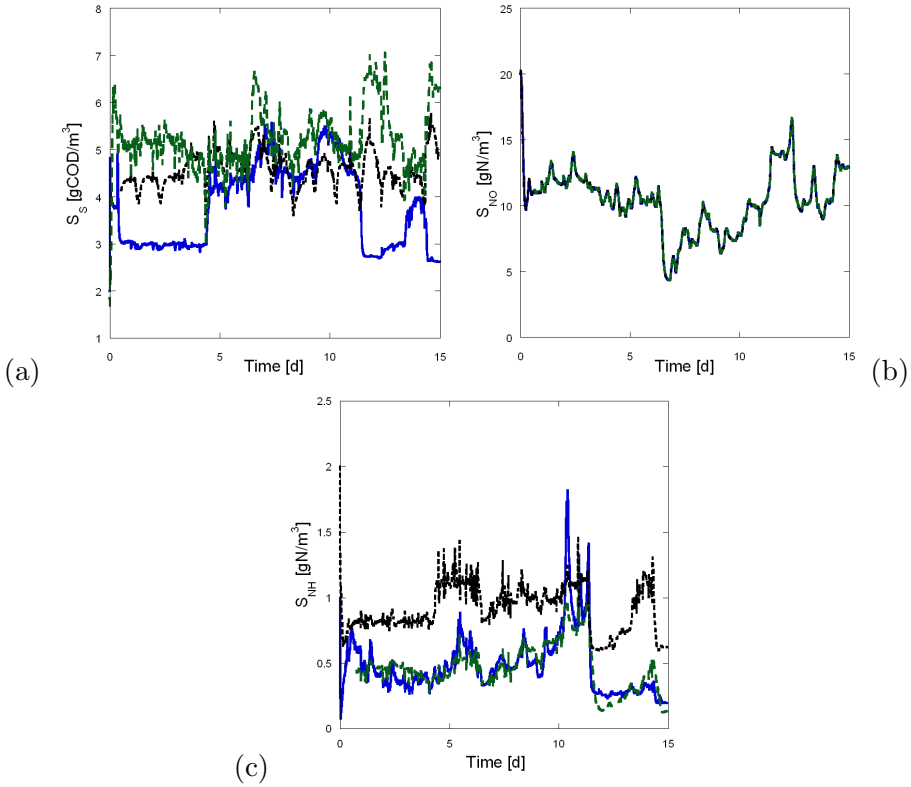


Figure 7.3: Estimator performances with the GQ *et al.* [Full model (solid); EKF (dot-dashed); NGO (dashed)]

Comparison between NGO and EKF

In Figure 7.3 some results for both NGO and EKF are reported for comparison to the full complex model. In Figure 7.3a the estimators outputs for the concentration of the readily biodegradable sustrate is shown. As we can note, there is not a good agreement with the full complex model, even if the EKF infers better than the NGO. This is probably due to an excessive mismatch between the reduced model and the detailed one. In Figure 7.3b the nitrate/nitrite concentration is reported, the good convergence is quite obvious, being the state incorporated as measured variable. In Figure 7.3c the ammonia concentration behavior at the bioreactor exit is shown. As we can note, in this case good results are obtained with respect to the full complex model and this is especially true when the ammonia is inferred with the EFK. However, it should be noted that even if an offset is present the resulting error remains inside the measure's precision region.

7.2.2 Case 2: Application with the JOMO$_2$ Model

The NGO is applied to infer the effluent organic compound and ammonia concentrations predicted by the JOMO$_2$ (see Section 6.3.1). Recalling the model equations in Table 6.3, the state partition is expressed as:

$$\begin{aligned} \mathbf{x}_I &= [X^p_{COD},\, S^p_O,\, X^n_{COD},\, S^n_O,\, S^n_{NO},\, S^n_{NH}]^T \\ \mathbf{x}_{II} &= [X^p_{BH},\, X^p_{BA},\, S^p_{NO},\, S^p_{NH},\, X^n_{BH},\, X^n_{BA}]^T \end{aligned} \tag{7.15}$$

According to Alvarez and Lopez [3], the motion of $\mathbf{x}(t)$ of the reactors in robustly exponentially detectable, since the following conditions are met:

1. The map $\Phi(\mathbf{x},\, \mathbf{d},\, \mathbf{p})$ is invertible for $\mathbf{x}_I$:

$$\Phi = [S^p_O,\, \frac{dS^p_O}{dt},\, S^n_O,\, \frac{dS^n_O}{dt}, S^n_{NO},\, \frac{dS^n_{NO}}{dt}].$$

2. The motion of the unobservable dynamics $\mathbf{x}_{II}$ is stable.

Also in this case, the construction of the geometric observer follows from a straightforward consequence of the detectability property and also the gain matrix present the same form as in equation 7.14.

In Figure 7.4, some results are reported. The NGO outputs are compared with the JOMO$_2$ and with the full complex model. We note that, even if the agreement is not very good, the observer tries to better infer the effluent COD (Figure 7.4a) and ammonia (Figure 7.4c) concentrations. This is interpreted as due to the excessive mismatch existing between reduced and full complex model. Also in this case, the nitrate/nitrite (Figure 7.4b) concentration shows an obvious good fitting, in fact this variables is incorporated as measured variable.

7.2.3 Case 3: Application with *Model2*

For the low-order model proposed in Chapter 4 also a simple Luenberger-like observer with constant gains is implemented. This was motivated by the fact that the model showed a good convergence with respect to the full ASM1 model. Also in this case, the observer estimates the organic compound and ammonia predicted by the reduced model, using equation 7.10 as measured states. Recalling the state variables for the reduced system *Model2*, firstly the system observability has been tested. Then, the observability matrix has been computed showing that $rank(\mathcal{O}) = 3$. This means that the system is only detectable and that the state partition can be represented as following:

$$\begin{aligned} \mathbf{x}_I &= [S^p_O,\, S^n_O,\, S^n_{NO},\, S^n_{NH}]^T \\ \mathbf{x}_{II} &= [S^p_I,\, X^p_S,\, S^p_{NO},\, S^p_{NH},\, S^n_I,\, X^n_S]^T \end{aligned} \tag{7.16}$$

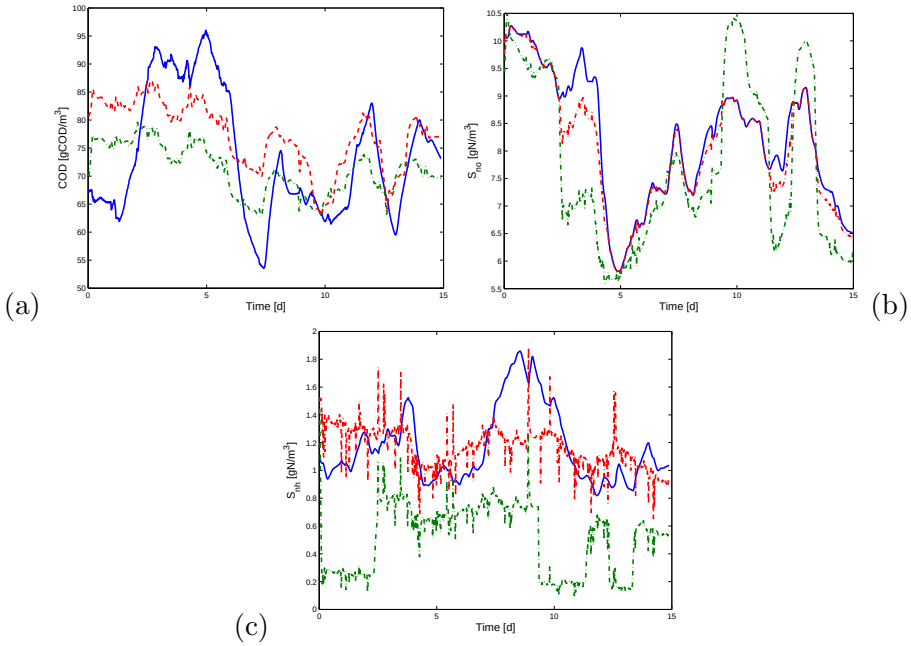


Figure 7.4: Estimator performance with the $JOMO_2$ [Full model (solid); Reduced Model (dot-dashed); NGO (dashed)]

This is obvious, if we think that neither dissolved oxygen nor nitrate/nitrite state variable depend on inert soluble substrate and on slowly biodegradable substrate (which in turn are not observable with states in 7.10).

The soft-sensor is tested with the experimental data (Figure 7.5) and with the periodic data (Figure 7.6) inputs.

The system shows the same good tendency for the effluent COD concentration for both experimental (Figures 7.5a) and periodic data (Figure 7.5a and Figure 7.6a). The same holds for the nitrate/nitrogen effluent concentration (Figure 7.5b and 7.6b), but also in this case it has been used as measured variable. The effluent ammonia concentration (Figure 7.6c) showed a not so good agreement, even if in both cases (Figure 7.5c and 7.6c) the observer tries to correct the output estimates. However, considering that the measurement errors in a common potentiometric ammonia sensor are around 10% the error in soft sensor estimates is not too large.

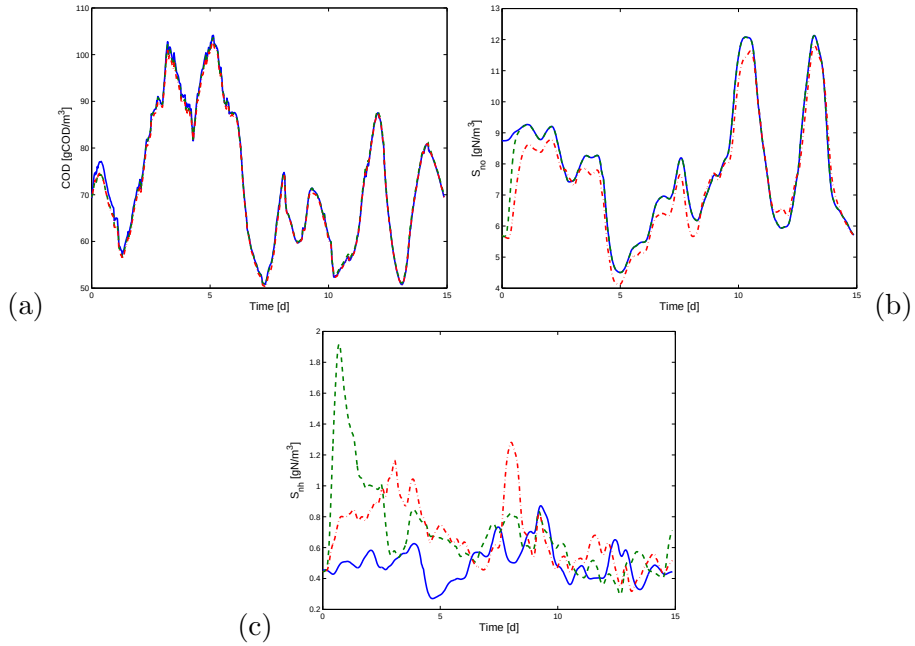


Figure 7.5: Estimator performance with *Model2* and experimental input data [ASM1 model (solid); Reduce Model (dot-dashed); Observer (dashed)]

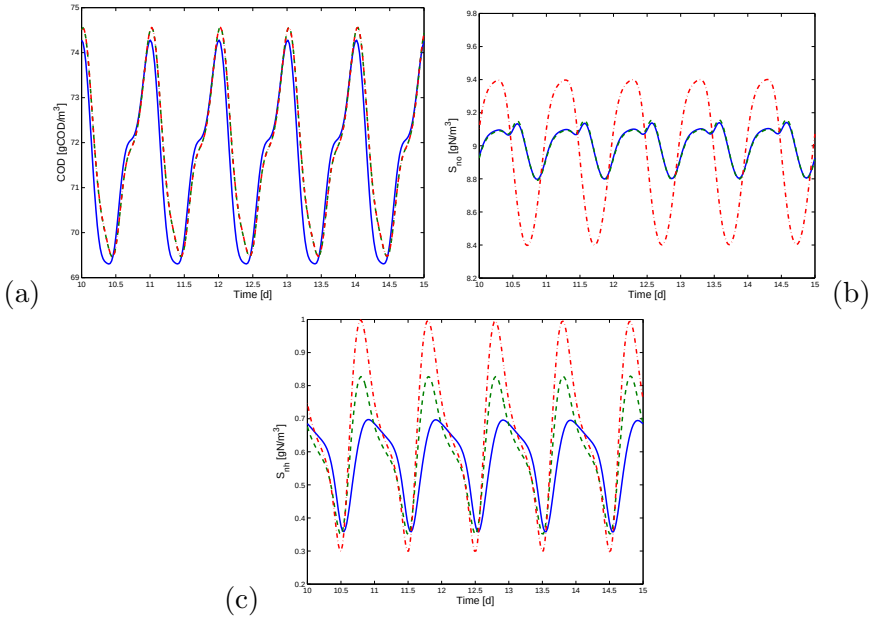


Figure 7.6: Estimator performance with *Model2* and periodic input data [ASM1 model (solid); Reduced Model (dot-dashed); Observer (dashed)]

Chapter 8

Conclusions

In this thesis several aspects and problematics associated with modelling and control of activated sludge process have been outlined and investigated.

To represent the process, the state of the art model has been adopted with regard to the bioreactor to characterize the removal of organic and nitrogen pollutants, and the double exponential velocity has been used to characterize the settling process. Using Matlab/Simulink and a commercial software, the activated sludge process has been simulated and calibrated with respect to experimental data collected from a real wastewater treatment plant.
The calibration has been performed by means of sensitivity analysis and optimization procedures. This task appeared very difficult because data from the real plant were essentially off-line data and too sparse to have a perfectly calibrated model. If the goal of this work were to obtain a perfect representation of the real plant situation, then a experimental campaign to collect data should have been planned. Otherwise soft-sensors to estimate the model parameters should have been designed. However, the final and main goal of this work was apply advanced control structures and design soft-sensors to the estimate the unmeasured variables.

In order to find the optimal operation for the activated sludge process an economical cost function was initially defined. The estimated costs for running pumps and blowing air were optimized for both the specific operating conditions and disturbances. A *self-optimization control* procedure was then applied in order to find the minumum of the cost function and the corresponding optimal operating conditions of the process according to the given constraints.
A first optimization procedure has been performed in order to find the optimal setpoints for the dissolved oxygen concentration when a DO controller

is present. The resulting operating costs have been considerably reduced. Furthermore, the optimization procedure was applied in two other different situations: i) in the presence of equalization tank (this means constant influent flowrate at the ASP and in turn the system has a single degree of freedom) and ii) in the absence of an equalization tank, which means that the influent flowrate can not be considered constant and that the system has two degrees of freedom for optimization.
The two situations resulted in different control strategies characterized by controlled variables that give rise to the optimal configuration. When only one degree of freedom is available, the resulting optimum consists of controlling the mixed liquor suspended solids by manipulating the waste activated sludge flowrate. On the other hand, for the two degrees of freedom case, the resulting configuration consists of manipulating both the wasted sludge and the recycled sludge flowrate. For this case, a decentralized structure has been chosen and several configurations have been considered by applying the minimum singular value rule and a number of analysis on the total cost and the process. Controlling the mixed liquor suspended solids by manipulating the wasted sludge flowrate and controlling the nitrate/nitrite concentration in the last anoxic zone demonstrated to be the configuration with the best self-optimizing properties for our system.

Since the lack in sensors has been recognized as one of the main problems in improving automation and control in wastewater treatment plants, in this thesis a number of different *soft-sensors* have been designed in order to estimate the unknown concentrations. Being based on robust hard-sensors and a mathematichal modelling, the soft sensor can be developed using a reduced but still representative ASM1 model. In such a way the main drawbacks of the full ASM1 model (i.e. lack of identifiably, nonlinearities, many parameters to be estimated) can be exceeded.
Several models of reduced order have been found in literature. Two of those have been studied more extensively and applied to our ASP configuration. In particular, the model proposed by Jeppsson have been modified in order to take into consideration the dissolved oxygen as state variable. Furthermore, several systematic techniques for model order reduction have been studied and applied to the activated sludge model. Modal residualization was found to provide a simple reduced model very representative of ASP.
This model and also the two models from literature have been used to design and test different soft-sensosing approaches (extended Kalman filter, nonlinear geometric observer and Luenbergher-like observer). The obtained positive results showed it is not necessary true that the use of a full model is the best way to obtain good process representation for monitoring and control purposes.

Appendix A

ASM1 Petersen Matrix

Most biological process models follow the standard matrix notation, in Table A.1. The notation makes clear the processes incorporated in the model and the state variables involved. The matrix is usually referred to *Petersen Matrix* and is well known by the modellers of biological wastewater treatment system. Here, the state variables are denoted with the subscript "i" ($i = 1, \ldots, 13$) and are numered and listed across the top. The important processes, designed by ρ_j ($j = 1, \ldots, 8$) are shown in separated rows. The actual process rate is shown in the rightmost column:

- ρ_1 is the aerobic growth of heterotrophs;
- ρ_2 is the anoxic growth of heterotrophs;
- ρ_3 is the aerobic growth of autotrophs;
- ρ_4 is the decay of heterotrophs;
- ρ_5 is the decay of autotrophs;
- ρ_6 is the ammonification of soluble organic nitrogen;
- ρ_7 is the hydrolysis of entrapped organics;
- ρ_8 is the hydrolysis of entrapped organic nitrogen.

The entries within the table are the stoichiometric parameters, denoted by ν_{ij}, as used in defining the net process rate for a component. The parameters define the mass relationship between components: if a process does not directly affects a component rate then the corresponding table cell is empty. The net reaction rate of a component, denoted by "r_i" is the sum of all the process rates which cause a change in the mass of that component:

$$r_i = \sum_j \nu_{ij} \rho_j$$

$i \rightarrow$ $\downarrow j$	1 S_I	2 S_S	3 X_I	4 X_S	5 X_{BH}	6 X_{BA}	7 X_P	8 S_O	9 S_{NO}	10 S_{NH}	11 S_{ND}	12 X_{ND}	13 S_{ALK}	Reaction
ρ_1		$-\frac{1}{Y_H}$			1			$-\frac{1-Y_H}{Y_H}$		$-i_{XB}$			$-\frac{i_{XB}}{14}$	$\mu_H \frac{S_S}{K_S+S_S}$ $\frac{S_O}{K_{OH}+S_O} X_{BH}$
ρ_2		$-\frac{1}{Y_H}$			1				$-\frac{1-Y_H}{2.86Y_H}$	$-i_{XB}$			$\frac{1-Y_H}{14 \cdot 2.86 Y_H}$ $-\frac{i_{XB}}{14}$	$\mu_H \frac{S_S}{K_S+S_S}$ $\frac{K_{OH}}{K_{OH}+S_O}$ $\frac{S_{NO}}{K_{NO}+S_{NO}} \eta_g X_{BH}$
ρ_3						1		$-\frac{4.57-Y_A}{Y_A}$	$\frac{1}{Y_A}$	$-i_{XB} - \frac{1}{Y_A}$			$-\frac{i_{XB}}{14} - \frac{1}{7Y_A}$	$\mu_A \frac{S_{NH}}{K_{NH}+S_{NH}}$ $\frac{S_O}{K_{OA}+S_O} X_{BA}$
ρ_4				$1-f_P$	-1		f_P					i_{XB} $-f_P i_{XP}$		$b_H X_{BH}$
ρ_5				$1-f_P$		-1	f_P					i_{XB} $-f_P i_{XP}$		$b_A X_{BA}$
ρ_6										1	-1		$\frac{1}{14}$	$k_a S_{ND} X_{BH}$
ρ_7		1		-1										$k_h \frac{X_S/X_{BH}}{K_X+X_S/X_{BH}}$ $[\frac{S_O}{K_{OH}+S_O}+$ $\eta_h \frac{K_{OH}}{K_{OH}+S_O}]$ $\frac{S_{NO}}{K_{OH}+S_{NO}}] X_{BH}$
ρ_8											1	-1		$\rho_7 (X_{ND}/X_S)$

Table A.1: Tabular format for the ASM1

Appendix B

ASP Model Simulation Parameters

The parameters in the activated sludge model in Chapter 2 are repoted. Table B.1 shows the influent fraction used to transform the measured variables into state variable. Table B.2 gives details about the settler parameters used at steady-state condition. In Table B.3 the kinetic and stoichiometric parametrs in the ASM1 model are shown.

Inert frac. of soluble COD	f_{S_I}	0.35	–
Substrate frac. of particulate COD	f_{X_S}	0.75	–
Heter. frac. of particulate COD	f_{X_S}	0.2	–
Ammonia/TKN ratio	f_{X_S}	0.36	–
Part. Org. N/ Total Org. N ratio	f_{xn}	0.9	–
Particulate COD/Volatile SS	i_{cv}	1.4	$gCOD/gVSS$
Volatile SS/ Total SS	f_{xn}	0.6	$gVSS/gTSS$

Table B.1: Influent Fractions

$v_0^{'}$	274	m/d
v_0	210	m/d
r_h	0.000401	m^3/gSS
r_p	0.0025	m^3/gSS
f_{XP}	0.001	–

Table B.2: Settler model parameters

Y_H	0.67	$g(cellCOD\,formed)/g(cellCOD\,oxidized)$
Y_A	0.24	$g(cellCOD\,formed)/g(cellN\,oxidized)$
f_P	0.08	
i_{XB}	0.086	$gN/gCOD$
i_{XP}	0.06	$gN/gCOD$
μ_H	6	$1/d$
K_{SH}	20	$gCOD/m^3$
K_{OH}	0.2	gO_2/m^3
K_{NO}	0.5	gN/m^3
b_H	0.62	$1/d$
η_g	0.8	
μ_A	0.8	$1/d$
K_{NH}	1.0	gN/m^3
K_{OA}	0.4	gO_2/m^3
b_A	0.2	$1/d$
k_a	0.08	$m^3/gCOD/d$
k_h	3.0	$g\,slowly\,biodegr.COD/(gcellCOD)/d$
K_X	0.03	$g\,slowly\,biodegr.COD/(gcellCOD)/d$
η_h	0.4	

Table B.3: Stoichiometric and kinetic parameters for the TecnoCasic ASM1 model

Appendix C

Nomenclature

Symbols

b_A	Autotrophic decay rate
b_H	Heterotrophic decay rate
$\mathbf{d}$	Disturbances vector
f_P	Fraction of biomass yielding particulate products
i_{XB}	(Mass N)/(Mass COD) in biomass
i_{XP}	(Mass N)/(Mass COD) prod. from biomass
i_vt	VSS/TSS ratio
k_a	Ammonification rate
k_h	Max. specific hydrolisys rate
v_0	Maximum theoretical settling velocity
$v_0^{'}$	Maximum practical settling velocity
r_h	Parameter for hindered settling zone
r_p	Parameter for flocculant settling zone
$\mathbf{u}$	Inputs vector
$\mathbf{x}$	State variables vector
$\mathbf{y}$	Measurements vector

A	Settler cross-sectional area
J	Cost function
K_{NH}	Ammonia half saturation coefficient for autotrophs
K_{NO}	Nitrate half saturation coefficient for heterotrophs
K_{OA}	Oxygen half saturation coefficient for autotrophs
K_{OH}	Oxygen half saturation coefficient for heterotrophs
K_{SH}	Half saturation coefficient for heterotrophs
K_X	Half saturation coeffic. for hydrolysis of slowly biod. substr.
N_m	Dynamic degrees of freedom
N_{opt}	Optimization degrees of freedom
S_{ALK}	Alkalinity
S_I	Soluble inert organic matter
S_O	Dissolved oxygen in ASM1 notation
S_{ND}	Soluble biodegradable organic N
S_{NH}	Free and Ionized Ammonia
S_{NO}	Nitrite and nitrate Nitrogen
V	Bioreactor Volume
$\mathbf{W}_c$	Controllability gramian
$\mathbf{W}_o$	Observability gramian
X_{min}	Minimum attainable suspended solids concentration
X_{BA}	Active autotrophic biomass
X_{BH}	Active heterotrophic biomass
X_I	Particulate inert organic matter
X_{ND}	Particulate biodegrabable organic N
X_P	Particulate products from biomass decay
X_S	Slowly biodegradable substrate
Y_A	Autotrophic Yield
Y_H	Heterotrophic Yield

Greek Letters

γ	Condition number
η_g	Correction factor for growth for heterotrops
η_h	Correction factor for anoxic hydrolysis
μ_A	Autotrophic maximum specific growth rate
μ_H	Heterotrophic maximum specific growth rate
λ	Eigenvalue
σ	Singular value
σ^H	Hankel Singular value

Superscripts

eff	effluent
n	nitrification
r	residualized
ref	reference
p	pre-denitrification
sp	setpoint
ss	steady-state

Subscripts

in	influent
r	recycle
w	wasted

Acronyms

ASM1 Activated Sludge Model No. 1

ASP Activated Sludge Process

BOD Biochemical Oxygen Demand

COD Chemical Oxygen Demand

DO Dissolved Oxygen

DOF Degree Of Freedom

F/M Food to Microorganisms ratio

MLSS Mixed Liquor Suspended Solids

SBH Sludge Blanket Height

SRT Sludge Retention Time

TKN Total Kjeldahl Nitrogen

TN Total Nitrogen

TSS Total Suspended Solids

VSS Volatile Suspended Solids

WWT Wastewater treatment

Bibliography

[1] V. Alstad. *Studies on selection of controlled variables*. PhD thesis, Dep. of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, 2005.

[2] J. Alvarez. Nonlinear state estimation with robust convergence. *Journal of Process Control*, 10:59–71, 2000.

[3] J. Alvarez and T. Lopez. Robust nonlinear state estimation of nonlinear plants. *AIChE Journal*, 45:107–122, 1999.

[4] A. C. Antoulas, D. C. Sorensen, and S. Gugercin. A survey of model reduction methods for large scale systems. In *Structured matrices in operator theory, numerical analysis, control, signals and image processing*. AMS Pubblications, 2001.

[5] E. Ayesa, J. Carstensen, U. Jeppsson, and P. Vanrolleghem. Identification of the dynamic processes in WWTP and control of WWTP. Technical report, COST 682 - Working Group Meeting, 1994. San Sebastian, Spain.

[6] E. Ayesa, G. Oyarbide, L. Larrea, and Garcia-Heras J. L. Observability of reduced order models - Application to model for control of α process. *Water Science and Technology*, 31(2):161–170, 1995.

[7] R. Baratti, A. Bertucco, A. Da Rold, and M. Morbidelli. Develepment of a composition estimator for binary distillation columns. *Chemical Engineering Science*, 50:1541–1550, 1995.

[8] R. Baratti, A. Bertucco, A. Da Rold, and M. Morbidelli. A composition estimator for multicomponent distillation columns-development and experimental tests on ternary mixture. *Chemical Engineering Science*, 53:3601–3612, 1998.

[9] G. Bastin and D. Dochain. *On-line estimation and adaptive control of bioreactors*, volume 1 of *Process Measurements and Control*. Elsevier, 1990.

[10] F. Benazzi, K. V. Gernaey, U. Jeppsson, and R. Katebi. On-line concentration and detection of abnormal substrate concentration in WWTPs using a software sensor: A benchmark study. In *Proceedings of the 2nd Conference on Instrumentation, Control and Automation (ICA2005)*, 20 May -2 June 2005. Busan, Korea.

[11] M. A. Branch and A. Grace. *MATLAB, Optimization Toolbox User's Guide*, 1996.

[12] E. H. Bristol. On a new measure of interaction for multivariable process control. *IEEE Transaction on Automatic Control*, 11:133–134, 1966.

[13] Y. Cao. Controlled variables selection for static self-optimizing control. In *Proceedings of IFAC Symposium on Advanced Control of Chemical Process (ADCHEM)*, pages 63–71, 2004.

[14] J. Cartensen, P. Harremoes, and H. Madsen. Statistical identification of Monod-kinetic parameters from on-line measurements. *Water Science and Technology*, 31(2):125–133, 1995.

[15] J. Cartensen, P. Vanrolleghem, E. Ayesa, U. Jeppsson, A. Arrutikoetexea, and B. Vanderhaegen. Objective function for wastewater treatment design and operation. Technical report, COST 682 - Working Group Meeting, 1994. Bruges, Belgium.

[16] B. Chachuat. *Méthodologie d'optimisation dynamique et de command optimale des petites stations d'épuration à boues activées*. PhD thesis, Istitut National Polytechnique de Lorraine, 2001.

[17] B. Chachuat, N. Roche, and M. A. Latifi. Dynamic optimization of small size wastewater treatment plants including nitrification and denitrification processes. *Computers and Chemical Engineering*, 25:585–593, 2001.

[18] S. H. Cho, F. Colin, M. Sardin, and C. Prost. Settling velocity model of activated sludge. *Water Research*, 27(7):1237–1242, 1993.

[19] A. Conley and M. E. Salgado. Gramian based interaction measure. In *Proceedings of the 39th IEEE Conference on Decision and Control*, pages 5020–5022, 2000. Sydney, Australia.

[20] J. B. Copp. COST action 624 - the COST simulation benckmark: Description and simulator manual. Technical report, European Community, 2000.

[21] E. Davison. A method for simplifying linear dynamic systems. *IEEE Transaction on Automatic Control*, 11:93–101, 1966.

[22] R. A. DeCarlo and R. Saeks. A root locus techniques for interconnected systems. *IEEE Transaction on Systems Man and Cybernetics*, SMC-9(1):53, 1979.

[23] S. Diehl and U. Jeppsson. A model of the settler coupled to the biological reactor. *Water Research*, 32(2):331–342, 1998.

[24] D. Dochain and P. Vanrolleghem. *Dynamical modelling and estimation in wastewater treatment process.* IWA Publishing, 2001.

[25] P. L. Dold, G. A. Ekama, and G. V. R. Marais. A general model for the activated sludge process. *Prog. Water Technology*, 12:47–77, 1980.

[26] G. A. Ekama, J. L. Barnard, Gunthert F. W., P. Krebs, J. A. McConquodale, D. S. Parker, and E. J. Wahalberg. Secondary settling tanks: theory, modelling, design and operation. Technical Report 6, IWA Publishing, 1997.

[27] H. K. Engelien. *Process integration applied to the design and generation of distillation columns.* PhD thesis, Dep. of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, 2004.

[28] D. Enns. Model reduction with balanced realization: an error bound and a frequency weighted generalization. In *Proceeding of the 23rd IEEE Conferences on Decision and Control*, pages 127–132, 1984.

[29] K. V. Fernando and H. Nicholson. Singular perturbational model reduction of balanced systems. *IEEE Transaction on Automatic Control*, 27:466–468, 1982.

[30] J. P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems. application to bioreactors. *IEEE TAC*, 37(6):875, 1992.

[31] S. Gillot, D. De Clercq, D. Defour, F. Simoens, K. Gernaey, and P. Vanrolleghem. Optimization of a wastewater treatment plant design and operation using simulation and cost analysis. In *Proceeding of 72nd Annual WEF Conference and Exposition*, 1999. New Orleans, USA.

[32] B. Glemmestad, S. Skogestad, and T. Gundersen. Optimal operation of heat exchange networks. *Computer and Chemical Engineering*, 23:509–522, 1999.

[33] K. Glover. All optimal Hankel norm approximation of linear multivariable systems and their L_∞-error bounds. *International Journal of Control*, 39:1115–1193, 1984.

[34] C. Gomez-Quintero, I. Queinnec, and J. P. Babary. A reduced nonlinear model of an activated sludge process. In *Proceeding of the International Symposium on Advanced Control on Chemical Processes (ADCHEM)*, June 14-16 2000. Pisa, Italy.

[35] M. S. Govatsmark. *Integrated optimization and control*. PhD thesis, Dep. of Chemical Engineering, Norwegian University of Science and Technology, 2003.

[36] W. Gujer, M. Henze, T. Mino, and M. C. M. van Loosdrecht. Activated Sludge Process Model No. 3. *Water Science and Technology*, 39(1):183–193, 1999.

[37] J. Hahn and T. F. Edgar. An improved method for nonlinear model reduction using balancing of empirical gramians. *Computer and Chemical Engineering*, 26(10):1379–1397, 2002.

[38] J. Hahn, T. F. Edgar, and W. Marquardt. Controllability and observability covariance matrices for the analysis and order reduction of stable nonlinear systems. *Journal of Process Control*, 13(2):115–127, 2003.

[39] I. J. Halvorsen. *Minimum Energy requirements in complex distillation arrangement*. PhD thesis, Dep. of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, 2001.

[40] I. J. Halvorsen, S. Skogestad, J. C. Morud, and V. Alstad. Optimal selection of controlled variables. *Industrial Engineering and Chemical Research*, 42(14):3273–3284, 2003.

[41] M. Henze, W. Gujer, T. Mino, G. V. R. Marais, and T. Matsuo. Activated Sludge Process Model No. 2. Scientific and Technical Report 3, IAWQ, 1995. London, UK.

[42] M. Henze, W. Gujer, T. Mino, T. Matsuo, M. C. Wetzel, G. V. R. Marais, and M. C. M. van Loosdrecht. Activated Sludge Process Model No. 2d. *Water Science and Technology*, 39(1):165–182, 1999.

[43] M. Henze, C. P. Leslie Grady, W. Gujer, G. V. R. Maris, and T. Matsuo. Activated Sludge Process Model No. 1. Scientific and Technical Report 1, IAWQ, 1987. London, UK.

[44] L. N. Hopkins, P. A. Lant, and B. Newell. Operating space diagrams: a toll for design of wastewater treatment plant. *Water Science and Technology*, 44(2-3):69–76, 2001.

[45] Hydromantis. *GPS-X Technical Reference*, 2001.

[46] Meltcalf & Eddy Inc. *Wastewater Engineering: treatment, disposal and reuse*. McGraw Hill, 1991.

[47] P. Ingildsen. *Realising full-scale control in wastewater treatment system using in situ nutrient sensors*. PhD thesis, Dep. of Industrial Electrical Engineering and Automation, Lund University, 2002.

[48] S. Isaacs. Short horizont control strategies for automating activated sludge process. *Water Science and Technology*, 34(1-2):203–212, 1996.

[49] S Isaacs and D. E. Thormberg. A comparison between model and rule based control of a periodic activated sludge process. *Water Science and Technology*, 37(12):343–352, 1998.

[50] A. H. Jazwinski. *Stocastic processes and filtering theory*. Academic Press, New York, 1970.

[51] U. Jeppsson. On the verifiability of the activated sludge dynamics. Master's thesis, IEA, Lund Institute of Technology, Lund, Sweden, 1993.

[52] U. Jeppsson. *Modelling aspects of wastewater treatment processes*. PhD thesis, IEA, Lund Institute of Technology, Lund, Sweden, 1996.

[53] U. Jeppsson and S. Diehl. An evaluation of the dynamic model of the secondary settler. *Water Science and Techonology*, 34(5-6):19–26, 1996.

[54] U. Jeppsson and S. Diehl. On the modelling of the dynamic propagation of biological components in the secondary clarifier. *Water Science and Technology*, 34(5-6):85–92, 1996.

[55] U. Jeppsson and G. Olsson. Reduced order models for on-line parameter identification of the activated sludge process. *Water Science and Technology*, 28(11-12):173–183, 1993.

[56] S. Julien, J. P. Babary, and J. P. Lessard. Theoretical and practical identifiability of a reduced order model in an activated sludge process doing nitrification and denitrification. *Water Science and Technology*, 37(12):309–316, 1998.

[57] S. Julien, P. Lessard, and J. P. Babary. A reduced order model for control of a single reactor activated sludge process. *Mathematical and Computer Modelling of dynamical Systems*, 5(4):337–350, 1999.

[58] J. C. Kabouris and A. P. Georgakakos. Accounting for different time scale in activated sludge process control. *Water Science and Technology*, 26(5-6):1381–1390, 1992.

[59] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[60] K. J. Keesman, H. Spanjers, and G. van Straten. Analysis of endogenous process behaviour. *Water Science and Technology*, 37(12):227–235, 1998.

[61] P. V. Kokotovic, R. E. O'Malley, and P. Sannuti. Singular perturbation and order reduction in control theory - an overview. *Automatica*, 12:123–132, 1976.

[62] A. Krener and A. Isidori. Linearization by output inijection and nonlinear observers. *Systems and Control Letters*, 3:47, 1987.

[63] S. Lall, J. E. Marsden, and S. Glavaski. Empirical model reduction of controlled nonlinear systems. In *Proceedings 14th IFAC Word Congress*, 1999.

[64] S. Lall, J. E. Marsden, and S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal of Robust and Nonlinear Control*, 12:519–535, 2002.

[65] T. Larsson, M. S. Govatsmark, and S. Skogestad. Control structure selection for reactor, separator and recycle processes. *Industrial and Engineerig Chemistry Research*, 42:1225–1234, 2003.

[66] T. Larsson, K. Hestetun, E. Hovland, and S. Skogestad. Self-optimizing control of a large-scale plant: the Tennessee Eastman process. *Industrial and Engineering Chemistry Research*, 40:4889–4901, 2001.

[67] T. T. Lee, F. Y. Wang, and R. B. Newell. Robust model-order reduction of complex biological processes. *Journal of Process Control*, 12(7):807–821, 2002.

[68] C. F. Lindberg. *Control and estimation strategies applied to the activated sludge process*. PhD thesis, System and Control Group, Uppsala University, Sweden, 1997.

[69] L. Ljung and T. Soderstrom. *Theory and practice of recursive identification*. MIT Press, Cambrige, USA, 1983.

[70] T. Lopez. *Estimacion y control no linear de reactores de copolimerizacion*. PhD thesis, Universidad Autonoma Metropolitana, Mexico, 2000.

[71] T. Lopez, A. Pulis, M. Mulas, and R. Baratti. A software sensor for a wastewater treatment plant. In *Proceeding of the International Symposium on Advanced Control on Chemical Processes (ADCHEM)*, 11-14 January 2000. Hong Kong.

[72] T. Lopez, S. Tronci, R. Baratti, and J. Alvarez. State estimation in a catalytic reactor via a recostructive approach. In *Proceedings of IFAC 15th Word Congress*, 2002. Barcelona, Spain.

[73] S. Marsili-Libelli. Modelling, identification and control of the activeted sludge process. In *Advances in Biochemical Engineering/Biotechnology*, volume 38, pages 90–148. Springer-Verlang, Berlin, 1989.

[74] P. S. Maybeck. *Stochastics models, estimation and control*, chapter 1. Academic Press, Inc., 1979.

[75] T. J. McAvoy, Y. Arkun, R. Chen, D. Robison, and P. D. Schnell. A new approach to defining a dynamic relative gain. *Control Engineering Practice*, 11(8):907–914, 2003.

[76] J. J. Monge and C. Georgakis. The effect of operating variables on the dynamics of catalytic cracking processes. *Chemical Engineering Comm.*, 60:1–26, 1987.

[77] B. C. Moore. Principal component analysis in linear system: controllability, observability and model reduction. *IEEE Trans. Automat. Control*, 26:17–31, 1981.

[78] M. Morari, Y. Arkun, and G. Stephanopoulos. Studies in the synthesis of control structures for chemical processes, Part I: Formulation of the problem, process decomposition and classification of the control task, analysis of the optimizing control structure. *AIChE Journal*, 26(2):220–232, 1980.

[79] M. Mulas, A. M. Pulis, and S. Tronci. State estimation issues on activated sludge process. In *Proceeding of CHISA2004, 16th International Congress and Process Engineering*, 2004. Prague.

[80] M. Mulas and S. Skogestad. Control structure analysis for an activated sludge process. In *Proceedings of ICheap7, Chemical Engineering Transaction*, pages 173–178, 2005.

[81] M. Mulas and S. Skogestad. Control structure analysis for an activated sludge process. In *AIDIC conference series*. Reed Business copyright, 2005.

[82] M. Mulas, S. Tronci, T. Lopez, and R. Baratti. State estimation of a biological reactor: comparison between an EKF and a NGO. In *Proceeding of ECCE2003, European Conference on Chemical Engineering*, 2003. Granada, Spain.

[83] E. V. Munch, P. Land, and J. Keller. Simultaneous nitrification and denitrification in bench-scale sequencing batch reactors. *Water Research*, 30(2):277–284, 1996.

[84] A. Nierderleski. A heuristic approach to the design of linear multivariable interacting control system. *Automatica*, 7:691–701, 1971.

[85] B. A. Ogunnaike and W. H. Ray. *Process dynamics, modeling and control*. Oxford University Press, 1994.

[86] G. Olsson and B. Newell. *Wastewater treatment systems - Modelling, diagnosis and control*. IWA Publishing, 2002.

[87] G. Olsson, M. K. Nielsen, Z. Yuan, A. Lyngaard-Jensen, and J. P. Steyer. Instrumentation, control and automation in wastewater systems. Technical Report 15, IWA Publishing, 2005.

[88] W. H. Ray. *Advanced Process Control*. McGraw-Hill, 1981.

[89] G. A. Robertson. *Mathematical modelling of statup and shutdown operations of process plants*. PhD thesis, The University of Queensland, Australia, 1992.

[90] G. A. Robertson and I. T. Cameron. Analysis of dynamic process models for structural analysis insight and model reduction - Part 1. structural identification measures. *Computers and Chemical Engineering*, 21(5):455–473, 1997.

[91] P. Samuelsson. *Control of Nitrogen Removal in Activated Sludge Process*. PhD thesis, Uppsala Dissertation from the Faculty of Science and Technology, Uppsala, Sweden, 2005.

[92] P. Samuelsson, B. Halvarsson, and B. Carlsson. Cost-efficiency operation of a denitrifying activated sludge process - an initial study. Technical report, Div. of System and Control, Dep. of Information Technology, Uppsala, Sweden, 2005.

[93] J. M. A. Scherpen. Balancing of nonlinear systems. In *Proceedings of the European Control Conference*, pages 1838–1843, 1993.

[94] H. Schmidt and E. W. Jacobsen. Selecting control configuration for performance with independent design. *Computers and Chemical Engineering*, 27(1):101–109, 2003.

[95] H. Siegrist, P. Krebs, R. Bushler, I. Purtshert, C. Rock, and H. Rufer. Denitrification in secondary clarifiers. *Water Science and Technology*, 31(2):205–214, 1995.

[96] H. Siegrist and M. Tschui. Interpretation of experimental data with regard to the activated sludge model No.1 and calibration of the model for municipal wastewater treatment plants. *Water Science and Technology*, 25(6):167–183, 1992.

[97] A. K. Singh and J. Hahn. On the use of empirical Gramians for controllability and observability analysis. In *Proceedings of the American Control Conference*, June 8-10 2005. Portland, USA.

[98] S. Skogestad. Plantwide control: the search for the self-optimizing control structure. *Journal of Process Control*, 10:487–507, 2000.

[99] S. Skogestad. Control structure design for complete chemical plants. *Compurters and Chemical Engineering*, 28:219–234, 2004.

[100] S. Skogestad and M. Morari. Implication of a large RGA elements on control performance. *Industrial Engineering in Chemical Research*, 26:2323–2330, 1987.

[101] S. Skogestad and I. Postlehwait. *Multivariable feedback control*. John Wiley Sons, New York, 1996.

[102] J. J. E. Slotine, J. K. Hedrick, and E. A. Misawa. On sliding observers for nonlinear systems. *Journal of Dynamic Systems, Measurements and Control*, 109:245, 1987.

[103] I. Y. Smets, V. J. Haegebaert, R. Carrette, and J. F. Van Impe. Linearization of the activated sludge model ASM1 for fast and reliable prediction. *Water Reseach*, 37(8):1831–1851, 2003.

[104] O. A. Z Sotomayor, S. Won Park, and C. Garcia. Software sensor for on-line estimation of the microbial activity in activated sludge system. *ISA Transaction*, 41(2):127–143, 2002.

[105] M. A. Steffens, P. A. Lant, and R. B. Newell. A systematic approach for reducing complex biological wastewater treatment models. *Water Research*, 31(3):590–606, 1997.

[106] G. Stephanopoulos. *Chemical Process Control*. Prentice Hall, Englewood Cliff, New Jersey, 1984.

[107] I. Takacs, G. G. Patry, and D. Nolasco. A dynamic model of the clarification-tickening process. *Water Reasearch*, 25(10):1263–1271, 1991.

[108] M. S. Tombs and I. Poslethwaite. Truncated balanced realization of stable, non-minimal state-space systems. *International Journal of Control*, 46:1319–1330, 1987.

[109] S. Tronci, F. Bezzo, M. Barolo, and R. Baratti. Geometric observer for a distillation column: development and experimental testing. *Ind. Eng. Chem. Res.*, 44:9884–9893, 2005.

[110] S. Tronci, M. Mulas, T. Lopez, and R. Baratti. State estimation of a biological reactor. In *Proceedings of the 6th Italian Conference on Chemical and Process Engineering*, 8-11 June 2003. Pisa, Italy.

[111] V. Van Breusegem and G. Bastin. Reduced order dynamical modelling of a reaction system: a singular perturbation approach. In *Proceedings of the 30th Conference on Decision and Control*, 1991. Brighton, England.

[112] P. Vanrolleghem. Building blocks for wastewater treatment process control: a review. In *Advanced course on Environmental Biotechnology*, May 25-June 3 1994. Delf, The Netherlands.

[113] P. Vanrolleghem and S. Gillot. Robustness and economic measure as control benckmark performance criteria. *Water Science and Technology*, 45(4-5):117–126, 2002.

[114] Z. Z. Vitasovic. *An integrated control strategy for the activated sludge process*. PhD thesis, Rice University, 1986. Houston, Texas.

[115] S. Weijers. *Modelling, identification and control of activated sludge plants for Nitrogen removal*. PhD thesis, Technische Universiteit Eindhoven, The Netherlands, 2000.

[116] M. F. Witcher and T. J. McAvoy. Interacting control systems: steady-state and dynamic measurements of interaction. *ISA Transaction*, 16(3):35–44, 1977.

[117] B. Wittenmark and M. E. Salgado. Hankel-norm based interaction measure for input-output pairing. In *Proceedings of 15th IFAC World Conference*, 2002. Barcelona, Spain.

[118] O. Yamanaka, T. Obara, and K. Yamamoto. Total cost minimization control scheme for biological wastewater treatment process and its evaluation based on COST Benchmark process. In *Proceedings of the 2nd IWA ICA Conference*, 29 May - 2 June 2005. Busan, South Korea.

[119] H. Zhao, S. H. Isaacs, H. Soeberg, and M. Kummel. An analysis of nitrogen removal and control strategies in an alternating activated sludge process. *Water Research*, 29(2):535–544, 1995.

[120] Z. X. Zhu. Variable pairing selection on individual and overall interaction measures. *Industrial Engineering in Chemical Research*, 35:4091–4099, 1997.