Data-Driven Control Design of Wastewater Treatment Systems

Alberto Sánchez

Industrial Control Centre
Department of Electronic and Electrical Engineering
University of Strathclyde

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy

14th June 2004

Declaration of Author's Rights

Abstract

The research in this thesis covers three fields of control theory; identification, control design, and real-time control applied to activated sludge wastewater treatment plants. The study is carried out using simulation and a full-scale implementation in Swinstie wastewater treatment plant from Scottish Water.

Subspace algorithms are explored to obtain adequate models for dissolved oxygen and nutrient dynamics. Results presented in this area are the outcome of a number of simulations and full-scale plant experiments, which have lead to the formulation of standard recommendations to the identification of linear models for the activated sludge process. Part of the work has also provided an evaluation of a number of subspace identification algorithms, although this has not been an objective within the thesis. The thesis also contains some insight into the modelling of the activated sludge for an intermittent aeration process.

The control design part of the thesis employs a two level hierarchical control approach. The low level control is usually a proportional integral derivative controller (PID) type. This thesis presents the development of three new tuning algorithms for PID type controllers: iterative feedback tuning (IFT), linear quadratic gaussian (LQG) and data-driven. The first two methods are developed for continuous time systems, while the last is a discrete time data-driven method which uses subspace identification. The second control level employs linear model predictive control (MPC). MPC is used for dissolved oxygen and nitrogen removal in a simulation level. Linear models of nutrient removal obtained by identification and by model reduction are used to implement controllers for continuous aeration and intermittent aeration plants.

Real-time control is implemented by developing a software platform. The software platform contains algorithms for subspace identification, MPC control design and execution and process monitoring. The software is developed using LabVIEW and MATLAB. The user frontend and the communication with the PLC are implemented in Lab-

VIEW. The PLC communication employs OPC tehcnology. Many of the algorithms required for identification, control design, and process monitoring are programmed in MATLAB and linked to LabVIEW by using several technologies as: Activex and Dynamic Link Libraries (DLL). The thesis finally presents results obtained by real-time execution of the identification and control algorithms.

Dedication

To my parents José and Ivette; my brothers Pepe, Daniel, and Andrés and finally in loving memory of Llallo, my cousin.

I also owe much gratitude to the COST action 624 and 682 workgroup for the invitations to join their workgroup, in particular to Dr. Ulf Jeppsson and Dr. Jens Alex for letting me use their MATLAB\SIMULINK and SIMBA implementations of the simulation benchmark.

To all the staff, past and present, at Scottish Water I would like to thank for their support. In particular, thanks are given to Johnny Tyler for sharing his 'practical knowledge' about the process, and to Linda Brady and Glen Dickson for co-ordinating the activities within the company and the university.

A recognition is due to Paul Reed and Arantxa Gomez-Arnau; and finally but undoubtfully not the least, my thanks are given to God, for letting me live every single day of my life.

Alberto Sánchez

Glasgow, March 2004

Acknowledgements

I will like to express my most profound gratitude to Dr. Reza Katebi, my supervisor, for his continuous encouragement, enlightment and support with his knowledge and experience during the last three years. The same thanks are also given to Prof. Michael Johnson, my co-supervisor, whose experience and meticulous work have taught me how to do research. I appreciate as well the high standards imposed by both of my supervisors, which have raised the level of my expectations.

I will also like to thank my parents for their continuous and unconditional advice and support in matters beyond engineering which have helped me to grow not only intellectually but spiritually.

Special thanks are also due to Matthew Wade, my project colleague, with whom discussions and mutual learning became a day to day routine. Dr. Leonardo Giovanini, whose knowledge and experience always opened my mind to different ways of looking at things. To Giuseppe and Giulliana Conti, whose friendship was of critical importance during the hardest days, and from whom I learned about traditional Italian cuisine besides pasta and pizza.

Thanks are also due to all my friends and colleagues who have worked in the Industrial Control Centre over the past few years, Massimo Calligaris, Peter Martin, Mohammed Moradi, Helen Marku, Mary Gillie, Sergio Dominguez, Marie O'Brien, Farid Benazzi, Flavia Camilleri, Arkadiusz Dutka, Jonas Balderud, Drew Smith, John Russell and many others whose name I may have missed, but nevertheless have given me their friendship.

A special recognition is due to all the SMAC project people, with whom the continuous exchange of knowledge has significantly contributed to the content of this thesis, and the European Commission under whose contract EVK1-CT-2000-00056 the SMAC project and this work has been performed.

Nomenclature

ASM	Activated sludge model
CVA	Canonical variate analysis
DDWF	Daily dry weather flow
DO	Dissolved oxygen
DLL	Dynamic link library
ERAS	External return activated sludge
IFT	Iterative feedback tuning
LQG	Linear quadratic gaussian
MIMO	Multiple-input, multiple- output
MLSS	Mixed liquor suspended solids
MOESP	Multivariable output-error state space
MPC	Model predictive control
N4SID	Numerical algorithms for subspace state-space identification
OPC	OLE for process control
PID	Proportional intergral derivative
PLC	Programmable logic controller
PRBS	Pseudo random binary signal
RAS	Return activated sludge
SCADA	Supervisory control and data acquisition
SISO	Single-input, single-output
SS	Suspended solids
STAR	Superior tuning and reporting
SV	Singular value
svd	Singular value decomposition
vaf	Variance accounted for
WWTP	Wastewater Treatment Plant

Contents

Chapter 1

Introduction

The construction and operation of water drainage and wastewater treatment has historically served two main purposes: (a) to avoid flooding and (b) to maintain public hygiene. It is just recently that pollution has became important due to environmental awareness. Therefore, the treatment of wastewater before discharging is a necessity as a way of reducing pollutant loading into biological active water bodies. Since then, the science and engineering of wastewater treatment has evolved considerably with different technologies and methods developed over the years.

The activated sludge treatment process is probably one of the most common processes employed for urban sewage. The treatment process consists in the bio-degradation of organic material, and bio-chemical decomposition of nutrients. The biochemical processes involved are complex and until recently the mechanics involved have not been clearly defined. It is not until the late 80s, with the work of Henze *et al.* (1987) that a more scientific understanding has been achieved by the development of mathematical models.

Modelling of the activated sludge process has not only provided a wider understanding; but has also provided scientists and engineers with a powerful tool which can be used to optimise and even predict the behaviour of the system under certain conditions. Although these models represent the state of the art in the mathematical understanding

of the treatment process, their applicability is limited due to assumptions of ideal conditions, which are not achievable in practice. On the other hand, the development of sophisticated mathematical tools, such as identification and advanced process control, can provide a wider set of tools for possible process improvement which are not necessarily dependent on a detailed process knowledge, as in the activated sludge model (ASM) case.

From the point of view of the wastewater industry, process optimisation, and therefore the efficient control of the treatment process, has become a necessity with the implementation of more stringent environmental regulations around the globe. Operational costs of treatment plants has also contributed to increase investment in the implementation of advanced control technologies. In some countries, governmental policies have lead to a link between environmental concerns and responsibility in the efficient treatment of sewage. Denmark, for example, charges wastewater companies a tax over the amount of pollutants discharged. Therefore, treatment plants are also required to monitor their process so pollutant loads can be accurately estimated.

On the other hand, the multivariable nature and complexity of the process presents a challenge for process control engineers and scientists. The use of advanced process control algorithms like model predictive control (MPC) combined with identification are alternatives which have been successfully exploited in other industries like the petrochemical, and whose application in the wastewater industry has been very limited.

Due to the nonlinearity of the process, and continuous changing conditions, it might be necessary also to have well proven and understood control laws which can be easily readjusted. Even more, the use of historical data, which is usually recorded and almost never used, might provide sufficient information to adjust controllers in an optimal way.

In summary, the wastewater industry has been led into a process of change to improve its operation on two fronts: the environmental and the economical, for which the use of

advanced process control can be beneficial. This thesis explores the use of identification, model predictive control and controller tuning applied to the wastewater industry. The work presented in the thesis conveys the results of three years of research in theoretical and real-time control design. The two 'faces' of this thesis are complementary, and in many cases theoretical results and experience gained by simulation have been corroborated in practice.

All the research in this thesis has been developed within a bigger project framework called the SMArt Control of Wastewater Treatment Systems (SMAC), which is summarised in the following section.

1.1 Project summary

The SMAC project is European Commission funded project under the Fifth Framework for research, technological development and demonstration activities. The SMAC project aims at expanding the control functions to become a smart and all-embracing control as presented in Figure (1.1) .

1.1.1 Objectives

Wastewater systems, meaning sewer network and wastewater treatment plants, are subject to large fluctuations in flow and concentrations of the wastewater. During stormwater situations large amounts of pollution are diverted untreated to the receiving waters and sudden changes in load deteriorate the removal of nitrogen and phosphorus.

Biological wastewater treatment relies on micro-organisms. New knowledge on the potentials and limitations of these still needs to be put into action in the wastewater systems. Also, most systems are designed for peaks, thus a spare capacity is available in normal conditions, which is frequently not exploited.

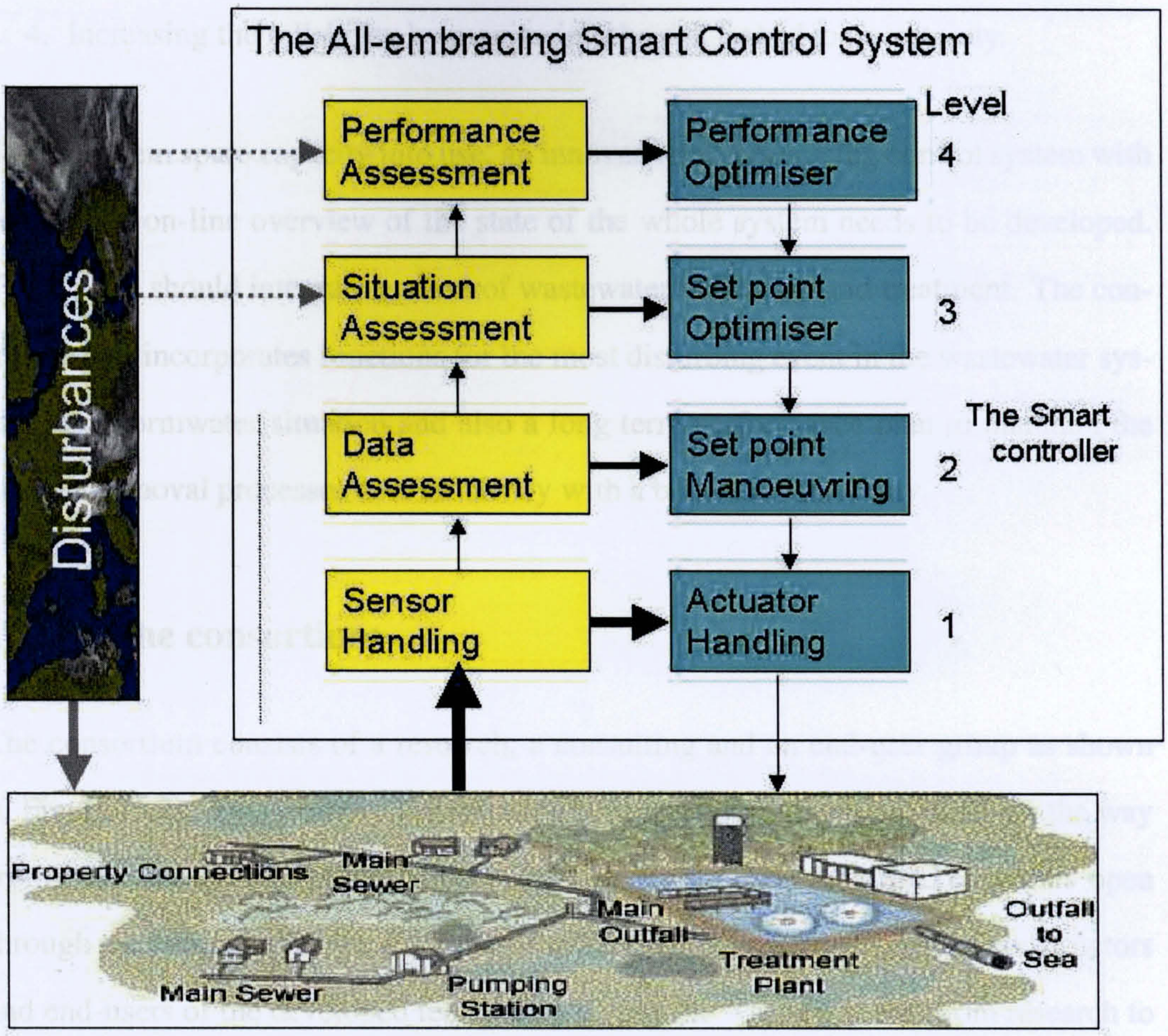

Figure 1.1: SMAC all-embracing control.

Today the control of wastewater systems is performed in local sub-optimal units. Due to the SISO control structures, the results of one control action affect other loops and there is usually no mechanism to counteract this interaction. Also, the sewer system and the treatment plant maintenance and control is normally not co-ordinated. Therefore, the overall objective of the project is to optimise the wastewater system operation by

1. Maximising the system capacity and availability dynamically.
2. Reducing the operational costs by improving the control system.
3. Minimising the pollutant load to the receiving waters.

4. Increasing the reliability by monitoring the risk linked to uncertainty.

To put system spare capacity into use, an innovative all-embracing control system with continuous on-line overview of the state of the whole system needs to be developed. The system should integrate control of wastewater collection and treatment. The control system incorporates functions for the most disturbing event in the wastewater system, the stormwater situation and also a long term performance plan to maintain the nutrient removal processes cost efficiency with a better sustainability.

1.1.2 The consortium

The consortium consists of a research, a consulting and an end-user group as shown in Figure (1.2). This primary organisation is to ensure an 'open pipeline' all the way from basic research to exploitation. The way from research to market is thus open through the consortium. The expertise of different universities, designers, constructors and end-users of the developed technology covers the 'vertical scale' from research to demonstration and commercialisation in this project.

1.1.3 Project description

The first part of the project is to describe the needs of the end-users seen in the light of possible control actions in the wastewater system. Wastewater systems have to be defined according to their design, the type of wastewater they handle, the disturbances and their dynamic behaviour under dry and wet weather and other disturbing conditions. The control systems and variables available or currently implemented have also to be identified, along with the constraints and limitations relative to the available technology and instrumentation.

The second part is to scrutinise the whole data management procedure. The optimal distribution of intelligence which is a part of 'smart control' assures a constant

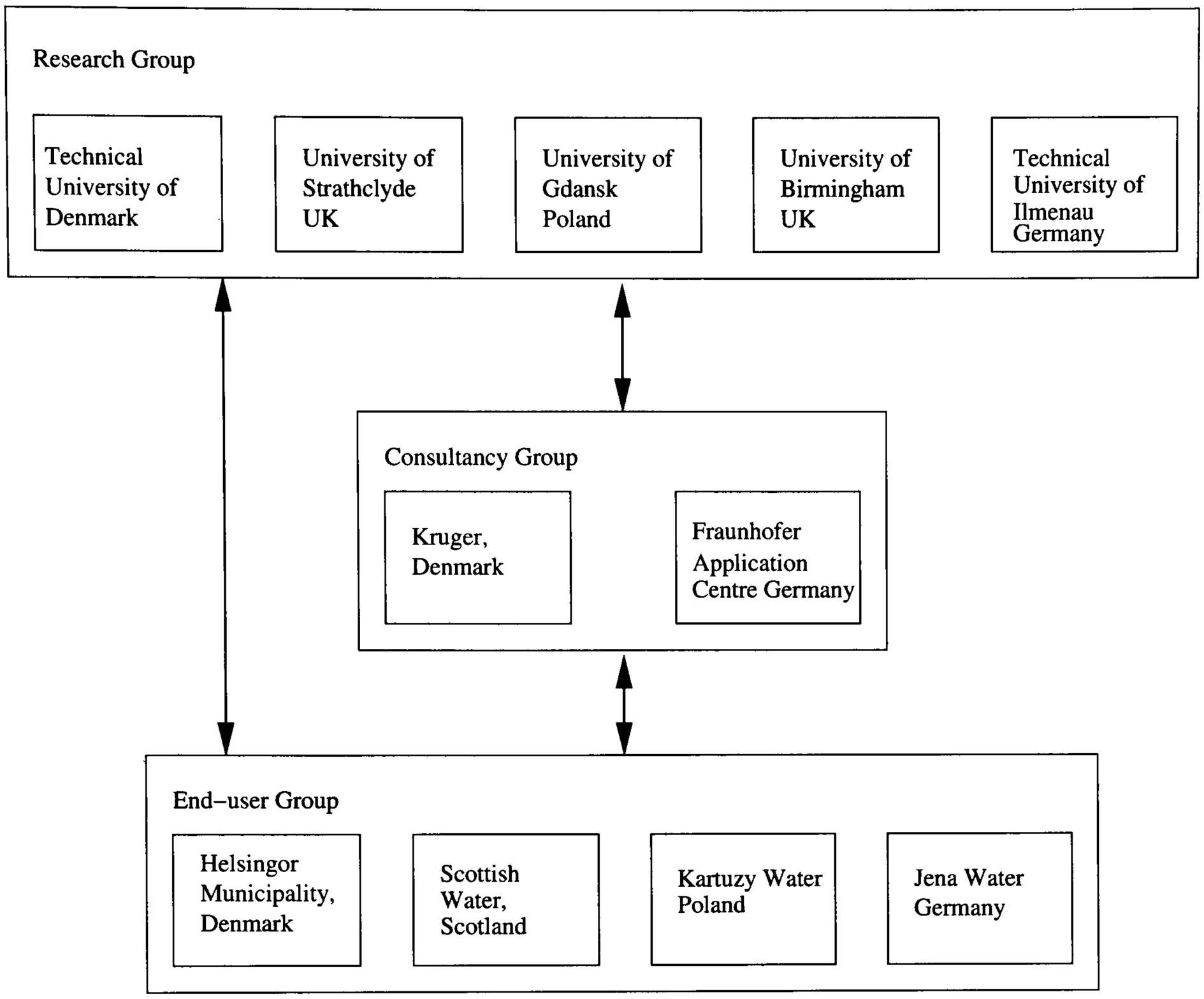

Figure 1.2: SMAC consortium

data quality control. Multivariable cross check will be a new approach to assure that measurements are valid for control purposes. The location of sensors in relation to information dynamics and delays of the signal are also of importance.

The third part concerns the development of the controls. The overall state assessment is the basis of an overall performance plan deciding which operational situation will be in focus for the forthcoming minutes and hours. In dry weather situations a long term plan will assure the best conditions for the micro-organisms development and economic operation. The development of prediction and preparation algorithms for the stormwater situation, both in sewers and at the plant, are also an important part for the implementation of disturbance rejection mechanisms.

In order to implement such a controller, a hierarchical control structure has been devised by the SMAC consortium. Due to the different time scales at which processes

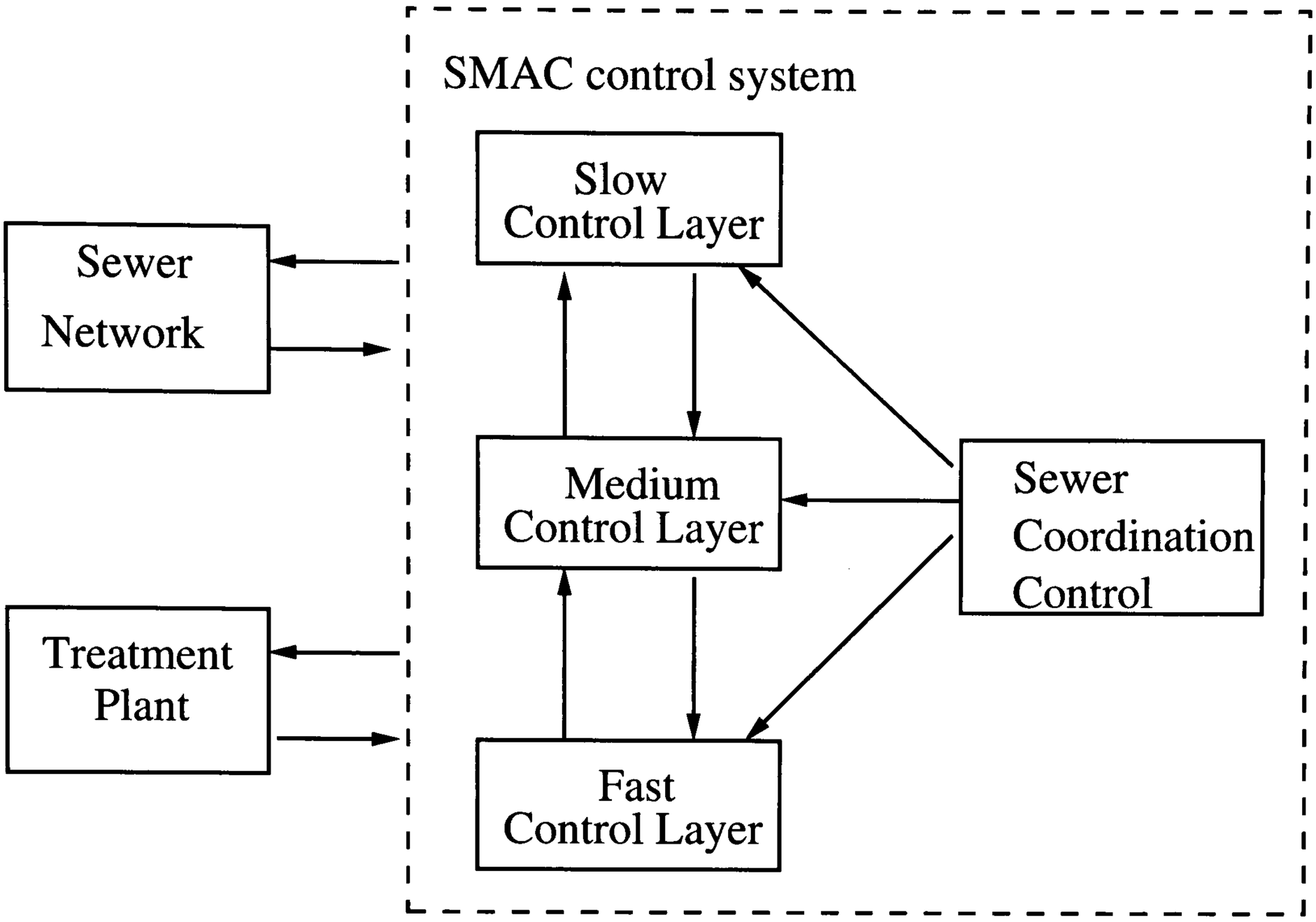

Figure 1.3: SMAC control structure

and events occur, it is sensible to look at the entire system in three time frames: slow, medium and fast. The slow control layer has the purpose of maintaining the plant sustainability by assuring long-term sludge inventories in a weeks to months time scale. The medium control layer's purpose is to optimise the pollutant removal by calculating optimal setpoints and has a time scale of several hours to days. Finally, the fast control layer main functions is to ensure effective plant operation by assessing the low-level control loops and trying to achieve the desired setpoints in an efficient way; this has a time frame of minutes to hours. Figure (1.3) presents an schematic diagram of the SMAC control system. The work presented in this thesis has been primarly developed for the *Fast Control Layer*; however, nutrient removal has also been the subject of a limited study.

The fourth part is the implementation of the results in four end-user wastewater systems. In Denmark the integrated control of wastewater collection and treatment during stormwater situations will be the focus. In Poland co-ordinated control will be de-

veloped to assure the quality of the discharge of treated wastewater before drinking water intakes to a large city. In Germany the integrated management of sewer system and wastewater treatment plant including optimisation of energy and chemicals will be investigated. Finally, in Scotland the efficient control of dissolved oxygen levels and on-line process monitoring will be the focus of implementation.

1.2 Motivation of the thesis research

As described in the previous sections, the research contained in this thesis is part of a more complete and all-embracing control architecture whose major objectives have been enumerated in section 1.1.1. The main motivation for the work developed in this thesis is synthesised in the following statements,

1. The use of subspace identification methods appears as good alternative to deterministic models. Deterministic models of the activated sludge process are only valid under very specific conditions, which are usually very difficult to achieve. In addition, they contain a high degree of uncertainty and unidentifiable terms, which can only be approximated by laboratory experiments. The calibration of these models can take a long period of time, which in practice is not useful, since the wastewater composition can change on a day to day basis. Wastewater treatment is a multivariable process, in which different sub-processes have different time scales, therefore the use of a multivariable identification technique is a straightforward choice. Also, the existence of large amounts of historical data in supervisory control and data acquisition (SCADA) databases can be exploited, as means to provide the sufficient data for the identification routines under certain circumstances. In this context subspace identification methods appear to be an interesting choice.

2. Due to the multivariable nature of the process, it will be beneficial to investigate

the possible improvements of adopting a multivariable control technique. If it is possible to obtain simple models for the process, then these can be used in a model predictive control structure. Different processes operate in different time scales, therefore the use of a hierarchical predictive control structure seems to be a good approach to pursue.

3. In most cases it is impossible or expensive to replace a control algorithm which has been programmed in a programmable logic controller (PLC). Therefore, the use of a second level of control implemented on a more computationally efficient machine, can increase the performance without sacrificing system integrity, security and reducing implementation costs. This type of architecture will also allow operators to have a plant wide perspective and monitor the complex interactions within the treatment process employing advanced multivariate statistical tools like principal component analysis (PCA).

4. The operation of many low level control loops, usually proportional-integral-derivative (PID) type, can significantly benefit from a simple re-tuning, when the plant is operating in a different condition. Moreover, a significat improvement in the higher control levels can only be achieved if the low level control systems are operating in an efficient and healthy way. The paradigm of tuning has been a subject of research for many years, and many methods have been developed. Within the work presented in this thesis three new tuning algorithms for PID-type controllers have been developed. Depending on the necessity and the application, these algorithms allow the calculation of the parameters either by a model-free approach, employing subspace identification, or with an optimal approach. Two of these methods have been extended to multivariable systems, therefore covering a wider spectrum of problems and achieving one of the main objectives of the project.

5. Another, important issue which has motivated some of the research presented in

this thesis, is the assessment of control loops. One of the most accepted methods to assess the performance of control loops, is by comparison with a theoretically optimal (or suboptimal depending the case), benchmark. Benchmarking can be used for many purposes. In particular, it could give indication of when is it necessary to re-tune controllers. Benchmarking is a complex issue for which much research has been performed. However, much of it has been left as an off-line approach; with very few exceptions. This thesis presents some initial work in this area applied to SISO continuous-time control structures, which can be used in an online approach.

6. Much of the work in the literature refering to the use of advanced process control methods has been left in simulation. One of the main aims of this thesis is to bridge the gap between theory and practice by presenting experiences in the implementation, design and operation of such controllers in real-time. The implementation of such controllers on an industrial scale conveys the use of sophisticated communication protocols and robust hardware and software architectures, within other issues, to maintain reliability in the operation of the plant. This thesis explores the implementation, design and operation of advanced control systems in real-time in a full-scale wastewater treatment plant (WWTP).

1.3 Outline of the thesis

The thesis is composed of eight chapters covering identification, control design, tuning and real-time implementation. The thesis is organised in the following way:

Chapter 2 provides a brief introduction to the activated sludge treatment process by describing the unit operations and their functions. Then, the chapter includes a description of the WWTP simulation plant employed including the modifications and assumptions used in the thesis. In addition, the chapter provides a description of the

two full scale WWTPs where the implementation of the control and analysis of data has been performed. A summary of the topics covered and developments presented in the chapter are included at the end.

Chapter 3, covers models and identification of the activated sludge wastewater. Two objectives drive the content of this chapter: (a) finding appropriate models for dissolved oxygen and nitrogen removal (nutrients) which can be later used in a model predictive control structure, and (b) exploring the suitability of different subspace algorithms for the identification of the activated sludge process. For the case of dissolved oxygen, only the use of subspace identification is explored; however, for the nutrients case, deterministic modelling and subspace identification are employed. The deterministic model has been developed only for the special case of an alternating wastewater treatment plant (intermittent aeration). The chapter also contains a section on identification of dissolved oxygen and nutrients with *a posteriori* analysis of data from the Helsingor WWTP (Denmark). This study has yielded interesting results which corroborate with the simulation results from the previous sections. The chapter ends with a brief summary of the chapter content and its results.

Chapter 4, concerns the design of model predictive controllers using the models developed in the previous chapter. The content of the chapter begins with a brief review of a standard formulation of model predictive control, which is used throughout the chapter. Later, the design of model predictive controllers for dissolved oxygen in a SISO and MIMO control structure are presented, followed by the design of MPCs for nitrogen removal. This last section, covers two control strategies: intermittent and continuous aeration. The first employs a deterministic model and the second uses a black box model both developed and identified in the previous chapter. Finally, the chapter ends with a summary conveying the main topics presented and the obtained results.

Chapter 5, presents the development of two continuous-time tuning techniques for PID-type controllers. The first part of the chapter presents a deterministic continuous time formulation for iterative feedback tuning (IFT). The section begins by introducing

a SISO formulation of the method, and then an extension to MIMO systems, which is the main contribution. The second part of the chapter presents the formulation of an optimal LQG tuning method for SISO type control systems. The development of the method is presented using a polynomial approach for the solution of the optimisation problem. This leads into an explicit solution of the optimal problem. Also, an algorithm for real-time monitoring of control systems using an optimal restricted structure LQG benchmark is developed. The algorithm is specially designed to use input-output information from the control system and perform an on-line assessment for possible re-tuning. Both methods covered in the chapter include several simulation case studies. The chapter finalises with a summary compiling the main results.

In Chapter 6 a new method for tuning of multivariable restricted-structure control systems is presented. The method is developed within a subspace framework, thus providing a transparent approach from the identification to the parameter calculation. The chapter begins by giving an introduction into the subspace framework employed. Later, the SISO case is examined and later the method is extended to a more general MIMO formulation. Several case simulation studies are presented towards the end of the chapter, and finishing with a summary of the main results.

Chapter 7 presents the real-time implementation of identification algorithms, controllers and monitoring algorithms in Swinstie WWTP. The chapter begins by presenting the development and operation of a software platform which allows the implementation of these advanced process control techniques. Some of the most important features developed include, (a) an identification module which allows the design of a real-time experiment, and the subsequent identification using two subspace algorithms. This module also allows the analysis, simulation and validation of the obtained models, (b) an MPC control module, which allows the design of an observer and a constrained or unconstrained MPC controller. The module has the advantage of allowing the user to fine-tune the parameters while operating on-line, and (c) a process monitoring module which allows a statistical process analysis to be performed in real-time by using recur-

sive principal component analysis (RPCA). Later in the chapter, the platform is used to perform identification, control design and real-time control, of dissolved oxygen in Swinstie WWTP. The chapter ends with a summary of the achieved results.

Finally, the thesis ends with conclusions from the work of the thesis and a future work programme is outlined in Chapter 8.

1.4 Main contributions in the thesis

The research study presented in the thesis covers a wide range of topics aimed at developing a data-driven approach to modelling, and control design, by using subspace identification, model predictive control, several tuning methods, and real-time implementation and experimentation in two full-scale wastewater treatment plants and simulation. The work also contains new developments in tuning of restricted-structure controllers. The following list gives the main contributions presented in this thesis organised by areas:

1. **Modelling and identification:**

   (a) A comprehensive study by simulation assessment of three subspace algorithms used to identify dissolved oxygen and nutrients was undertaken. The results are obtained:

      i. by simulation using the COST WWTP simulation benchmark (Copp, 2002), and

      ii. by using real plant data from historical SCADA archives.

   (b) The development of a systematic procedure for the identification of models suitable for use in the design of a model predictive controller.

   (c) The development of a model for an alternating activated sludge wastewater treatment plant, which can be used for the design of a model predictive controller.

2. **MPC control design:**

   (a) A comprehensive study by simulation of the behaviour of a standard model predictive controller for dissolved oxygen under different weather conditions was conducted. The study includes SISO and MIMO MPCs.

   (b) An MPC for nitrogen removal evaluated under dry weather conditions was designed.

   (c) The design of an MPC for nitrogen removal for an intermittent aeration control approach was performed.

3. **Tuning:**

   (a) The extension of the iterative feedback tuning (IFT) algorithm to multivariable deterministic systems in continuous-time, and its application to tuning of dissolved oxygen controllers.

   (b) An explicit solution of the restricted-structure optimal LQG problem for a SISO control system by using a first order model representation of the plant and its application to tuning of dissolved oxygen controllers.

   (c) A data-driven tuning algorithm for multivariable restricted-structure controllers using a subspace identification framework. The developed algorithm provides a tool for a direct calculation of a multivariable controller from closed-loop input-output data, without the need of calculating the system matrices.

4. **Benchmarking:**

   The formulation of a monitoring algorithm for the performance assessment of SISO restricted structure controllers using a restricted LQG benchmark. The LQG benchmark is calculated by assuming a first order model of the plant.

5. **Software:**

   The development of a software platform, programmed using LabVIEW, for the

testing of advanced process control and data quality management in real-time which includes the following features:

(a) Identification module with options for:

i. On-line experiment setup.

ii. Two identification algorithms: robust N4SID 'SV' and robust N4SID 'CVA'.

iii. Model analysis, simulation and validation.

(b) MPC control module with options for:

i. Observer design by pole-placement.

ii. Constrained/unconstrained MPC design using a standard formulation.

(c) Data quality management and process monitoring

i. Multivariate statistical process monitoring using recursive principal component analysis (RPCA).

ii. Diagnosis and process analysis.

6. **Real-time implementation:**

The real-time implementation of subspace identification, MPC control and monitoring tools. The thesis conveys results obtained from the real-time testing in Swinstie WWTP of:

(a) System identification using subspace identification for dissolved oxygen.

(b) The design of a MPC for dissolved oxygen control.

1.5 Publications arising from the research

This section presents a listing of scientific papers and project reports published and written by the author, which have been the results of the work presented in this thesis. The section is divided into two parts: public documents and project reports which

are confidential due to intellectual property rights and possible future commercial exploitation.

1.5.1 Public

1.5.1.1 Book chapters

A. Sánchez and M.R. Katebi. Chapter 10 Tuning of multivariable restricted structure controllers using subspace identification. In M.A. Johnson and M.H. Moradi, Eds. (2004). PID Control. Springer Verlag London.

1.5.1.2 Journal

A. Sánchez, M.R. Katebi and M.A. Johnson (2004). A tuning algorithm for multivariable restricted structure control systems using subspace identification. *Int. J. Adapt. Control Signal Process.* Accepted for publication in special issue on subspace identification.

A. Sánchez, M.J. Wade and M.R. Katebi (2004). On real-time control and process monitoring of wastewater treatment plants: real-time control. Submitted to *Trans. Inst. of Measurement and Control.*

M.J. Wade, A. Sánchez, and M.R. Katebi (2004). On real-time control and process monitoring of wastewater treatment plants: real-time process monitoring. Submitted to *Trans. Inst. of Measurement and Control.*

1.5.1.3 Conference

A. Sánchez, M.R. Katebi and M.A. Johnson (2003). Subspace Identification Based PID Control Tuning. In *Proc. of the* 13^{th} *IFAC Symposium on System Identification.* 27-29 August. Rotterdam - The Netherlands.

A. Sánchez and M.R. Katebi (2003). Predictive Control of Dissolved Oxygen in an Activated Sludge Wastewater Treatment Plant. In *Proc. of the European Control Conference ECC 2003*. 1-4 September. Cambridge - UK.

A. Sánchez, M.R. Katebi and M.A. Johnson (2003). Design and Implementation of a Control Platform for the Testing of Advanced Control Systems and Data Quality Management in the Wastewater Industry. In *Proc. of the* 4^{th} *IEEE International Conference on Control & Automation ICCA'03*. 10-12 June. Montreal - Canada. pp. 68-74.

M.A. Johnson and A. Sánchez (2003). Process Control Loop Tuning and Monitoring using LQG Optimality with Applications in Wastewater Treatment Plant. In *Proc. of the* 4^{th} *IEEE International Conference on Control & Automation ICCA'03*. 10-12 June. Montreal - Canada. pp. 84-90

A. Sánchez, M.R. Katebi and M.A. Johnson (2002). Optimal Control of an Alternating Aerobic-Anoxic Wastewater Treatment Plant. In *Proc. of the* 15^{th} *IFAC World Congress*. Vol Q: Modelling and Control of Agricultural, Biological and Chemical Systems. 21-26 July. Barcelona - Spain.

K. Mahathanakiet, M.A. Johnson, A. Sánchez and M. Wade (2002). Iterative Feedback Tuning and an Application to Wastewater Treatment Plant. *Asian Control Conference*. 25- 27 September. Singapore.

1.5.1.4 Book reviews

T.L. Blevins, G.K. McMillan, W.K. Wojsznis and M.W. Brown (2003). Advanced Control Unleashed: Plant Performance Management for Optimum Benefit. The Instrumentation Systems and Automation Society. *Reviewed by Michael Johnson and Alberto Sánchez* in IEEE Control Systems Magazine, 23(6), p.p. 88-89.

1.5.2 Project reports contributions

M. Wade and A. Sánchez (2004). Deliverable 17: Prototype Participant - Scottish Water. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

A. Sánchez, Editor (2003c). Deliverable 13: SMArt Control System Design, Algorithm and Software Report. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

A. Sánchez, (2003b). Chapter 3 Data-based loop controller tuning and multivariable dissolved oxygen control. In Deliverable 12: External WWTP Flow Rate Control Functions, Algorithm, Design. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk

A. Sánchez, (2003a) Development of a MPC for the Fast Control Layer (several sections of the report). In Deliverable 11: Internal WWTP Flow Rate Control Functions, Algorithms, Design. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk

A. Sánchez, (2004). Chapter 4 Operational planning procedures at the Swinsite test site. In Deliverable 10: Report on Operational Planning Procedures. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk

A. Sánchez, (2002c). Chapter 5 Fast Control Layer. In Deliverable 9: Coordination Control System Architecture and Design. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

A. Sánchez, (2002b). Chapter 5 Situation Assessment at Fast Control Layer. In Deliverable 7: Report on measures and algorithms for risk and situation assessment. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

A. Sánchez, (2002a). Section 2.1 Models for estimation and control: Models for Fast Layer. In Deliverable 6: Algorithms for System Monitoring. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

A. Sánchez, (2001). Chapter 3 Fast Control Layer. In Deliverable 5: Definition of System Performance Assessment Criteria, and Selection of Models for Monitoring and Control. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Chapter 2

The activated sludge treatment process and plant descriptions

An urban wastewater treatment system is comprised of three main components: a sewer network, a treatment plant and a receiving water body. The sewage produced by each household is placed into the sewer network which transports the wastewater into the treatment plant so it can be later discharged into the receiving waters, as in Figure (2.1).

Urban activated sludge wastewater treatment plants (WWTPs) are facilities which process sewage almost entirely by biological means before discharging it into a receiving water body. The main mechanism to achieve pollutant reduction is to maintain the active sludge suspended in the wastewater by stirring or aeration. The suspended solids are composed of living biomass and organic and inorganic particles. The biomass will feed from the organic particles by using oxygen or other oxidation agents, thus removing the organic material from the wastewater. Even though this simple explanation provides a basic knowledge of the mechanics involved, the real processes are complex biological systems that are difficult to describe mathematically.

This chapter introduces some basic knowledge of activated sludge wastewater treat-

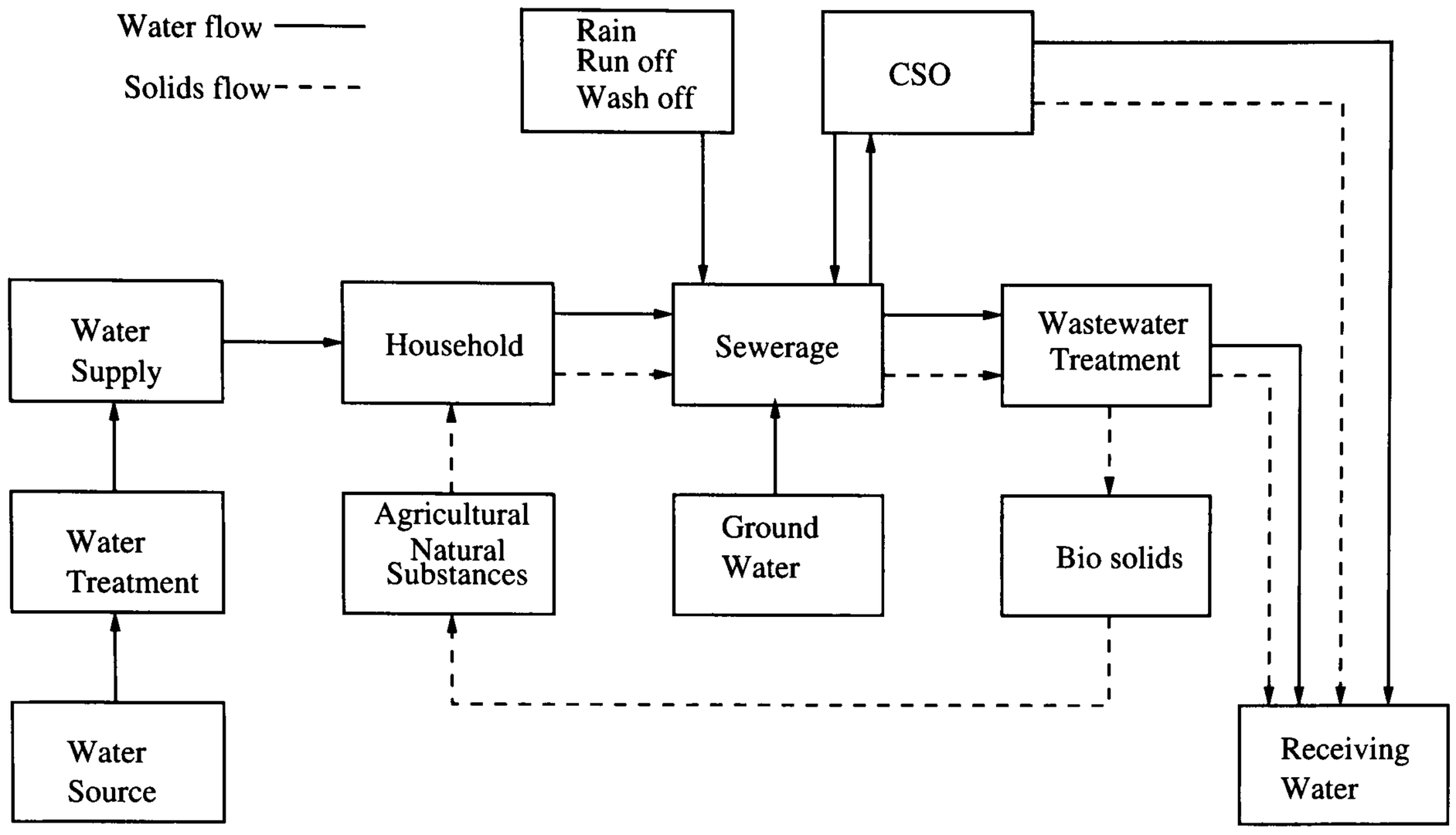

Figure 2.1: Wastewater treatment system

ment plants configurations and describes how they work. To do so, a brief description on modelling of activated sludge treatment plants is given in section 2.1. Later the COST simulation benchmark model for activated sludge wastewater treatment plants (Copp, 2002) is briefly introduced in section 2.2. Additionally two full scale treatment plants operating in Scotland and Denmark are described in sections 2.3 and 2.4.

Most of the work performed throughout the thesis employs the COST simulation benchmark model; however, real data from Helsingor WWTP (Denmark) is analysed and real-time experiments are performed at Swinstie WWTP (Scotland). Finally, the chapter ends with a brief summary of the contents previously discussed.

2.1 Generalities of activated sludge wastewater treatment plants

Modern activated sludge WWTP are normally composed of the following treatment stages (Metcalf and Eddy, 1991),

- **Preliminary treatment:** Initially, the waste-water enters the treatment plant and passes through mechanical screens to remove coarse material and solid debris. Grit removal by sedimentation or flocculation reduces potentially damaging large and heavy particles such as sand and gravel, for example. Other preliminary treatment operations include grease removal and flow equalisation. In general, preliminary treatment is a pre-treatment stage that ensures the wastewater passing to the subsequent stages is free from material that could disrupt the plant operation.

- **Primary treatment:** Organic material is partially removed by passing the wastewater through primary sedimentation tanks or primary clarifiers. Sedimentation occurs when solids that have a higher specific gravity than the liquid settle to the base of the tank, where the settled solids are removed for sludge treatment. Design of the clarifiers must account for the flow velocity and load. If the velocity of the flow is too high, the solids retention time (SRT) will be less than desirable, resulting in excess solids passing to the secondary treatment phase and exerting an increased demand on the process. Typically, 30-40% of the influent biological oxygen demand (BOD_5) and 60-75% of the influent suspended solids (SS) is removed prior to secondary treatment, Wilson (1981).

- **Secondary treatment:** The driving process of wastewater treatment occurs in the secondary treatment stage. The major biological unit operations are implemented to provide removal of organic waste and nutrients. The three biological processes that can be employed during this phase are:

  - *Aerobic processes:* Aeration of the wastewater results in oxidation of the carbonaceous and nutrient material (substrate) by chemical reactions initiated when the biomass utilise these components for biological growth. The carbonaceous material is oxidised to CO_2 and the nutrients to more benign forms of the compound. The chemical expressions for oxidisation of or-

ganic matter by micro-organisms, with (2.1) and without (2.2) nitrification, are presented as follows, Henze (1997):

$$C_{18}H_{19}O_9N + 19.5O_2 \rightarrow 18CO_2 + 9H_2O + H^+ + NO_3^- \qquad (2.1)$$

$$C_{18}H_{19}O_9N + 17.5O_2 + H^+ \rightarrow 18CO_2 + 8H_2O + NH_4^+ \qquad (2.2)$$

- *Anaerobic processes:* In the absence of free oxygen or nitrate, micro-organisms breakdown the complex organic material by hydrolysis to smaller molecules. Acid-forming bacteria break these fat, protein and carbohydrate molecules into long-chained fatty acids and amino acid, amongst others. The products of this process are acetic acid, formic acid, ethanol and methanol, which are further broken down into hydrogen, carbon dioxide (CO_2) and methane (CH_4). This process requires a number of different types of bacteria to perform the different degradation stages, all of which are sensitive to factors such as pH, temperature, toxicity or even the presence of oxygen. Hence, design of anaerobic treatment processes requires careful selection of conditions to enable the appropriate operational performance. One benefit of anaerobic digestion is the production of biogas (CH_4), which can be used as a source of energy, on-site or supplied to the national electricity grid, if the quantity is large.

- *Anoxic processes:* In anoxic conditions, free oxygen is absent, but nitrate is present, providing a source of oxygen for denitrifying bacteria. The process of denitrification may be written as, Metcalf and Eddy (1991):

$$6NO_3^- + 5CH_3OH \rightarrow 5CO_2 + 3N_2 + 7H_2O + 6OH^- \qquad (2.3)$$

The principle of the activated sludge plant is that mass flow of wastewater is

in kept in continuous motion through the plant by gravity, pumping, mixing and aeration. In this way, treatment is performed in an effective and controllable manner. However, it is necessary to maintain the biology in the secondary phase long enough for biomass growth through contact with the substrate and the subsequent associated reactions. The length of time, or mean cell residence time, that the biomass remains in the secondary treatment stage is known as the sludge age. The hydraulic retention time (HRT) and sludge age must be balanced so that the process kinetics can take place. The Return Activated Sludge (RAS) feedback loop recycles sludge from the secondary clarifier to the aeration tank in order to maintain the sludge concentration. Excess sludge is wasted from the secondary clarifier and is treated separately with sludge collected from the primary clarifiers. An internal nitrate recycle may also be used in secondary treatment to supplement the nitrate concentration in the anoxic zone. Typically, the anoxic (denitrification) zone is situated prior to the aeration (nitrification) tank and the internal nitrate recycle is a loop between the end of the aeration tank and the inlet to the anoxic zone. However, it is possible to have different configurations based on the design criteria and treatment objective.

- **Tertiary treatment:** After the secondary stage, additional treatment may be required to remove and remaining undesirable substances, such as suspended solids, inorganic ions, heavy metals and synthetic organic matter. The last three items may be particularly pertinent where the wastewater contains effluent from industrial manufacturing such as pharmaceutical production, pulp mills or the metals industry, for example. Common advanced wastewater operations include filtration, microstraining, air stripping and reverse osmosis.

Figure (2.2) presents a summary of the treatment stages by means of a flow-chart.

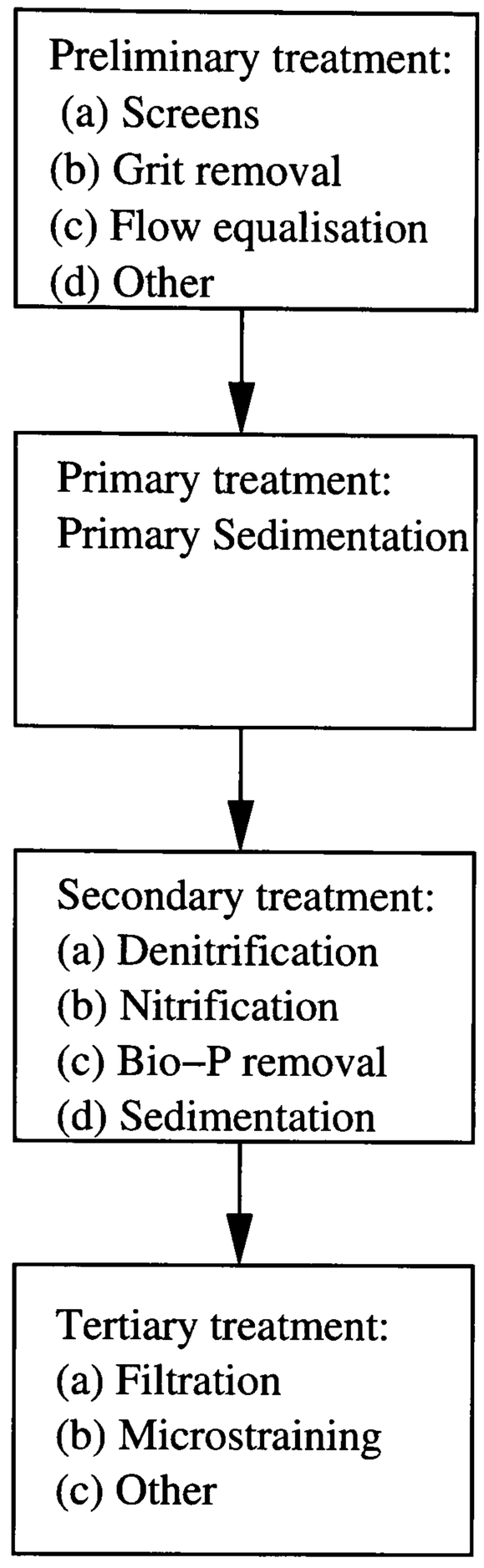

Figure 2.2: Treatment stages.

Most of the control priorities are centred in the secondary treatment, which is where most of the biological treatment occurs. The secondary treatment is as well the most sensitive part of the treatment process. As discussed before, the main objective in the secondary treatment is to keep an acceptable concentration of suspended solids in the wastewater. This however, might be difficult to achieve under certain conditions. For example, if the plant receives a high hydraulic load, the SRT might be sharply decreased, thus not allowing enough time for treatment. An even worst scenario would be that the hydraulic load would drag a large quantity of suspended solids with it, leaving the plant with reduced treatment capability, which can take up to weeks to restore.

This type of effects caused by external disturbances can be somehow mitigated by using some ingenious control strategies as step-feed or aeration tank settling (Nielsen *et al.*, 2000, 1996). It is argued that by using these types of control strategies, it is possible to increase significantly the plant capacity without any risk to solids loss. This claim is however still under research.

Under normal operating conditions, that is the operation of the plant under normal daily influent flows, other control objectives might be more important. About 10% of a plant operating costs is spent in energy (electricity), from which a considerable amount is employed for aeration. Most plants employ a constant aeration, trying to keep dissolved oxygen at a concentration of 2 mg/l. This however, has been demonstrated in practice to be unnecessary. Plants like Helsingor in Denmark, aerate the sludge in a cyclic manner and only when required (high NH_4 concentration). In addition, the oxygen setpoint is also calculated accordingly to the ammonia level, thus providing further savings. However, this might still leave for improvement since the setpoint is calculate only on present information and not using models which can somehow predict future plant behaviour. The use of chemicals for phosphorus release is also a heavy burden in the operation budget of a plant. Therefore the minimisation of chemical expenditure by improving biological phosphorus (Bio-P) release is an active research topic nowadays, specially because the process seems to be very sensitive and unstable(Nielsen *et al.*, 2002).

Many of these problems to improve efficiency in the use of resources in the control and operation of wastewater treatment plants has lead to the use of advanced process control and monitoring techniques. Among the advanced process control techniques, model predictive control (MPC) has been a strategy which has been very successful in several process industries. However, as pointed out by Yuan *et al.* (2001), this method is still in its infancy in the area of wastewater treatment. One of the reasons for its lack of use, amongst others, is the lack of simple models which can be easy employed. Normally, the scientific community has been attracted to the use of well established

models like the Activated Sludge Model No.1 (ASM1) Henze *et al.* (1987).

The ASM1 model describes the degradation of carbonaceous material as well as nutrient removal. Mathematically, the model is composed of a set of 13 non-linear differential equations with 19 parameters (in its original version), most of them with a high degree of uncertainty. Extensions to the original ASM1 model, have appeared in time to include more complex phenomena like phosphorus precipitation and Bio-P (Henze *et al.*, 1995, 1999; Gujer *et al.*, 1999). However, complexity has also increased considerably. Due to these characteristics the applicability of the ASM models is restricted to benchmarking for simpler models and research.

Several researchers have tried to derive simpler models based on the ASM models. Jeppsson (1995) derived a set of reduced order non-linear models based on the original ASM1 with the purpose of nutrient control. However, parameter estimation still required a great amount of effort. Later, Anderson *et al.* (2000) obtained a linear reduced order model from ASM1, for an alternating aerobic-anoxic process also with the purpose of nutrient control. Lindberg (1997) used simple models extracted from ASM1 for dissolved oxygen control, and begun studying the possibility of black-box parametric identification of models for nutrient control by using subspace identification. A similar work has been recently reported by Sotomayor *et al.* (2003).

Other attempts to model activated sludge wastewater treatment process are grey-box models. These models employ part of the deterministic structure of the ASM models, and use special techniques to model the remaining uncertain parts as in (Carstensen, 1996; Bechmann, 1999). These models however, will still suffer from identifiability problems in the parameter estimation.

This thesis adopts a linear black box identification approach for dissolved oxygen and nutrient removal. Much of the effort is concentrated in demonstrating that, for many purposes, the use of sophisticated and complex identification algorithms and parameter estimation using Kalman filtering is unnecessary and time consuming for this applica-

Table 2.1: Sensors

Variable	Sensor
Level	Bubblers Sonic, ultrasonic and microwave Capacitance and impedance probes Float level instruments
Flow	Weirs Parshall flume Magnetic meters Sonic meters Turbine meters Venturi tubes and flow tubes Vortex shedding
Wastewater bio-chemical characteristics	Dissolved oxygen sensor pH sensor Suspended solids sensor Turbidity sensor Ammonia analyser ($NH_4 - N$) Nitrate analyser ($NO_3 - N$) Phosphate analyser (PO_4) Chemical oxygen demand (COD) sensor

tion. Subspace identification techniques are very powerful algorithms, numerically stable and extremely easy to use. Modelling and control of alternating wastewater treatment plants is also discussed.

2.1.1 Sensors and actuators

One of the main limiting factors for the implementation of any control technology is the availability of accurate on-line sensors. Sensor technology for bio-processes have evolved considerably in the past few years, therefore providing a wider scope of on-line measurements. The instruments employed in wastewater treatment systems are many and varied. Table (2.1) summarises some of the most useful variables and the type of sensors employed to measure them (Metcalf and Eddy, 1991; Marinaki and Papageorgiou, 2002).

Another limiting factor for control implementation is the number of available control handles. Unfortunately, wastewater treatment plants have a very limited number of

Table 2.2: Actuators

Variable	Sensor
Control valves	Globe valves Butterfly valves Ball valves Diaphragm valves Plug valves Diffusers
Pumps	Fans Blowers Compressors
Motors	Induction electric motors DC electric motors Combustion engines

control handles which are driven by actuators. Table (2.2) summarises some of the most common actuators employed (Metcalf and Eddy, 1991; Marinaki and Papageorgiou, 2002). Additionally, Figure (2.3) presents a sensor and actuator distribution in a generalised WWTP, according to the processes and control loops.

2.2 The COST simulation benchmark wastewater treatment plant

The WWTP simulation benchmark was developed by the COST actions 624 & 682 research group Copp (2002). COST was founded in 1971 as an intergovernmental framework for European Co-operation in the field of Scientific and Technical Research. The goal of COST is to ensure that Europe holds a strong position in the field of scientific and technical research for peaceful purposes, by increasing European co-operation and interaction in this field. COST action 682 'Integrated Wastewater Management' (1992-1998) focused on biological wastewater treatment processes and the optimisation of the design and operation based of dynamic process models. Action 624 was dedicated to the optimisation of performance and cost-effectiveness of wastewater management systems.

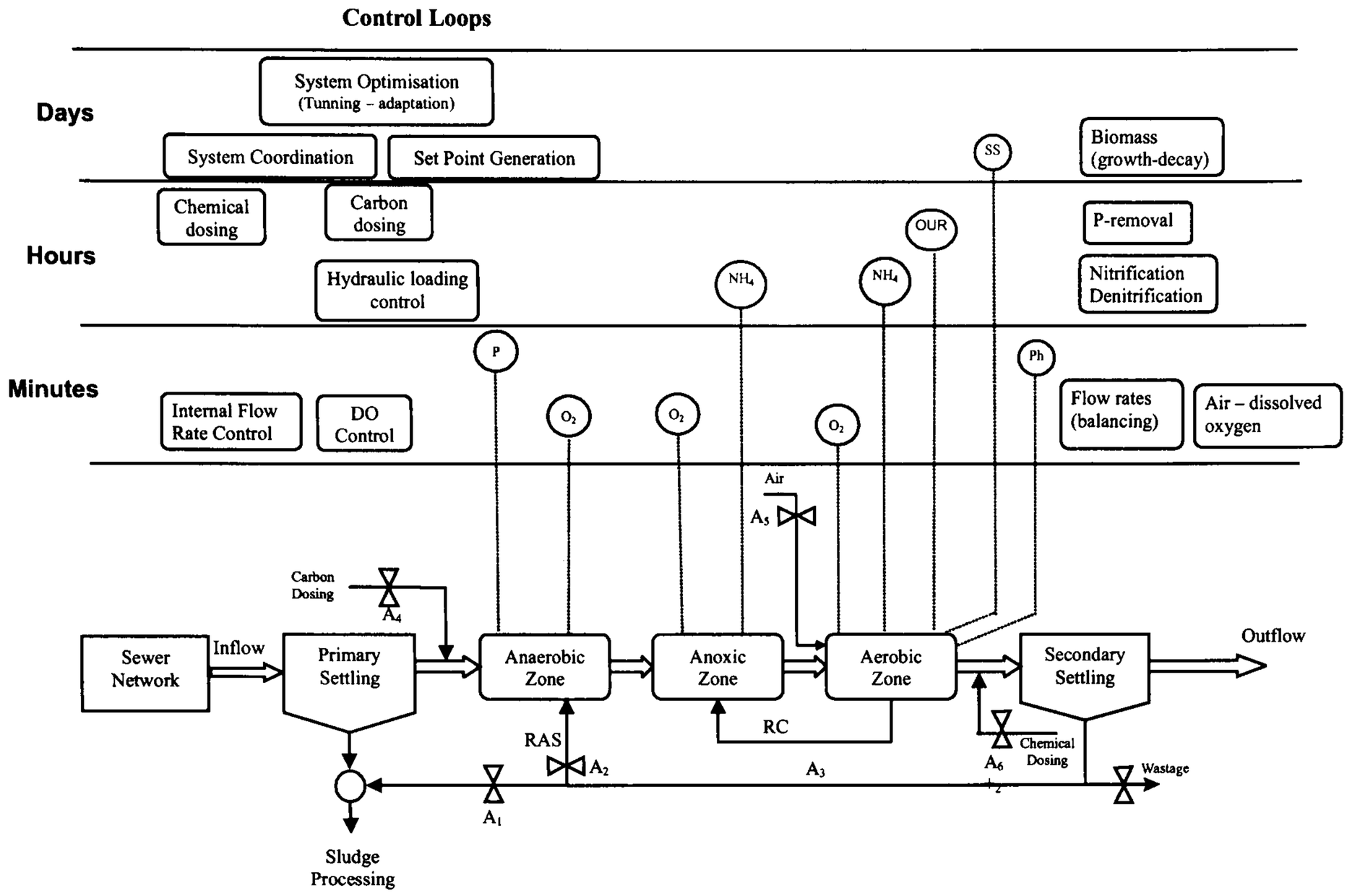

Figure 2.3: Sensor and actuator distribution according to processes and control loop.

The simulation benchmark is a fully defined simulation protocol and was developed as a tool for evaluating activated sludge wastewater treatment control strategies. The simulation benchmark is by itself platform independent and has been tested in several programming languages and simulation packages as MATLAB/SIMULINK, Fortran, SIMBA, STOAT, WEST, EFOR, GPS-X and BioWin. The Simba implementation of the simulation benchmark has been used in this thesis.

This section provides a brief summary of the COST simulation benchmark, focusing in the parts which have been modified with the purpose of the development of this thesis. Not all of the original definitions and indexes used to evaluate the control strategies have been employed since they were not of particular interest for this thesis.

2.2.1 Plant layout

The benchmark is composed of five cascade biological reactors and a 10-layer non-reactive secondary settling tank. The plant layout is presented in Figure (2.4). The

plant is fully defined and has the following characteristic features:

- 5 biological cascade tanks with a secondary settler
- Total biological volume of 5999 m^3.
  - Tanks 1 and 2, each of 1000 m^3 .
  - Tanks 3, 4 and 5 each of 1333 m^3.
  - Tanks 1 and 2 un-aerated, but fully mixed.
- Aeration of tanks 3, 4 and 5 achieved using a maximum of K_La of 360 d^{-1}.
- DO saturation level of 8 mg/l.
- Non-reactive secondary settler with a volume of 6000 m^3.
  - Area of 1500 m^2.
  - Depth of 4 m^2.
  - Subdivided into 10 layers.
  - Feed point to the settler at 2.2 m from the bottom.
- 2 internal recycles:
  - Nitrate recycle from the 5^{th} to the 1^{st} tank at a default flow rate of 55338 m^3d^{-1}.
  - RAS recycle from the underflow of the secondary settler to the front end of the plant at a default flow rate of 18446 m^3d^{-1}.
- WAS is continously pumped from the secondary settler underflow at a default rate of 385 m^3d^{-1}.

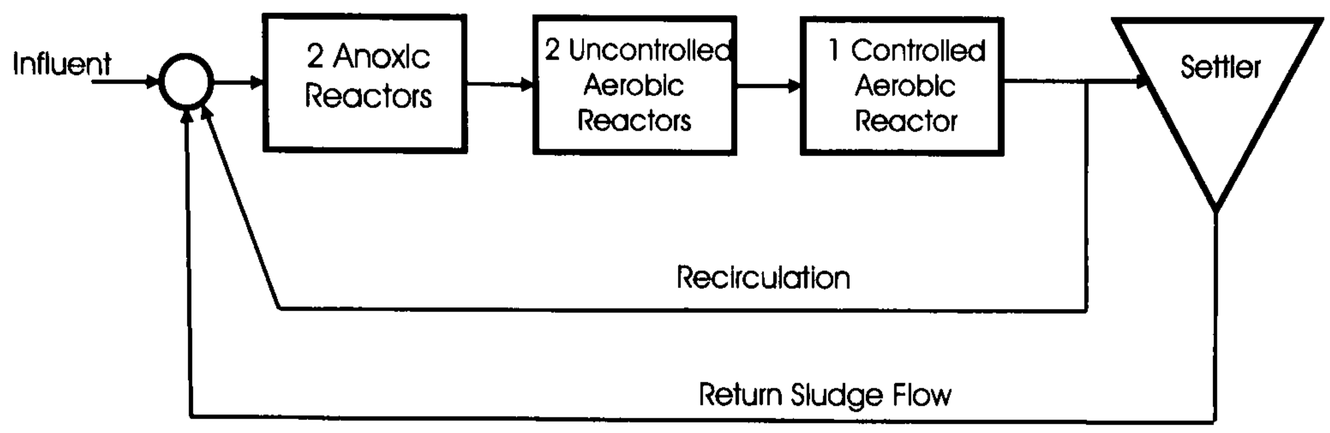

Figure 2.4: COST Simulation Benchmark - Plant Layout.

2.2.2 Process models

The biological tanks (aerated and un-aerated) are modelled using ASM1 (Henze *et al.*, 1987). The settler is modelled using a double-exponential settling velocity function (Takcs *et al.*, 1991).

The first two reactors are anoxic, while the last three are aerobic. The model also has a recirculation flow and a return sludge flow.

The dissolved oxygen sensor utilised in the simulations has a 1-minute time delay and 1 minute sampling time. Actuators have been modelled as physical limitations in the air supply equivalent to a maximum oxygen transfer (k_La) of 360 $\left[day^{-1}\right]$. The simulation benchmark also provides three files of dynamic influent data for dry, rain and storm conditions, and a file of constant influent data used to stabilise the plant.

2.2.3 Influent composition

There are three influent disturbances and each is meant to be representative of a different weather condition: dry, rain and storm. A constant influent condition has also been assumed for certain tests and simulations.

2.2.4 Sensors and actuators

The original definition of the COST benchmark was somehow explicit concerning different sensors; however, there is no definition of actuators. Table (2.3) and (2.4) sum-

Table 2.3: Sensors in the COST benchmark

sensor	Time delay (min)	Sampling time (min)	Units
Dissolved oxygen (DO)	0	1	mg/l
Nitrate (NO_3)	10	15	mg/l
Ammonia (NH_4)	10	15	mg/l
Total Nitrogen (TN)	10	15	mg/l
Flow (F)	0	15	$10^3\ m^3/d$

Table 2.4: Actuators in the COST benchmark

Actuator	Input	Range	Output	Range	Units
Blowers	Normalised capacity fraction	1-10	Airflow	0-99975	$m^3 d^{-1}$
Pumps	Normalised capacity fraction	1-10	Flow rate	0-10000	$m^3 d^{-1}$

marise the sensors and actuators characteristics employed throughout the thesis.

2.2.5 Dissolved oxygen controllers

Most of the identification exercises and control designs presented through this thesis, assume that there is an existing control level operating in the plant. As originally presented, the benchmark considered only a single continuous PI controller in the last aeration basin, while the other two basins were considered to be uncontrolled and with a fixed oxygen transfer rate ($k_L a$) of 10 $hour^{-1}$. This approach has been modified in order to include more advanced problems of identification and control. All the possible configuration used throughout the thesis are listed in Table (2.5).

Table 2.5: Dissolved oxygen PI control configurations

No.	Type	Reactors	$k_p(\times 10^4)$	$k_i(\times 10^4)$	T_s
1	discrete	5	1	0.93055	1 minute
2	discrete	3, 4 and 5	1	0.93055	1 minute
3	continuous	5	1	50	–

2.3 Swinstie wastewater treatment plant

Swinstie Wastewater Treatment Plant is located close to the town of Cleland, about 25km south-east of the City of Glasgow. The plant was commissioned in 1998 and has

a capacity of 20075 p.e. but the actual load on the plant is typically 13000 p.e. The influent wastewater is predominantly household effluent and very little is industrial discharge.

Following a process review by Scottish Water in May, 2003, a revised plant configuration was implemented for Swinstie WWTP due to over sizing of the main unit operations at the plant. The details of the actual plant configuration are provided below.

2.3.1 Sewer network

The sewer network serves a number of small communities close to Swinstie and the main trunk sewer enters the plant by gravity alone. At present there is no control on the volume of influent to the plant, it passes immediately into a channel that incorporates a control flume to limit the flow to the works in periods of high hydraulic loading.

2.3.2 Plant configuration

The plant inlet has no control mechanism for regulating and distributing the influent. All flows in excess of 4 daily dry weather flow (DDWF) bypasses the plant and is discharged into the River Calder at the south side of the site after passing through a 6mm mechanical screen.

Preliminary treatment consists of 6mm screening provided by duty standby mechanical raked screens. A hand-raked emergency bypass screen is also provided.

Grit removal is provided by 2 detritors downstream of the screens. There is no fat or grease removal in this unit due to the low amounts observed in the influent, although the primary tanks do have scum traps that remove the small amount of fatty material and debris that passes to the primary treatment stage.

Primary treatment consists of 2 primary settling tanks. Excess return activated sludge (ERAS) is returned to the primary tanks for co-settling. Scum and settled solids are re-

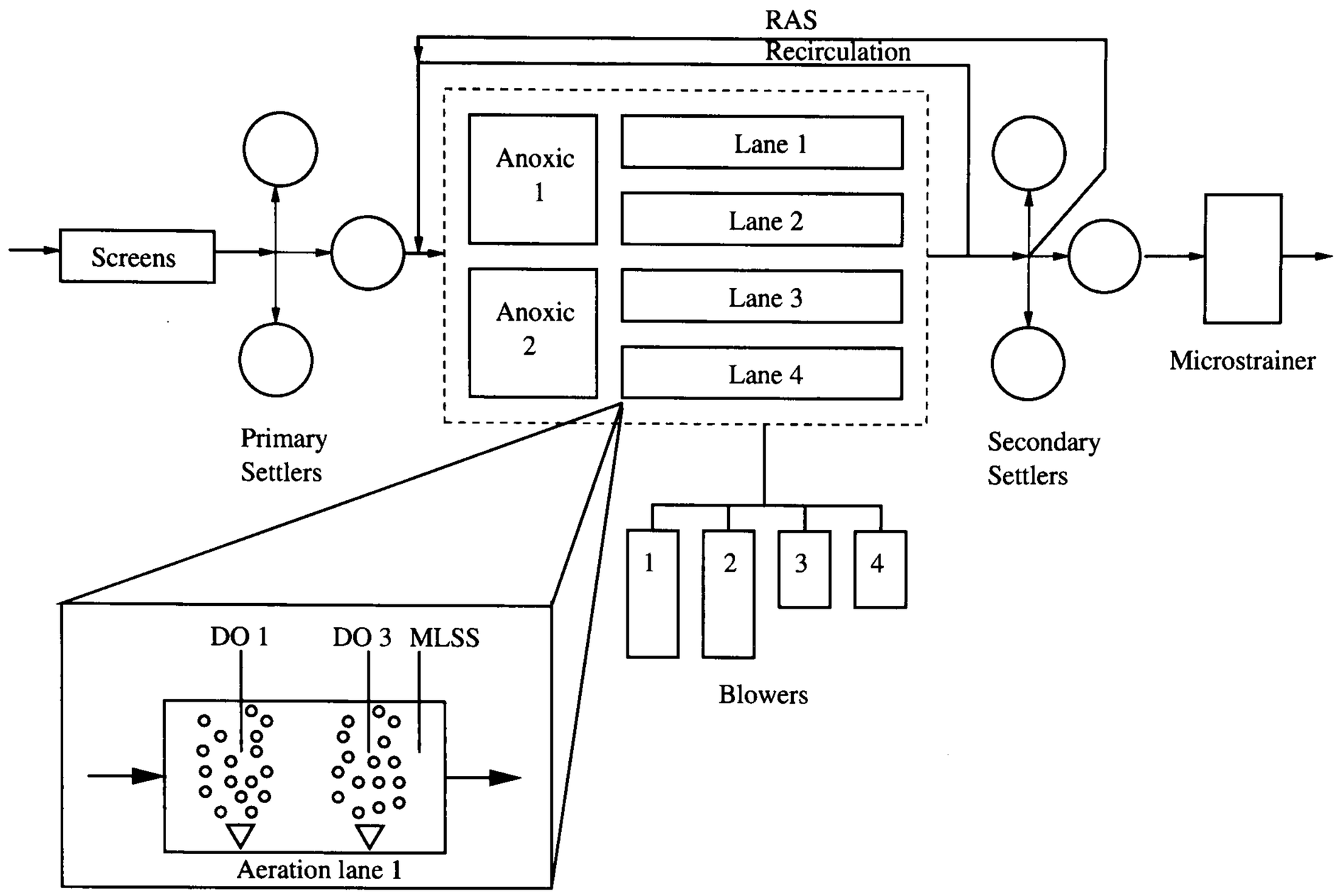

Figure 2.5: Swinstie WWTP plant layout

moved to a sludge holding tank and the wastewater is passed to a pre-treatment channel prior to the secondary treatment stage.

A flume in the channel leading to secondary treatment limits the flow to full treatment to 2 DDWF. Settled sewage in excess of this passes to the river via 6mm mechanical screens. The aeration is preceded with an anoxic zone (1 anoxic tank) for denitrification and sludge conditioning. Secondary treatment consists of 2 parallel aeration tanks. Aeration is provided by fine bubble diffused aerators (FBDA) situated in three zones down the length of each tank. This provides a stepped oxygen supply along the length of the aeration tank.

Mixed liquor from the aeration tank is dosed with ferric sulphate to remove phosphate prior to final clarification. An internal recycle facility is also provided within the aeration tanks.

Final clarification is provided by 2 final settlement tanks with equal distribution of the influent. A proportion of the settled solids is returned to the secondary treatment stage

as return activated sludge (RAS). The clarified water passes through a micro-strainer for polishing before discharge to the river. The residue is returned to the primary settling tanks for co-settling. All sludge is removed by tanker and treated off-site. Improvement of the control system relies on knowledge of the existing structure of the wastewater systems. Figure (2.5) presents the layout configuration of Swinsite WWTP.

2.3.3 Instrumentation and Control

The current available data acquisition and control system at Swinstie consists of an IC2000 SCADA unit with process and measurement visualisation capacity.

The works is controlled by a Siemens TI-565 PLC located in the power distribution house. The system contains all the algorithms for the control of the plant.

The SCADA software provides a limited amount of control by allowing set points to be altered and communicates with the user via mimic screens and associated alarms. The system has one workstation for user input located in the main control room. This comprises a VDU, keyboard, mouse and two printers.

There is no control provided except that the set points for the controlling parameters, e.g. desludging times, detritor operating times, DO levels, can be altered within design set limits.

The aeration control system employs 4 measurements of DO, 2 in each lane (inlet and outlet). The measurements are averaged and this value is compared with a high and low DO range. If the mean is above the range, the controller will decrease the blower speed by 10% every 1 minute for the high capacity blowers and 10 minutes for the low capacity blowers. Similarly, a 10% speed increase will occur if the DO mean is below the established range. This is not an ideal control system as there is no account for variation in DO between lanes and percentage speed variation is a very imprecise form of control.

A certain amount of aeration is required to maintain an appropriate mixture and overload the blowers. This however results in a very limited range of speed to which the blowers can actuate to perform control actions.

The anoxic zone has one DO measurement that is controlled independently. When the DO is above a set limit, the output penstocks to the aeration tanks will close fractionally, which will increase the hydraulic retention time and mean cell residence time in the anoxic zone, thus reducing the DO in this section.

ERAS is controlled in a simple manner. When the level of ERAS in the chamber reaches a set level, the pumps switch on and transports the ERAS to the primary settlement tanks until the ERAS level reduces to a set minimum.

There is currently no control structure in place for the RAS flow, although the screw pumps can be switched on or off depending on requirements.

2.4 Helsingor WWTP

Helsingor municipality is in the north of Zealand in Denmark. The treatment plant is a recirculating biological nutrient removal (BNR) plant with pre-clarification and sludge digestion, and with a capacity of 26000 pe.

The plant has three aeration tanks and five secondary settlers in parallel. Each aeration lane has three banks of fully controllable diffusers and three dissolved oxygen measurements. The plant also performs Bio-P removal by using anaerobic tanks. There is also access to flow meters and pumping information from the sewer network.

Helsingor uses an alternating aeration scheme. Dissolved oxygen setpoints, recirculation rates, carbon dosing, chemical dosing, and phase length are all controlled by STAR. STAR is the acronym for Superior Tuning and Reporting, which is a sophisticated control system and reporting machine. STAR is able to archive historical data

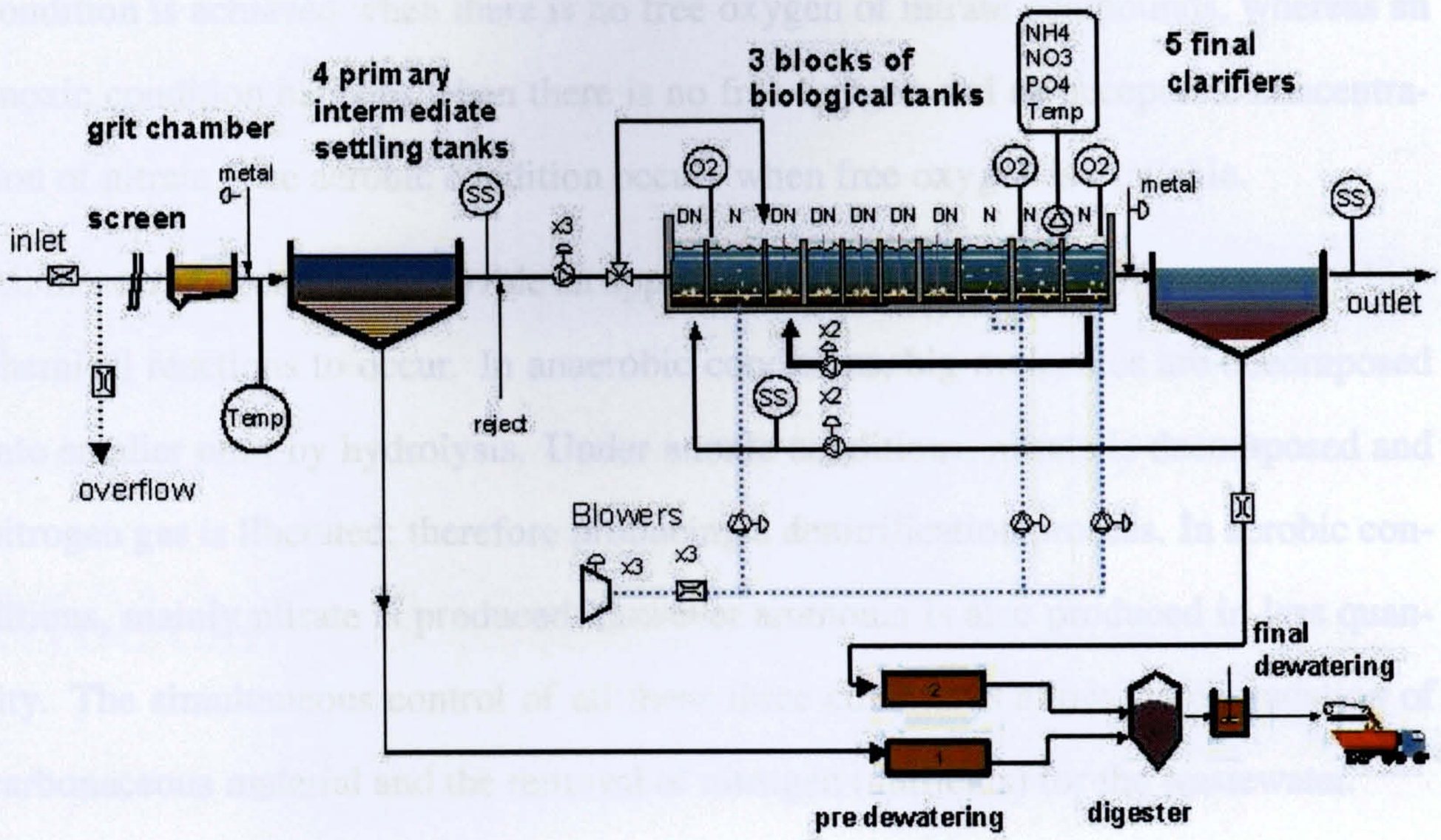

Figure 2.6: Helsingor WWTP layout. (taken from www.smac.dk/strardb/)

from several years which has been useful in this work to analyse and perform identification experiments. The plant layout is presented in Figure (2.6).

2.5 Summary

This chapter has provided a brief introduction into the activated sludge process, the COST simulation benchmark and a brief description of two-full scale treatment plants in Denmark and Scotland.

The treatment of wastewater using activated sludge is the most common process used to treat sewage. The treatment can be divided into four stages: preliminary, primary, secondary and tertiary. Most of the biochemical processes occur in the secondary treatment, where the primary objective is to keep a sustained concentration of suspended solids either by stirring or aeration.

The treatment process conveys several complex bio-chemical reactions from which three type of processes are relevant: anaerobic, anoxic and aerobic. An anaerobic

condition is achieved when there is no free oxygen of nitrate compounds, whereas an anoxic condition happens when there is no free oxygen and an acceptable concentration of nitrate. The aerobic condition occurs when free oxygen is available.

All these three conditions provide an appropriate environment for different type of biochemical reactions to occur. In anaerobic conditions, big molecules are decomposed into smaller ones by hydrolysis. Under anoxic conditions, nitrate is decomposed and nitrogen gas is liberated; therefore producing a denitrification process. In aerobic conditions, mainly nitrate is produced; however ammonia is also produced in less quantity. The simultaneous control of all these three conditions allow the degradation of carbonaceous material and the removal of nitrogen (nutrients) for the wastewater.

The description of the two plants in Scotland and Denmark is given in the chapter. The description provides information about several characteristics of both plants, in particular their instrumentation and control handles. The knowledge of this information is important for several reasons: (a) Chapter 3 uses historical data collected from Helsingor WWTP for identification and therefore the plant configuration is of vital importance. (b) A software platform developed for the real-time implementation of advanced process control uses Swinstie WWTP as its main target, thus precise information about plant layout, instrumentation and handles are important.

Chapter 3

Models and identification

The first model for the activated sludge process widely accepted by the scientific community, was the Activated Sludge Model No.1 (ASM1) developed by Henze *et al.* in 1987. The model is comprised of 13 non-linear differential equations and 19 parameters. ASM1 describes, among others, three important processes: (a) degradation of carbonaceous material (b) nitrification and (c) denitrification.

Further developments in the understanding of the internal processes and also the necessity of explaining more complex phenomena like phosphorus removal, either by precipitation or by biological means, lead to the development of ASM2 (Henze *et al.*, 1995) and ASM2d (Henze *et al.*, 1999), and later ASM3 (Gujer *et al.*, 1999) to solve some numerical problems which arose from the initial assumptions used in ASM1 and propagated to the ASM2 models. ASM3, however, does not consider the biological release of phosphorus; and special extensions are necessary to model this process. As the models grew to include more processes, so did their complexity. It has been internationally accepted that some of the ASM models parameters are unidentifiable, see for example Jeppsson (1996). Therefore, they can only be calibrated to represent the process up to a certain limit of accuracy. Also, many of the processes described in the model cannot be measured, as for example concentration of bacteria. This has limited the direct use of ASM models for the purpose of control, where simple models with a

reasonably good accuracy are required.

The main purpose of this chapter is to explore the possibility of obtaining linear time-invariant models for control of dissolved oxygen and nitrogen removal, either by parametric identification or model reduction and simplification.

The first approach is called black-box modelling, since the obtained model only represents the input-output characteristics of the system; thus its internal structure has no physical or chemical meaning. The second approach is called deterministic modelling, and is achieved by simplifying and reducing the ASM models according to the application. The use of the term 'deterministic' to denote these type of models is a common process engineer terminology which means that the model possesses some internal structure which has a physical or chemical meaning. This should not be interpreted as the more general case of a deterministic model in systems and control theory.

There is also a third possibility, which makes use of the system structure (knowledge of the process) but also includes some black-box identification. A model with such characteristics is called a grey-box model. A very limited amount of research has been performed in this area and it is still left to prove its suitability for control purposes due to the amount of on-line measurements required, see for example Carstensen (1996) or Bechmann (1999).

As described in previous chapters, the activated sludge process is inherently multivariable, and one of the main objectives, as outlined in the SMAC project description, was to try to counteract undesirable effects due to these couplings. Therefore, the use of a multivariable control technique was one of the main objectives. The aim of this chapter is to obtain simple models which can be used to design a multivariable control technique, as model predictive control. To obtain such models a multivariable identification method is preferred.

Probably one of the most successful class of identification algorithms for multivariable systems are the subspace identification methods. The main success of these algorithms

is that they transform a non-linear optimisation problem, which is the main bottle-neck in other algorithms, into a linear optimisation with excellent numerical properties; which make them fast and robust. There are still however many open issues in the development of these algorithms, especially regarding accuracy. This is probably the major drawback, since in contrast with prediction error methods (PEMs), subspace identification does not work in an iterative way over an error tolerance; therefore input-output uncertainty cannot be directly measured or minimised during the identification procedure. This however, has not limited its now extensive use, specially for multivariable systems; where PEMs considerably increase their numerical complexity with a relatively small number of input-output signals.

Regarding model reduction and simplification, there is a considerable amount of work performed in this area, as for example the work of Jeppsson (1995); Anderson *et al.* (2000); Huang and Hao (1996). The approach consists basically in reducing and simplifying the original ASM model by eliminating unmeasurable or unimportant variables depending on the application.

In summary, subspace identification and model reduction are employed to obtain adequate models for control of dissolved oxygen and nitrogen removal. The chapter employs data generated by the COST simulation model and real data collected from Helsingor WWTP. Thus, another of the objectives of this chapter is to cross-validate model characteristics obtained by simulation with their counterparts obtained from real data. This study has however been restricted to the dissolved oxygen case, and with a limited scope into nitrification and denitrification processes.

In addition, *open-loop* and *closed-loop* algorithms are employed to identify models of dissolved oxygen and nitrogen removal which can later be embedded in a more complex system structure; with the objective of including actuator and controller dynamics which can be used to design a MPC controller with the inclusion of constraints over particular signals.

The model reduction approach is used to obtain a model which can be used to control an intermittent aeration system using model predictive control, as will be described in the following chapter. The model has the characteristic of combining two model simplifications for anoxic and aerobic conditions. Therefore, the obtained model unifies both conditions seamlessly and simplifies the design of the controller in the next chapter.

Also, due to the large number of subspace algorithms available, this thesis has adopted a comparative approach in order to perform a qualitative assessment of three open-loop algorithms: N4SID, robust N4SID 'SV' type, robust N4SID 'CVA' type, and two closed-loop algorithms MOESP type and CVA. The significance of this abbreviations will be left to the next section.

The chapter is organised in the following manner: the first section presents a brief introduction to subspace identification algorithms, which are employed throughout this chapter. The next section, explores the identification of dissolved oxygen considering the univariate and the multivariable case. Both cases make use of the COST Simulation Benchmark introduced in Chapter 2.

In section 3.3 the identification and modelling of nutrients is discussed. A model for alternating aeration in a plug-flow single reactor is developed, as well as continuous aeration model based on the COST Benchmark. For this last case, only *open-loop* subspace algorithms are employed.

Section 3.4, explores the use of real data for the identification of dissolved oxygen and for nutrients in a limited scope. Several month of data are employed to obtain a more general perspective of the behaviour of the plant and the type of models expected. The results produced from this section are to be compared with the results obtained in the previous sections in which data from simulation was employed.

Finally, the chapter ends with a summary of the main results and conclusions.

3.1 Subspace identification

The beginning of the 90s experienced the birth of a new type of identification algorithms now known as subspace identification. These methods have their origin in state-space realisation theory, however the literature related to this method can be traced as far back to 1933 to principal component factor analysis (Hotelling, 1933). Seminal papers on the recovery of system matrices from impulse responses came later with contributions by Ho and Kalman (1966), and subsequent improvements by Zieger and McEwen (1974), and Kung (1978) who introduced the singular value decomposition to reduce the sensitivity to errors in the measured impulse response. These methods, however, required special inputs as impulse or white noise signals.

Methods to determine the system matrices directly from data without explicitly forming impulse responses were developed by De Moor *et al.* (1988), Moonen *et al.* (1989) and Verhaegen (1991) and called in the literature as direct subspace state-space system identification methods (4SID). These methods only give consistent estimates of the system matrices under certain restrictions on the noise characteristics.

Further developments using instrumental variables have been suggested to overcome this last drawback (Larimore, 1990; Viberg *et al.*, 1993; Verhaegen, 1994; van Overschee and De Moor, 1994). Later on, van Overschee and De Moor (1995) proved that the algorithms by Larimore (1990) (canonical variate analysis - CVA), Verhaegen (1994) (multivariable output-error state space - MOESP) and van Overschee and De Moor (1994) (numerical algorithms for subspace state-space system identification - N4SID) can all be formulated in one unifying framework and therefore are special cases of a single general algorithm.

The following section provides some background on the basic concepts involved in subspace system identification based on the work of van Overschee and De Moor (1995, 1996*b*).

3.1.1 Combined deterministic-stochastic identification

Three of the most commonly used subspace identification algorithms are CVA, MOESP and N4SID. CVA developed by Larimore (1990), makes an extensive use of principal angles and directions, while MOESP (Verhaegen, 1994) and N4SID (van Overschee and De Moor, 1994) are based on geometrical and linear algebra concepts. As demonstrated in (van Overschee and De Moor, 1995), all three algorithms are special cases of one general algorithm which is described in this section.

The subspace identification problem is formulated as follows. Let $u(k) \in \mathbb{R}^m$ and $y(k) \in \mathbb{R}^l$ be the observed input and output signals from the unknown system:

$$x(k+1) = Ax(k) + Bu(k) + w(k) \tag{3.1}$$

$$y(k) = Cx(k) + Du(k) + v(k) \tag{3.2}$$

with

$$E\left[\begin{pmatrix} w(k) \\ v(k) \end{pmatrix}\begin{pmatrix} w^T(l) & v^T(l) \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}\delta_{kl} \geq 0 \tag{3.3}$$

and A, $Q \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{l\times n}$, $D \in \mathbb{R}^{l\times m}$, $S \in \mathbb{R}^{n\times l}$, $R \in \mathbb{R}^{l\times l}$, and δ_{kl} the Kronecker delta. The signals $v(k) \in \mathbb{R}^l$ and $w(k) \in \mathbb{R}^n$ are unobserved, uncorrelated, Gaussian distributed, zero mean, white noise vector sequences. It is also assumed that the pair $\left\{ A, \quad C \right\}$ is observable and $\left\{ A, \quad \begin{bmatrix} B & Q^{1/2} \end{bmatrix} \right\}$ is controllable and the input $u(k)$ is persistently exciting (Ljung, 1987). Then, the problem is stated as: given a sufficiently large number of measurements of the input $\{u(k)\}$ and output $\{y(k)\}$ generated by the unknown system described by equations (3.1), (3.2) and (3.3), find A, B, C, D, Q, R and S to within a similarity transformation.

The matrix input-output equation for the system can be obtained by recursive substitu-

tion of equations (3.1) and (3.2).

$$Y_f = \Gamma_i X_i + H_i^d U_f + H_i^s M_f + N_f \tag{3.4}$$

where:

- Γ_i is the extended observability matrix ($i > n$)

$$\Gamma_i \triangleq \begin{bmatrix} C & CA & CA^2 & \cdots & CA^{i-1} \end{bmatrix}^T \tag{3.5}$$

- H_i^d is the deterministic lower block triangular Toeplitz matrix

$$H_i^d \triangleq \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & \cdots & D \end{bmatrix} \tag{3.6}$$

- H_i^s is the stochastic lower block triangular Toeplitz matrix

$$H_i^s \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ CA & C & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{i-2} & CA^{i-3} & \cdots & 0 \end{bmatrix} \tag{3.7}$$

- The input and output block Hankel matrices

$$U_{0|2i-1} \triangleq \frac{1}{\sqrt{j}} \begin{bmatrix} u_0 & u_1 & \cdots & u_{j-1} \\ u_1 & u_2 & \cdots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{2i-1} & u_{2i} & \cdots & u_{2i+j-2} \end{bmatrix} \tag{3.8}$$

$$Y_{0|2i-1} \triangleq \frac{1}{\sqrt{j}} \begin{bmatrix} y_0 & y_1 & \cdots & y_{j-1} \\ y_1 & y_2 & \cdots & y_j \\ \vdots & \vdots & \ddots & \vdots \\ y_{2i-1} & y_{2i} & \cdots & y_{2i+j-2} \end{bmatrix} \tag{3.9}$$

where it is assumed for statistical reasons that the number of columns $j \longrightarrow \infty$. As a shorthand notation, the following are employed throughout the thesis: $U_p \triangleq U_{0|i-1}$, $U_f \triangleq U_{i|2i-1}$, $Y_p \triangleq U_{0|i-1}$, and $Y_f \triangleq U_{i|2i-1}$. The subscripts p and f denote the past and the future respectively. An additional definition required is the past input-output matrix:

$$W_p \triangleq \begin{bmatrix} Y_p \\ U_P \end{bmatrix} \tag{3.10}$$

- The block Hankel matrices formed by the process and measurement noises are defined in a similar way as the input-output Hankel matrices and employ the same shorthand notation: $M_p \triangleq M_{0|i-1}$, $M_f \triangleq M_{i|2i-1}$, $N_p \triangleq N_{0|i-1}$, and $N_f \triangleq N_{i|2i-1}$.

$$M_{0|2i-1} \triangleq \frac{1}{\sqrt{j}} \begin{bmatrix} v_0 & v_1 & \cdots & v_{j-1} \\ v_1 & v_2 & \cdots & v_j \\ \vdots & \vdots & \ddots & \vdots \\ v_{2i-1} & v_{2i} & \cdots & v_{2i+j-2} \end{bmatrix} \tag{3.11}$$

$$N_{0|2i-1} \triangleq \frac{1}{\sqrt{j}} \begin{bmatrix} w_0 & w_1 & \cdots & w_{j-1} \\ w_1 & w_2 & \cdots & w_j \\ \vdots & \vdots & \ddots & \vdots \\ w_{2i-1} & w_{2i} & \cdots & w_{2i+j-2} \end{bmatrix} \tag{3.12}$$

- X_i is the state sequence

$$X_i \triangleq \begin{bmatrix} x_i & x_{i+1} & x_{i+2} & \cdots & x_{i+j-1} \end{bmatrix} \tag{3.13}$$

All subspace algorithms comprise two basic steps. The first step involves the weighted projection of the row space of the previously defined data Hankel matrices. Using this projection and the observability matrix Γ_i an estimate of the state sequence $\tilde{X}_i$ can be calculated. The second step consists in calculating the systems matrices A, B, C, D, Q, R and S. The first step is common to all algorithms while the second step can be different. Some algorithms use the observability matrix to calculate the system matrices, while others use the state sequence estimate; here only the methods using the state sequence are described.

3.1.1.1 Projection

Let $\mathcal{O}_i$ be the oblique projection of the row space of Y_f along the row space of U_f over the row space of W_P, defined as in equation (3.14).

$$\mathcal{O}_i \triangleq Y_f /_{U_f} W_p \tag{3.14}$$

and the singular value decomposition:

$$\begin{aligned} W_1 \mathcal{O}_i W_2 &= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \\ &= U_1 S_1 V_1^T \end{aligned} \tag{3.15}$$

Then the following statements hold (van Overschee and De Moor, 1995, 1996*b*):

1. The matrix $\mathcal{O}_i$ is equal to:

$$\mathcal{O}_i \triangleq \Gamma_i \tilde{X}_i \tag{3.16}$$

2. The extended observability matrix is equal to:

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2} T \tag{3.17}$$

where T is a similarity transformation.

3. The state sequence $\tilde{X}_i$ is equal to:

$$\tilde{X}_i = \Gamma_i^{\dagger} \mathcal{O}_i \tag{3.18}$$

Two important observations which are specific to identification problems are:

1. Optimal prediction: It is possible to obtain an optimal prediction of the future outputs by a linear combination of the past inputs and outputs and future inputs by using equation (3.19). The quality of the prediction is measured in the Frobenius norm presented in equation (3.20).

$$Y_f = L_w W_p + L_u U_f \tag{3.19}$$

$$\min_{L_w, L_u} \left\| Y_f - \begin{pmatrix} L_w & L_u \end{pmatrix} \begin{pmatrix} W_p \\ U_f \end{pmatrix} \right\|_F^2 \tag{3.20}$$

where the Frobenius norm of a matrix A is defined as $\|A\|_F^2 = \sqrt{\Sigma_i \Sigma_j \left|a_{ij}\right|^2}$.

2. Complexity reduction: This is achieved by reducing the subspace dimension to n (the order of the resulting system). The weighting matrices W_1 and W_2 help to determine what information of $\mathcal{O}_i$ is important, by examining the $rank(R)$, where R is defined as in equation (3.21).

$$R = W_1^{-1} U_1 S_1 V_1^T W_2^{\dagger} \tag{3.21}$$

The correct selection of weights W_1 and W_2 allows the algorithm to collapse into CVA, MOESP or N4SID as shown in Table (3.1). It is also possible to define other algorithms using this approach, see van Overschee and De Moor (1995); Favoreel *et al.* (2000) for more details. Also, for a particular choice of W_2, it is shown by van Overschee and De Moor (1996*b*) that $\tilde{X}_i$ is a Kalman filter estimate of X_i.

Table 3.1: Algorithms according to weight selection

Acronym	W_1	W_2
CVA	$\left[\left(Y_f/U_f^{\perp}\right)\cdot\left(Y_f/U_f^{\perp}\right)^T\right]^{-1/2}$	$U_f^{\perp T}\left(U_f^{\perp}U_f^{\perp T}\right)^{\dagger}U_f^{\perp}$
MOESP	I_l	$U_f^{\perp T}\left(U_f^{\perp}U_f^{\perp T}\right)^{\dagger}U_f^{\perp}$
N4SID	I_{li}	I_j

3.1.1.2 System matrices calculation

For those methods which use the state sequence (CVA, N4SID), the system matrices can be obtained by solving the over determined system of equation (3.22) in a least-squares sense.

$$\begin{bmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} \tilde{X}_i \\ U_{i|i} \end{bmatrix} + \begin{bmatrix} \rho_w \\ \rho_v \end{bmatrix} \tag{3.22}$$

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \frac{1}{j}\left[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix}\begin{pmatrix} \rho_w^T & \rho_s^T \end{pmatrix}\right] \geq 0 \tag{3.23}$$

where ρ are the residuals produced due to the approximation. van Overschee and De Moor (1996*b*, p.131), present a robust algorithm which has been tested in several indsutrial applications. As indicated in the reference, much of the symmetry and simplicity of the procedure described before is unfortunately lost, however, the algorithm gives excellent results in several practical situations.

An important feature desired in these methods is the stability of the A matrix. It is possible to obtain stable matrices from the extended observability matrix Γ_i, however, guaranteed stability methods such as the one presented by Chui and Maciejowski (1996) should be used carefully, since even unstable systems will be identified as stable.

The robust algorithm of van Overschee and De Moor (1996*b*, p.131), and its MATLAB implementation provided in a toolbox with the book has been used for many of the case studies examined in this thesis. Also, a stand-alone module, based on this toolbox, has been developed to run as a dynamic link library and will be presented in Chapter 7. The following is a summary of the algorithm.

Algorithm 3.1.1. Robust combined algorithm

1. Calculate the oblique projection:

   $$\mathcal{O}_i = Y_f / {}_{U_f} W_p$$

2. Calculate the weighted singular value decomposition

   $$W_1 \mathcal{O}_i W_2 = U_1 S_1 V_1^T$$

3. Calculate the extended observability matrix

   $$\Gamma_i = W_1^{-1} U_1 S_1^{1/2}$$

4. Compute the state sequence

   $$\tilde{X}_i = \Gamma_i^{\dagger} \mathcal{O}_i$$

5. Solve the following set of equations in the least square sense to find the system matrices,

$$\begin{bmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \tilde{X}_i \\ U_{i|i} \end{bmatrix} + \begin{bmatrix} \rho_w \\ \rho_v \end{bmatrix}$$

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \frac{1}{j} \left[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \begin{pmatrix} \rho_w^T & \rho_s^T \end{pmatrix} \right] \geq 0$$

3.2 Identification of dissolved oxygen

This work employs subspace identification as the method to generate models for prediction and estimation for dissolved oxygen control. These algorithms allow the identification of multivariable systems, therefore the use of subspace identification for the identification of dissolved oxygen dynamics for univariate and multivariable cases is explored throughly. The COST Simulation Benchmark is employed throughout this section, as described in Chapter 2.

3.2.1 Univariate identification

For this study assume that only the last reactor of the simulation benchmark allows manipulation of the air compressor. Figure (3.1) shows a diagram of the control structure in the last aeration tank. As described in section 3.1.1, in order to obtain a representative model of the dynamics, it is necessary to persistently excite the system. Pseudo-random binary sequences (PRBS) is a common signal employed to excite systems. This type of signal switches between two levels at random discrete points in time. Several authors point out that PRBS might have some shortcomings for particular applications. Tulleken (1990) demonstrates that PRBS gives more emphasis to high frequency components when used in conjuction with prediction error methods (PEMs), and proposes the use of a generalised binary noise signal (GBN). In Godfrey (1993), it is stated that large magnitude PRBS signals applied to non-linear systems may bias the estimation of the linear kernel, therefore a multi-level sequence (m-level) might be more suitable.

Some examples of the use of PRBS and m-level sequences applied to wastewater treatment plants prior to a subspace identification routine can be found in (Lindberg, 1997) and in (Sotomayor *et al.*, 2003) respectively. Both works report the excitation of the treatment plant by probing dissolved oxygen and other variables to identify nutrient dynamics (nitrate and ammonia). Given that the research of this work will investigate

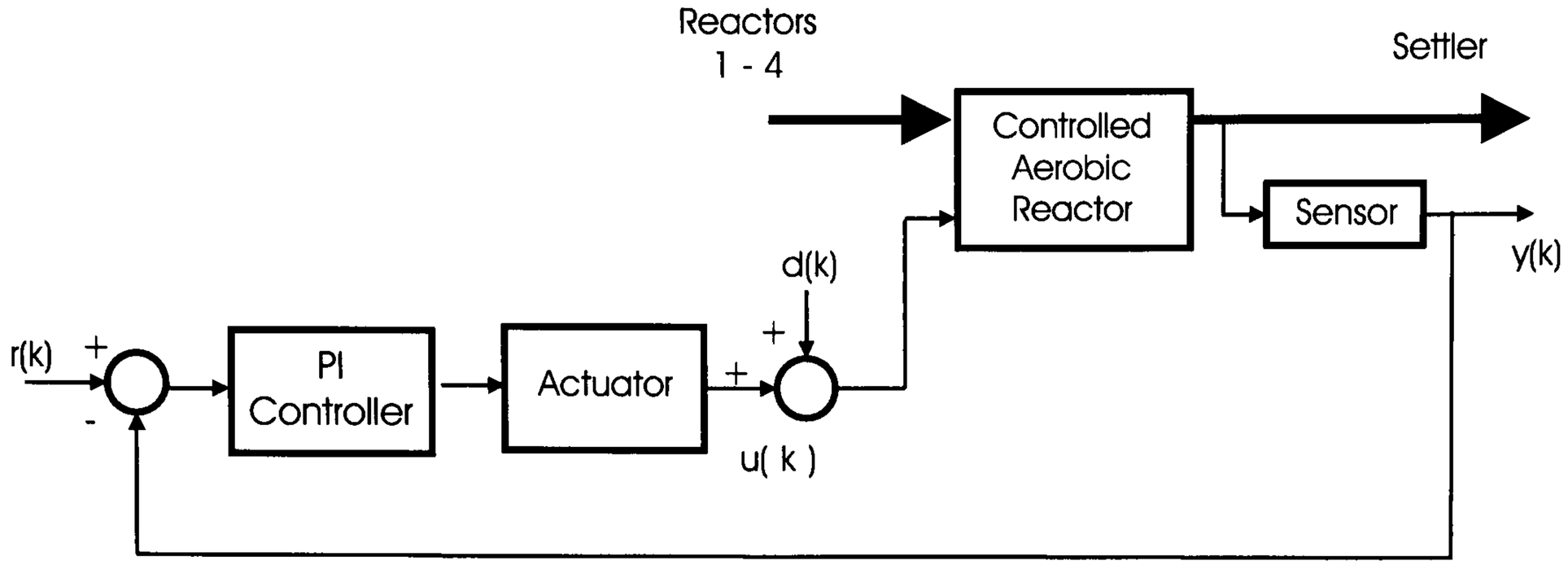

Figure 3.1: Dissolved oxygen control loop.

practical experimentation, excitation via PRBS, which has already been proven to be effective for industrial purposes, is employed.

3.2.1.1 Data collection

For the univariate case, the last reactor of the simulation benchmark was excited with a PRBS signal of 1 [mg/l] amplitude around an operating point of 2 [mg/l] during 1 day. All sensors and actuators are modelled as described in section 2.2.

Initially three models were identified using the algorithms in (van Overschee and De Moor, 1996*a*) and (van Overschee and De Moor, 1996*b*, p.131). The first model is an open-loop model (from *u(k)* to *y(k)*) identified from closed-loop data using the first algorithm presented in (van Overschee and De Moor, 1996*a*). The following two models are closed-loop models, identified from *r(k)* to *y(k)* using the *SV* and the *CVA* options of the combined deterministic-stochastic robust identification algorithm presented in van Overschee and De Moor (1996*b*). Table (3.2) presents a summary of the three identified models.

Table 3.2: Identified Models

No.	Model	Order	Algorithm	Notes
1	open loop	2	(van Overschee and De Moor, 1996*a*)	1st algorithm
2	closed-loop	3	(van Overschee and De Moor, 1996*b*, p.131)	SV
3	closed-loop	3	(van Overschee and De Moor, 1996*b*, p.131)	CVA

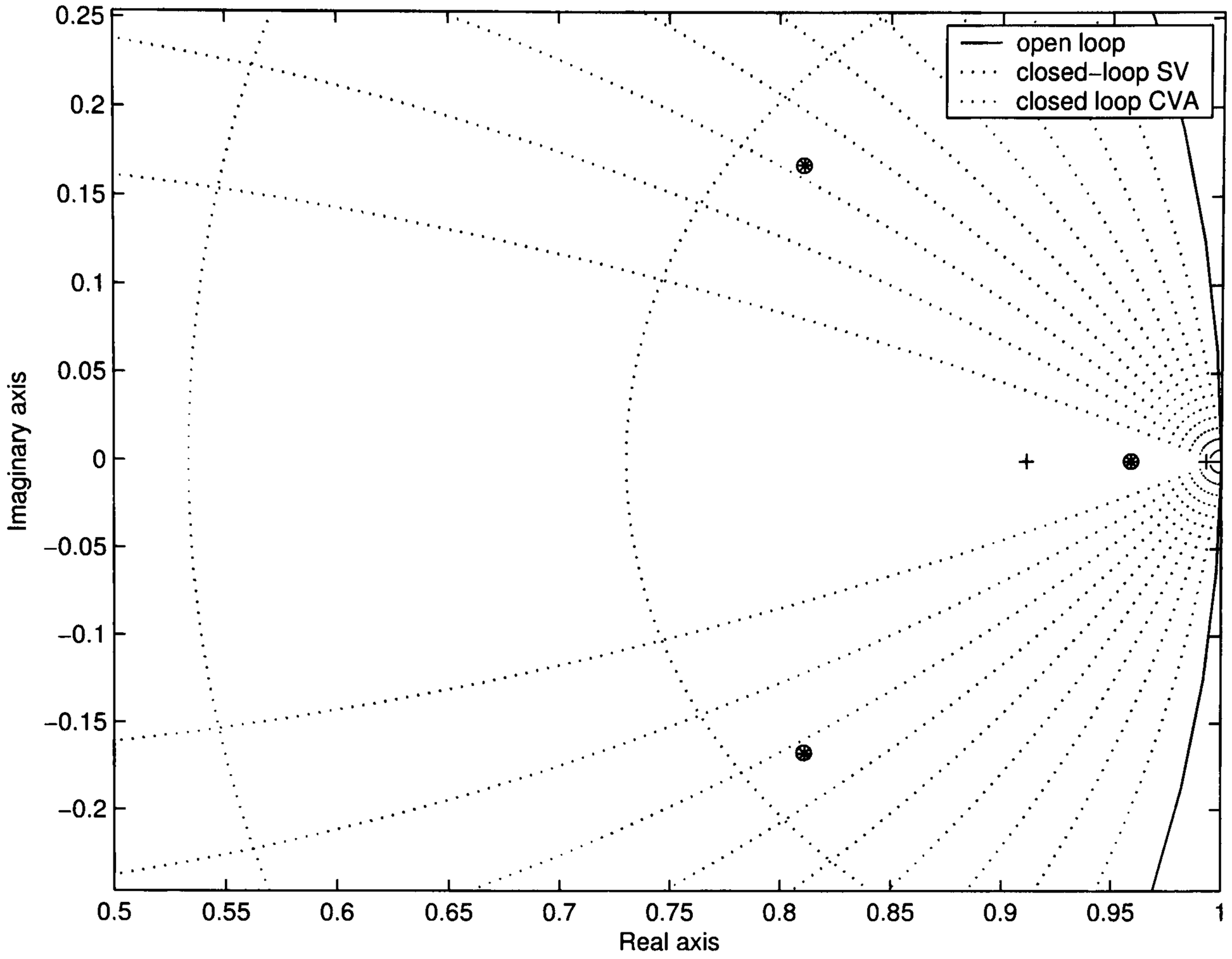

Figure 3.2: Pole location for the open-loop model (No.1) and closed-loop models (No.2 and 3).

The order of the models was chosen by examining at the singular values (SV case) and the principal angles (CVA case). This way of selecting the model order however, does not guarantee that the model will be stable. Therefore, it is necessary to check the eigenvalues of matrix A for modelling stability. There has been some work on how to generate stable systems. For example, Chui and Maciejowski (1996) propose an algorithm which generates either an asymptotic or a marginally stable system. The method consists of a data augmentation procedure to find stable approximations to least-squares problems. This method however will only be consistent if the process is known to be stable. Figure (3.2), shows the pole locations to be within the unit circle for the three models.

3.2.1.2 Actuator and disturbance models

None of the previously identified models include the actuator output signal $u(k)$. In order to include the physical limitations of the aeration system it is necessary to incorporate this signal into the model. Additionally, to compensate for the plant-model mismatch and un-modelled disturbances given by the changing influent load (daily variations and weather effects), it is necessary to introduce a disturbance model. This section discusses some ways in which it is possible include the actuator (controller) dynamics. This leads to the formulation of three new models which will be denoted as composite for the open-loop case and augmented for the closed-loop cases.

Composite and augmented models

Let the loop-controller be described by equation (3.24) and the plant model No.1 by equation (3.25). The loop-controller is of PID type controller represented in state-space.

$$\begin{aligned} x_c(k+1) &= A_c x_c(k) + B_c e(k) \\ u(k) &= C_c x_c(k) + D_c e(k) \end{aligned} \tag{3.24}$$

$$\begin{aligned} x_o(k+1) &= A_o x_o(k) + B_o u(k) \\ y(k) &= C_o x_o(k) + D_o u(k) \end{aligned} \tag{3.25}$$

where *e(k)* is the error between the oxygen measurement *y(k)* and the reference signal *r(k)*. Closing the loop and rearranging the matrices the following composite system is found:

$$\begin{aligned} X(k+1) &= A \cdot X(k) + B \cdot r(k) \\ Y(k) &= C \cdot X(k) + D \cdot r(k) \end{aligned} \tag{3.26}$$

where,

$$X(k) = \begin{bmatrix} x_o(k) \\ x_c(k) \end{bmatrix} \tag{3.27}$$

$$Y(k) = \begin{bmatrix} y(k) \\ u(k) \end{bmatrix} \tag{3.28}$$

$$A = \begin{bmatrix} A_o - B_o M D_o C_o & B_o M C_o \\ B_c D_o M D_c C_o - B_c C_o & A_c - B_c D_o M C_c \end{bmatrix} \tag{3.29}$$

$$B = \begin{bmatrix} B_o M D_c \\ B_c - B_c D_o M D_c \end{bmatrix} \tag{3.30}$$

$$C = \begin{bmatrix} C_o - D_o M D_c C_o & D_o M C_c \\ -M D_c C_o & M C_c \end{bmatrix} \tag{3.31}$$

$$D = \begin{bmatrix} D_o M D_c \\ M D_c \end{bmatrix} \tag{3.32}$$

$$M = (I + D_c D_o)^{-1} \tag{3.33}$$

For the case of the closed-loop models No.2 and 3, the controller dynamics and therefore the actuators limits are included in a different way. Considering the same loop-controller state-space representation and the closed-loop model of equation (3.34), and using the error definition it is possible to define an augmented model described by equations (3.35) to (3.41).

$$\begin{aligned} x_{cl}(k+1) &= A_{cl} x_{cl}(k) + B_{cl} r(k) \\ y(k) &= C_{cl} x_{cl}(k) + D_{cl} r(k) \end{aligned} \tag{3.34}$$

$$\begin{aligned} X(k+1) &= A \cdot X(k) + B \cdot r(k) \\ Y(k) &= C \cdot X(k) + D \cdot r(k) \end{aligned} \tag{3.35}$$

where,

$$X(k) = \begin{bmatrix} x_{cl}(k) \\ x_c(k) \end{bmatrix} \tag{3.36}$$

$$Y(k) = \begin{bmatrix} y(k) \\ u(k) \end{bmatrix} \tag{3.37}$$

$$A = \begin{bmatrix} A_{cl} & 0 \\ -B_c T_2 C_{cl} & A_c \end{bmatrix} \tag{3.38}$$

$$B = \begin{bmatrix} B_{cl} \\ B_c (T_1 - T_2 D_{cl}) \end{bmatrix} \tag{3.39}$$

$$C = \begin{bmatrix} C_{cl} & 0 \\ -B_c T_2 C_{cl} & C_c \end{bmatrix} \tag{3.40}$$

$$D = \begin{bmatrix} D_{cl} \\ D_c (T_1 - T_2 D_{cl}) \end{bmatrix} \tag{3.41}$$

The matrices T_1 and T_2 are chosen such that they map the output of the plant into the input of the controller so $e(k) = T_1 r(k) - T_2 Y(k)$, where $e(k)$ is a vector of error signals.

Validation of these models is performed by measuring the *percentage variance accounted for* (vaf) between the measured and predicted signals and defined in equation (3.42). The *vaf* coefficient is only a measure of the degree of similarity between the two signals and does not measure biases. Table (3.3) presents a summary of the obtained *vaf* coefficients for the oxygen concentration and actuator output predictions. Figures (3.3) and (3.4) show the signals of the three models validated against the DO measurement generated from the plant model.

$$vaf\% = \left[1 - \frac{var(y - \hat{y})}{var(y)}\right] * 100 \tag{3.42}$$

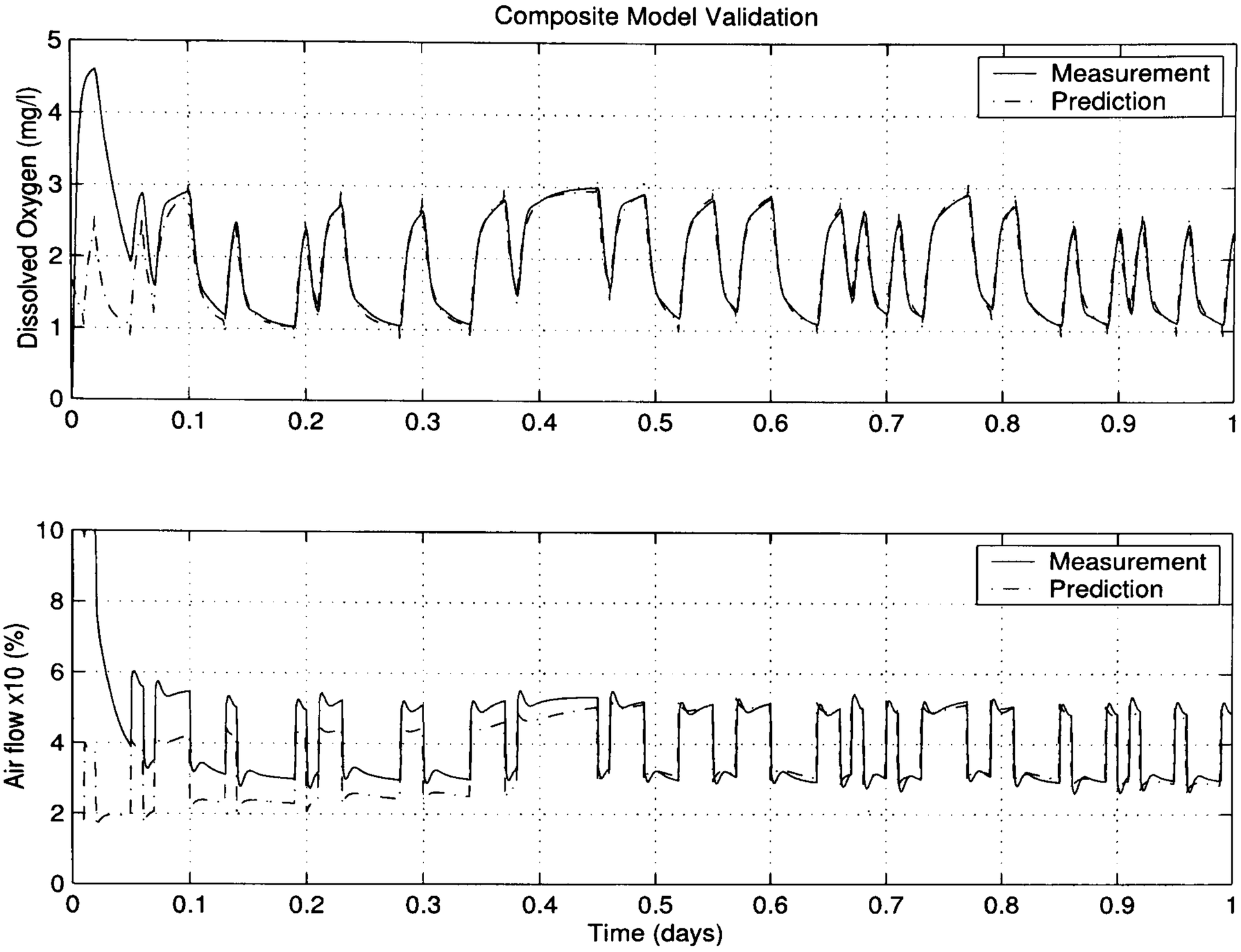

Figure 3.3: Composite model prediction

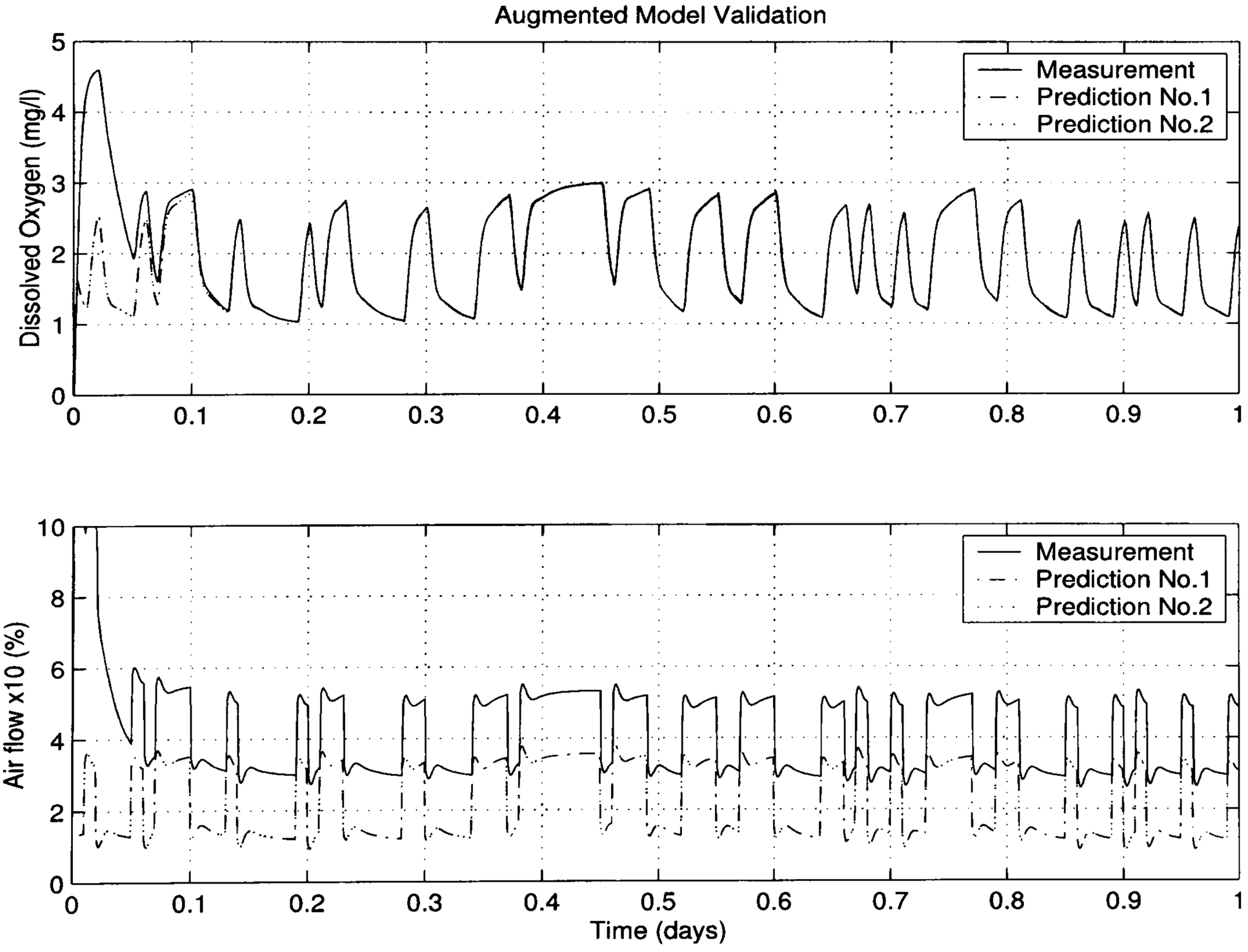

Figure 3.4: Augmented models predictions

Table 3.3: VAF coefficients for composite and augmented models

Model	$\hat{y}$ *vaf* (%)	$\hat{u}$ *vaf* (%)
composite	96.809	81.9695
augmented 1	99.95716	86.55406
augmented 2	99.95720	86.55406

Unmeasurable disturbance model

There are two important reasons to include an unmeasurable disturbance model in the DO dynamics model: (a) To compensate for changing load conditions due to influent concentrations and flow variations during the day and in meteorological events as rain or storm. (b) The models which are being used have been recovered from an identification procedure. Therefore, they are just an approximation to the real plant dynamics. Due to this plant-model mismatch, the augmented and the composite models can give significant errors in the prediction of the actuator (controller) output, as can be seen in Figure (3.4) for the augmented models.

Notice in Figure (3.4) that the measurement and the prediction of both identified models differ mainly in a bias. This bias is produced by the amplification of the small error of the prediction of $y(k)$ by the integrating effect of the controller.

In general, disturbance models have to be chosen accordingly to the expected load. It could be argued that in the case of WWTPs, the most common disturbance will have a cyclic daily fluctuation. However, the prediction horizon of a variable like dissolved oxygen is in the range of fractions of an hour. Therefore it is more realistic to assume either a constant disturbance or a slowly decaying model (Lindberg, 1997).

To introduce the disturbance effect into the composite and the augmented model formulations it is only necessary to redefine the state and output equation as in equations (3.43) to (3.49).

$$\begin{aligned} \xi(k+1) &= A_d\xi(k)+B_d r(k) \\ Y(k) &= C_d\xi(k)+D_d r(k) \end{aligned} \tag{3.43}$$

where,

$$\xi(k) = \begin{bmatrix} X(k) \\ d(k) \end{bmatrix} \tag{3.44}$$

$$d(k) = Y(k) - (CX(k) + Dr(k)) \tag{3.45}$$

$$A_d = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \tag{3.46}$$

$$B_d = \begin{bmatrix} B \\ 0 \end{bmatrix} \tag{3.47}$$

$$C_d = \begin{bmatrix} C & I \end{bmatrix} \tag{3.48}$$

$$D_d = D \tag{3.49}$$

Notice however, that this approach can only be implemented in conjunction with a state-observer which provides the initial plant-model mismatch with which it is possible to calculate the predictions. Therefore in order to evaluate these models it is necessary to implement them through state observers or estimators.

3.2.1.3 State observers

A fundamental part of the design of a MPC controller involves the design of a state observer or estimator. In the case of deterministic systems the most common approach is to design a state observer if the system is observable. If the system is of stochastic nature, the optimal solution would be a Kalman filter. However, in this process as in many industrial processes the noise characteristics are not known. Due the lack of this information it might be time consuming and difficult to calibrate a Kalman filter compared to a state observer. Therefore, this thesis does not address the advantages or disadvantages of using Kalman filtering. Instead state observers designed using pole placement are employed. Due to the high similarity between the augmented models, the CVA case model will be used as default throughout the rest of this section.

State observers can be described, in a general way, by equations (3.50) and (3.51).

$$\hat{x}(k+1) = [A-LC]\hat{x}(k) + \begin{bmatrix} B-LD & L \end{bmatrix} \begin{bmatrix} u(k) \\ y(k) \end{bmatrix} \tag{3.50}$$

$$\begin{bmatrix} \hat{y}(k) \\ \hat{x}(k) \end{bmatrix} = \begin{bmatrix} C \\ I \end{bmatrix} \hat{x}(k) + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u(k) \\ y(k) \end{bmatrix} \tag{3.51}$$

Table (3.4), presents the chosen observer gains for the augmented and composite cases with the inclusion of the disturbance models. These gains were chosen such that the slowest pole of the observer was at least 5 times faster than the slowest pole of the model. These are not necessarily the most effective gains, but they have produced satisfactory results.

Table 3.4: Observer gains

Model	L
composite	$\begin{bmatrix} 0.70 & 0.66 & 0.55 & 0.59 & 0.60 \end{bmatrix}$
augmented	$\begin{bmatrix} 0.80 & 0.77 & 0.81 & 0.85 & 0.60 & 0.50 \end{bmatrix}$

The response of these observers are presented in Figures (3.5) and (3.6). As can be seen the bias in the augmented model is corrected, and the observations of the outputs converge very fast to the measured signals. Table (3.5) contains the vaf coefficients of the observed signals for the composite and augmented models with disturbance.

Table 3.5: vaf coefficients of observed signals

Model	$\hat{y}$ vaf(%)	$\hat{u}$ vaf(%)
composite w. disturbance	96.51	70.61
augmented w. disturbance	99.42	82.18

3.2.2 Multivariable identification

Consider that the plant has three independent PI controllers regulating the dissolved oxygen concentration in reactors 3, 4, and 5 as presented in Figure (3.7). Each control

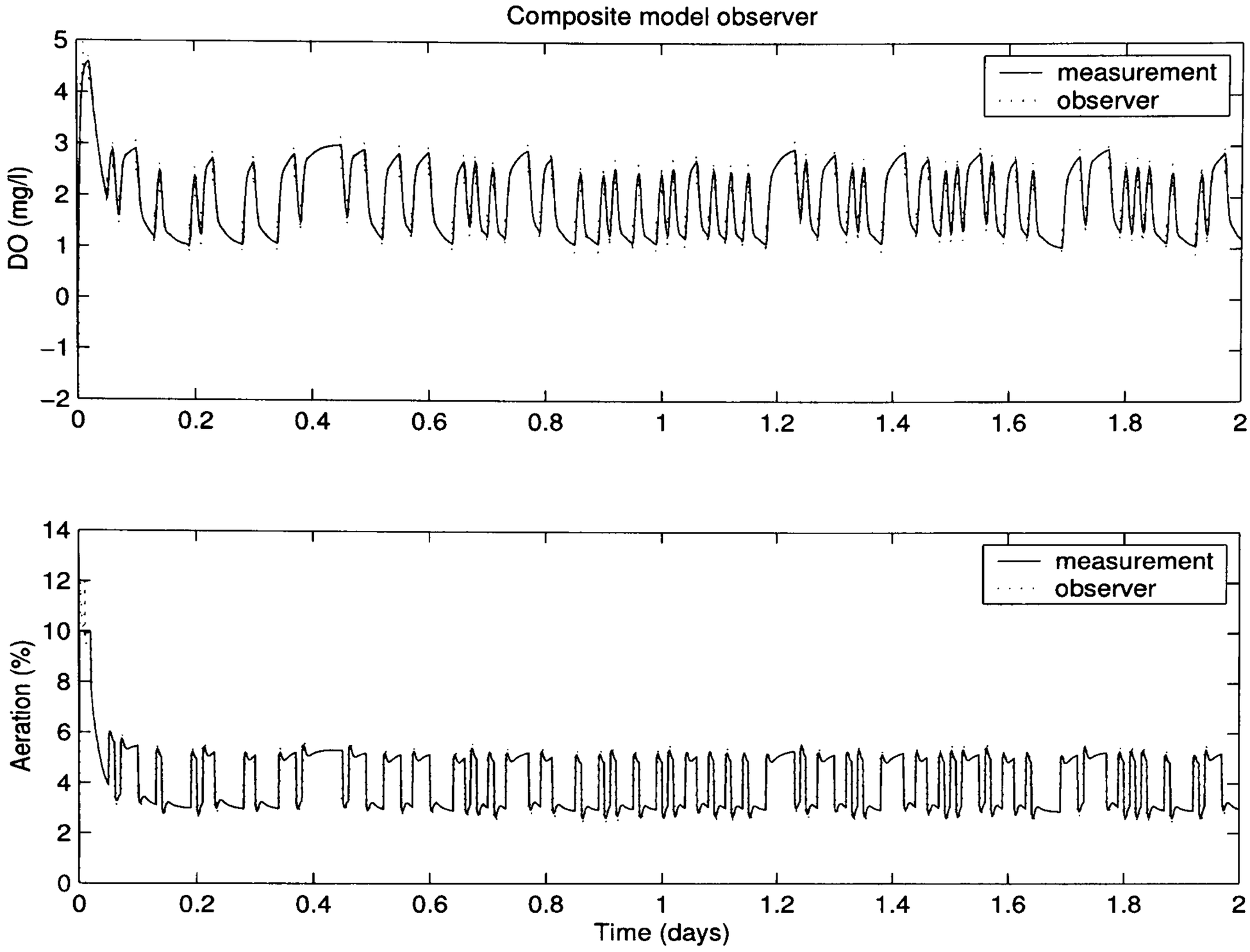

Figure 3.5: Composite model with disturbance observer

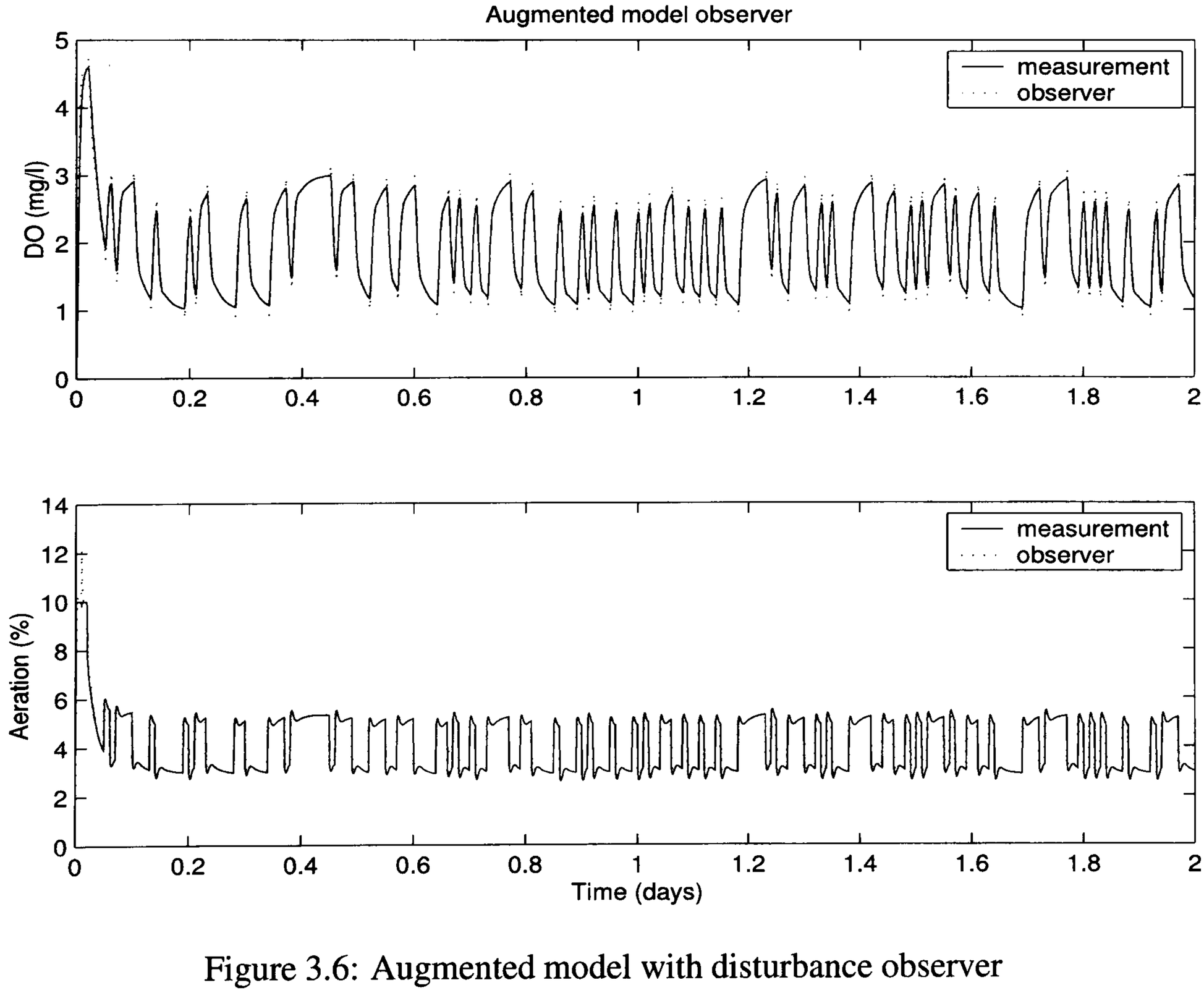

Figure 3.6: Augmented model with disturbance observer

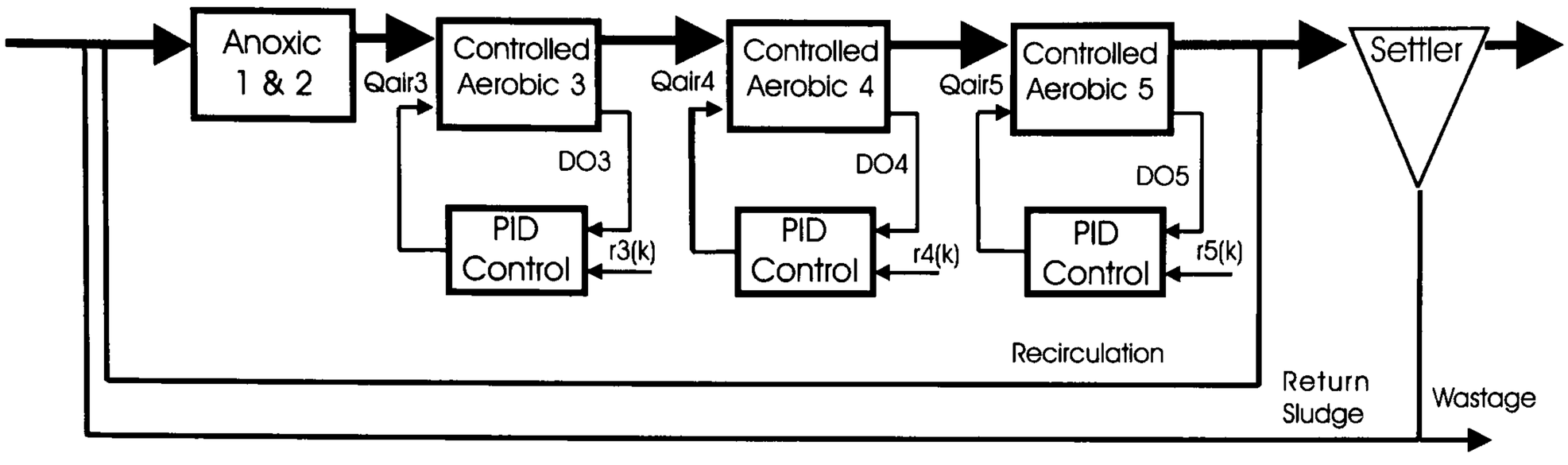

Figure 3.7: PI controllers arrangement

loop has the same structure as previously presented in Figure (3.1). This section follows the same approach and sequence as the previous with some differences due to the multivariable nature of the problem. Open-loop and closed-loop models are identified using CVA, MOESP and N4SID algorithms. Table (3.6), summarises the algorithms employed for the closed-loop and open-loop identification.

Table 3.6: Subspace algorithms for multivariable DO identification

Model No.	Algorithm	Acronym	Model type
composite 1	van Overschee and De Moor (1996*a*)	N4SID	open-loop
composite 2	Verhaegen (1993)	MOESP	open-loop
augmented 1	van Overschee and De Moor (1996*b*)	Robust N4SID 'SV' based	closed-loop
augmented 2	van Overschee and De Moor (1996*b*)	Robust N4SID 'CVA' based	closed-loop

3.2.2.1 Data collection

The data required for the identification of the multivariable model for dissolved oxygen is performed by persistently exciting the inputs simultaneously with three different PRBS signals with the characteristics presented in Table (3.7).

Table 3.7: PRBS input signals characteristics

Reactor	Mean (mg/l)	Amplitude (mg/l)
3	1.5	1
4	1.5	1
5	2	1

The mean value of the PRBS signal applied to the 5^{th} reactor is higher than that of

Table 3.8: Multivariable DO model validation (vaf coefficients)

Model	Order	DO3 *vaf %*	DO4 *vaf %*	DO5 *vaf %*	Q3 *vaf %*	Q4 *vaf %*	Q5 *vaf %*
composite 1	4(1)	41.89	73.93	82.53	63.76	82.19	78.26
	5(2)	43.68	76.56	97.55	62.20	82.59	82.47
composite 2	4(1)	39.77	75.34	84.04	2.90	-37.61	-58.09
	5(2)	48.99	81.10	96.04	64.27	85.35	84.59
augmented 1	9(6)	96.81	98.57	99.47	56.58	81.96	86.28
	6(3)	83.36	93.60	97.30	54.84	70.20	-39.40
	13(10)	99.89	99.18	99.70	60.78	83.21	77.23
augmented 2	9(6)	95.31	98.16	99.42	47.76	78.56	81.33
	6(3)	84.53	93.40	97.53	40.46	-1.06×10^2	-1.72×10^2
	15(12)	99.89	99.18	99.75	61.03	83.74	88.54

Note: The number in brackets is the order of the identified model, thus the order of the full system is that of the identified model plus the controller order (3).

the others. This is because a larger amount of air is required to be pumped into the first reactors to achieve the same oxygen concentration, and therefore more energy is required. This effect is caused by the oxygen-free wastewater incoming from the anoxic zone of the process. This extra effort by the aeration can lead to saturation of the actuators, and therefore the system would be operating in a non-linear region.

The system is excited over a period of 2 days (2880 data points) under constant influent conditions. Of the 2880 data points only those corresponding to the last day are employed for the identification. The remaining 1440 data points are used for validation.

Several tests were carried out to find an appropriate number of block rows in the Hankel matrices with the purpose of finding a good model without unnecessarily increasing the computation burden. It was found that the best number of block rows was 30. Table (3.8) gives a summary of some of the most significant results of several identification runs with the same data and with the listed algorithms. Note as well that composite and augmented models follow the same description as in section 3.2.1.2.

Results presented in Table (3.8), provide some interesting insight into the algorithms and their behaviour. The open-loop models (No.1 and No.2), could not be approximated to a higher order than 2, without producing an unstable realization. For the

Table 3.9: Models for control

Model	Order	Acronym
composite 1	5(2)	comp1-5
composite 2	5(2)	comp2-5
augmented 1	9(6)	aug1-9
	13(10)	aug1-13
augmented 2	9(6)	aug2-9
	15(12)	aug2-15

composite model No.2, if a first order model was employed, a bias was produced in the airflow signals. In general composite models No.1 and No.2 do not produce a good approximation of the DO signals (very low vaf coefficients).

With respect to the augmented models, as the order increased so did the vaf coefficients, therefore giving a better approximation. Figures (3.8) to (3.11), show the signals trends for all the identified models as presented in Table (3.8).

It is evident that only certain composite and augmented models produce a sufficiently good approximation to the system dynamics in order to be used in a model predictive control scheme. Therefore, only the selected models listed in Table (3.9) will be used in the following sections.

3.2.2.2 State observers

The design of the state observers is based on the models identified in the previous section. Additionally, the constant disturbance model introduced in section 3.2.1.2 has also been used for each of the cases. Therefore, a total of 12 observers (one for each model) have been designed by conventional pole placement. In addition, this section also examines the behaviour of the designed observers when the influent flow is of a dry weather type, instead of a constant influent as has been assumed so far. This will provide a more realistic assessment of the performance of the observers.

To simplify the presentation of results, the acronyms for the models defined in Table (3.9) are used. When the model also includes a disturbance model, then a *d* will be

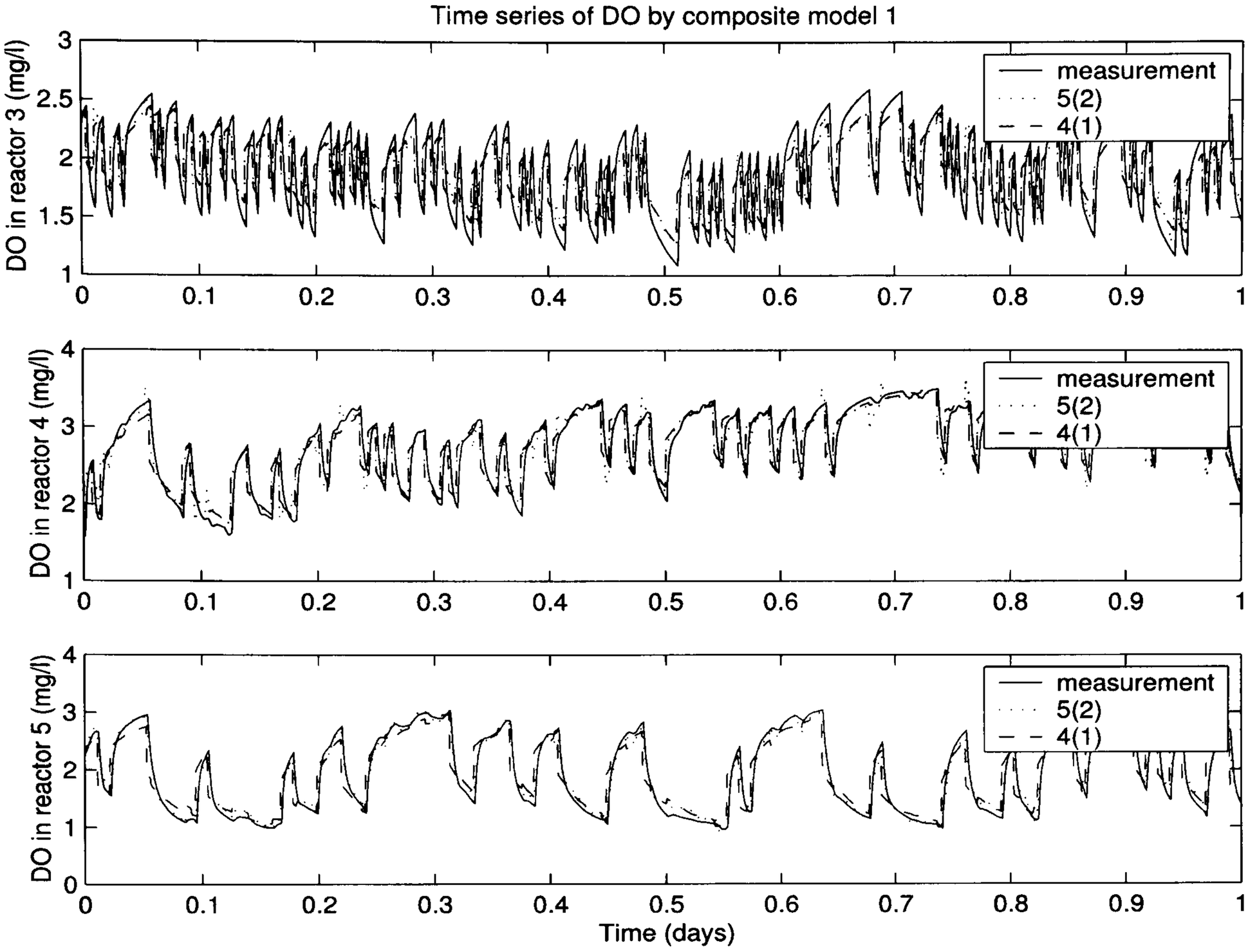

Figure 3.8: Composite model No.1 validation trends

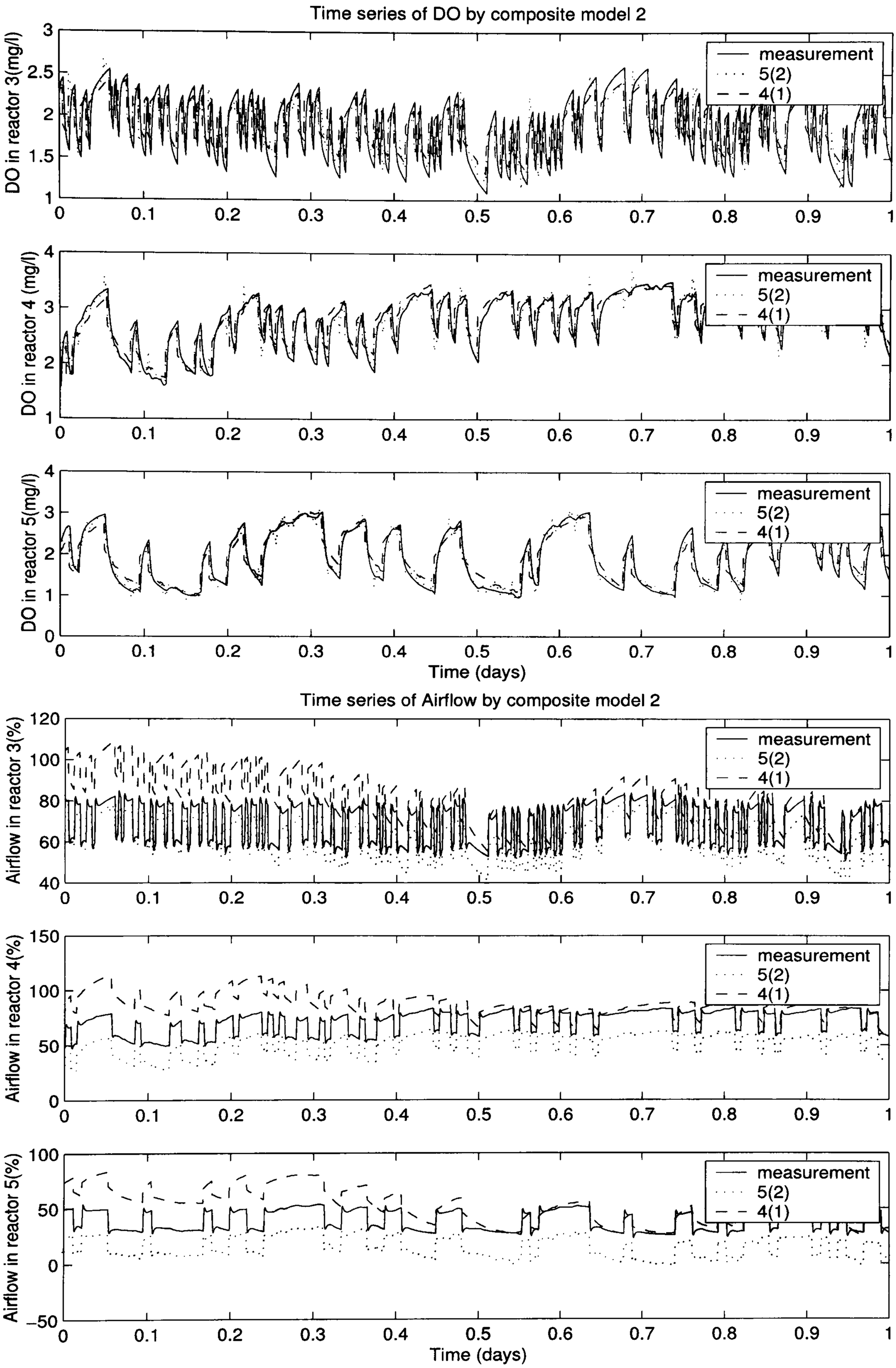

Figure 3.9: Composite model No.2 validation trends

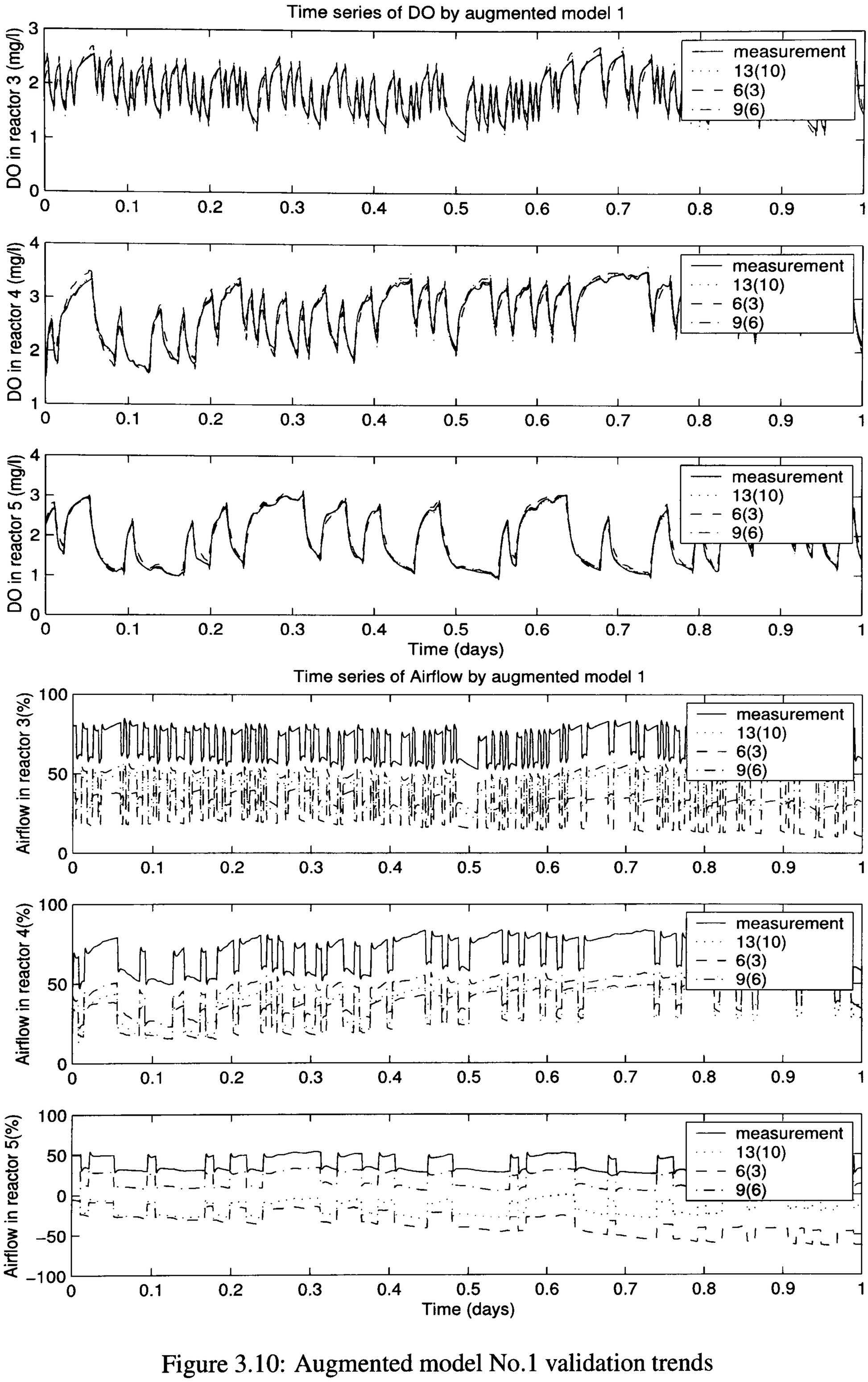

Figure 3.10: Augmented model No.1 validation trends

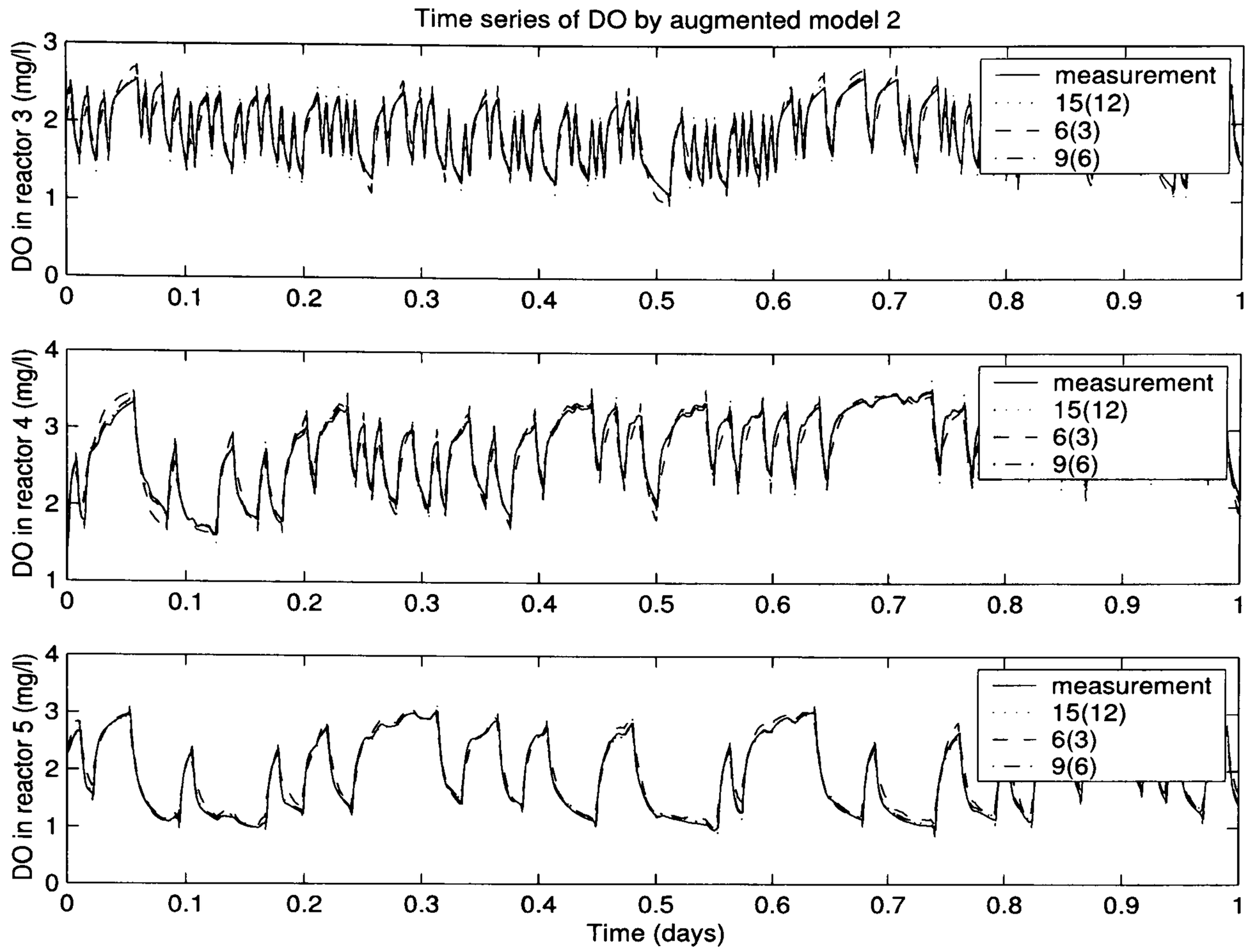

Figure 3.11: Augmented model No.2 validation trends

added to the acronym. For example, composite model No.1, which is of fifth order, will be abbreviated as comp1_5.

Table 3.10: Multivariable DO observers performance (constant influent)

Model Observer	DO3 vaf %	DO4 vaf %	DO5 vaf %	Q3 vaf %	Q4 vaf %	Q5 vaf %	Overall vaf %
comp1-5	65.139	64.008	67.428	55.536	80.636	67.435	66.697
comp1-5d	60.359	81.834	97.444	63.734	85.528	85.492	79.065
comp2-5	69.289	72.253	86.499	39.506	75.271	75.059	69.646
comp2-5d	66.804	86.768	97.086	55.018	84.634	82.721	78.838
aug1-9	95.880	98.929	99.617	43.499	77.098	82.711	82.956
aug1-9d	97.988	99.412	99.800	40.025	77.715	82.807	82.958
aug1-13	39.944	73.734	55.937	34.749	71.373	73.931	58.278
aug1-13d	78.477	97.375	98.993	53.662	80.880	83.201	82.099
aug2-9	95.420	98.234	99.511	44.990	77.147	83.113	83.0696
aug2-9d	97.522	98.892	99.631	42.881	77.335	83.888	83.358
aug2-15	66.068	-74.171	276.699	32.195	39.532	-101.754	–
aug2-15d	82.532	62.258	34.386	39.918	75.173	70.542	60.802

According to the results presented in Table (3.10), note that the composite model observers with a disturbance correction improve their performance significantly. The augmented model observers with disturbance correction also show improvement over their counterpart without disturbance correction; however, it is not as significant as in the composite case. The observers for higher order models do not show a significant improvement over their lower order versions.

Table 3.11: Multivariable DO observers performance (dry weather influent)

Model Observer	DO3 vaf %	DO4 vaf %	DO5 vaf %	Q3 vaf %	Q4 vaf %	Q5 vaf %	Overall vaf %
aug1-9	96.067	98.247	99.169	64.382	89.952	93.093	90.152
aug1-9d	98.036	99.027	99.353	62.165	90.183	93.089	90.308
aug1-13	28.514	61.705	38.887	34.749	86.960	90.413	60.764
aug1-13d	80.294	96.249	98.287	70.746	91.636	93.314	88.421
aug2-9	93.603	94.324	97.048	65.097	89.767	93.005	88.808
aug2-9d	97.402	98.245	98.955	63.950	89.988	93.456	90.335

Table (3.11) presents results of the state observers but working under dry-weather influent. The table however, only presents results for the 6 best observers from the constant influent case. In general, it appears that the observers with disturbance correction

appear to perform even better under dry-weather conditions, achieving in some cases vaf coefficients over 90%. Figures (3.12) and (3.13) present the responses for the observers aug1-9d and aug2-9d compared with the measurement for a 1 day simulation time.

This section has addressed the model identification of a multivariable DO model, and it has achieved the same conclusion as in the univariate case. Augmented models perform better than composite models identified using closed-loop subspace identification algorithms, for prediction and state observation. Therefore, these models, in particular the one identified with the robust N4SID using CVA, will be employed in Chapter 4, for the design of model predictive controllers.

3.3 Modelling and identification of nitrogen

This section focuses on the modelling and identification of nitrogen in an activated sludge wastewater treatment plant. The first task is the the development of a model for an alternating aeration plug-flow treatment plant under constant influent conditions. This case has been simplified to a single bioreactor modelled by the full ASM No.1. The second task considers the identification of nutrients (NO_3, NH_4 and total N) in the COST Simulation Benchmark, under dry-weather conditions.

Subspace algorithms continue to be the preferred choice for the identification of nutrient dynamics. However, only open-loop algorithms are employed. This is because there is usually no common control algorithm employed for nutrient control, apart from probably nitrate recirculation if the plant is properly equipped for this facility. Most commonly nutrient sensors, if available, are used only for monitoring.

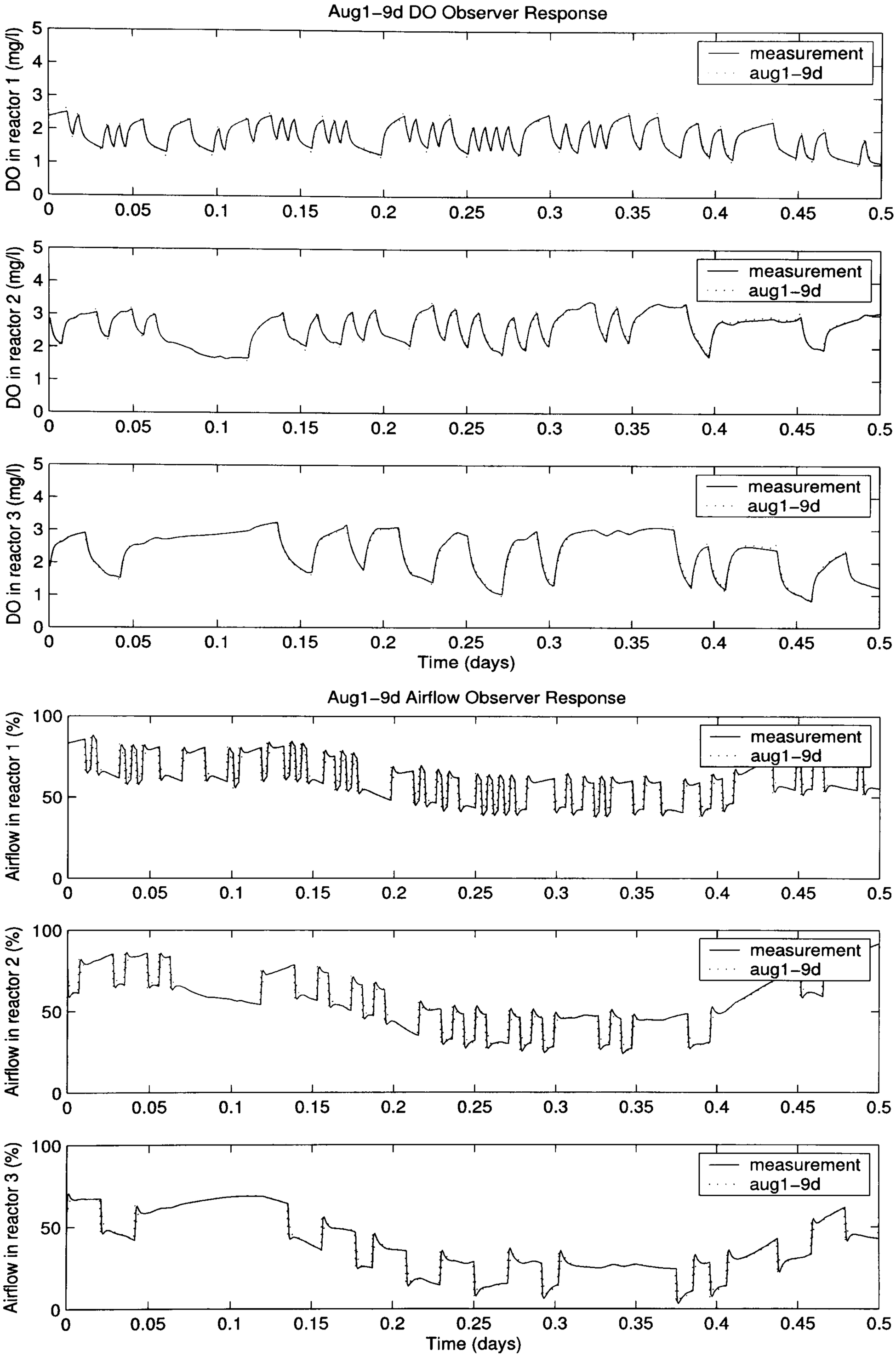

Figure 3.12: Augmented model No.1 observer with disturbance correction

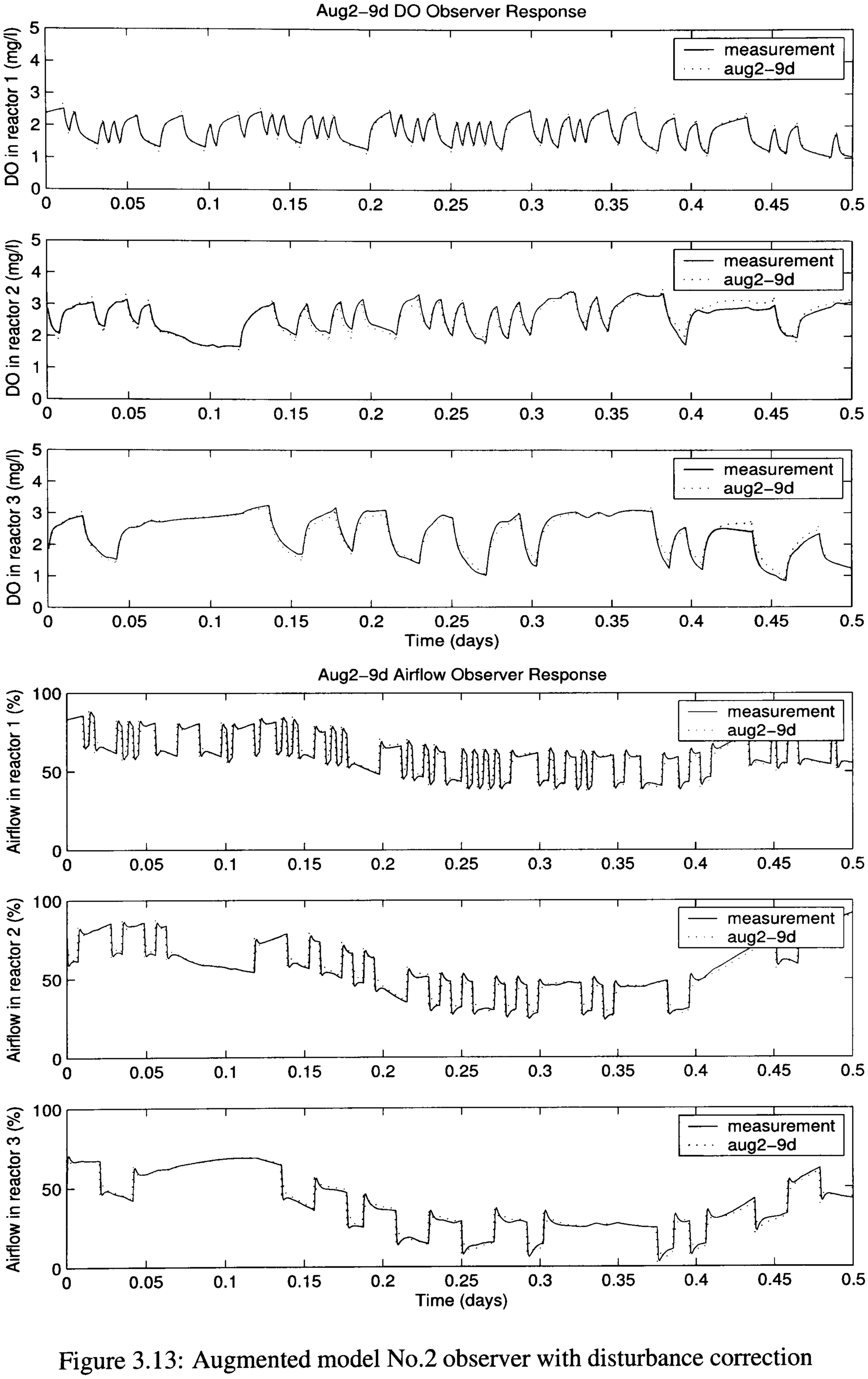

Figure 3.13: Augmented model No.2 observer with disturbance correction

3.3.1 Alternating aeration modelling

In an activated sludge process, different biochemical reactions occur. These processes can be classified into two different categories: aerobic and anoxic reactions. Aerobic reactions make use of the oxygen dissolved in the water body and the two main biochemical phenomena are the oxidation of the carbonaceous material and nitrification. Under anoxic conditions, denitrification reactions are predominant. These reactions make use of nitrate as the oxidation agent, instead of oxygen, to produce free nitrogen and other compounds.

Anderson *et al.* (2000) presented two linear approximations to the ASM No.1, one for each phase: aerobic and anoxic. These models were obtained by approximating the half saturation non-linear terms to linear terms. The model is also of reduced order, which is accomplished by omitting: soluble inert organic matter (S_I) and particulate inert organic matter (X_I), (which are decoupled from the system); dissolved oxygen (S_o), (which is assumed to be controlled); alkalinity (S_{alk}), (since denitrification can partially recover some alkalinity consumed through nitrification); and the growth of particulate products (X_P), which does not interact with the other variables. In addition, dissolved oxygen is not considered to be a limiting factor during the aerobic process and to be totally absent during the anoxic phase. The final system is composed of two state space representations, one for each phase, of eighth order, with the following form:

$$\dot{x}(t) = A_e x(t) + D_e x_{inf} \tag{3.52}$$

$$\dot{x}(t) = A_a x(t) + D_a x_{inf}$$

In equation (3.52), A_e and D_e denote the system matrix representation for the aerobic phase and, A_a and D_a are the system matrix representations for the anoxic phase. The variable $x(t)$ is the state vector of the system whose components are specified in equation (3.53), and x_{inf} is the corresponding vector of influent characteristics (con-

centrations) into the system, for each state variable. The last term in equation (3.52) should be considered as a measurable disturbance, since there is no possible control over the influent characteristics.

$$x(t) = \begin{bmatrix} S_s & X_s & X_{B,H} & X_{B,A} & S_{N,H} & S_{NO} & S_{ND} & X_{ND} \end{bmatrix}^T \tag{3.53}$$

Since the control principle of this type of structure demands the switching between the two models at a given frequency and duty cycle, an appropriate model representation for the system behaviour over the entire time domain is required. Equation (3.54) shows the proposed model representation.

$$\begin{aligned} \dot{x}(t) &= [g(t,\delta)\cdot A_e + (1-g(t,\delta))\cdot A_a]x(t) \\ &\quad + [g(t,\delta)\cdot D_e + (1-g(t,\delta))\cdot D_a]x_{inf} \end{aligned} \tag{3.54}$$

The switching function $g(t,\delta)$ represents a train of width modulated pulses of unit amplitude as shown in Figure (3.14). When the switching function is unity, the aerobic phase is said to be ON, and when it is zero the aerobic phase is said to be OFF. The switching function depends on time and duty cycle 'δ', which is defined as the relation between the time the aerobic phase is ON (T_{ON}), and the switching period ($T_{switching}$), as presented in equation (3.55).

$$\delta = \frac{T_{ON}}{T_{switching}} \tag{3.55}$$

For mathematical simulation, $g(t,\delta)$ can be expanded into a finite Fourier series, where the Gibbs phenomenon is eliminated by the use of a saturation function in the convergence point of the discontinuities. Equation (3.56) denotes the truncated Fourier Series expansion of $g(t,\delta)$.

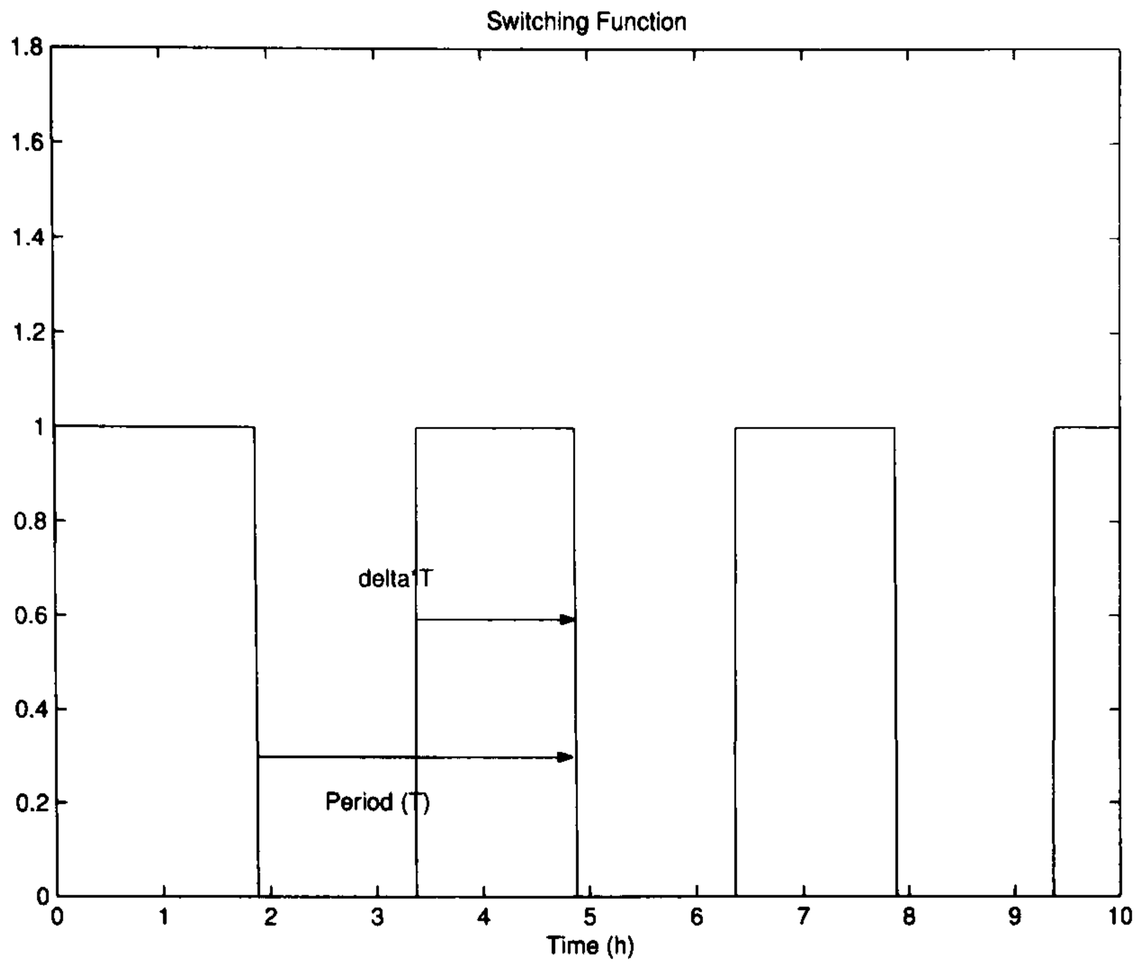

Figure 3.14: Switching Function.

$$g(t,\delta) = \delta + \sum_{n=1}^{N} \frac{1}{n\pi}[\sin(2n\pi\delta)\cdot cos(2n\pi f_o t) + (1-\cos(2n\pi\delta))\cdot sin(2n\pi f_o t)] \quad (3.56)$$

Equation (3.54) can be rearranged to give equation (3.57).

$$\dot{x}(t) = [A_1\cdot g(t,\delta)+A_2]x(t) + [D_1\cdot g(t,\delta)+D_2]x_{inf} \quad (3.57)$$

where $A_1 = A_e - A_a$, $A_2 = A_a$, $D_1 = D_e - D_a$ and $D_2 = D_a$. By substituting equation (3.56) into (3.57) and considering that the expressions contained within the brackets are time dependent, equation (3.57) can be rewritten as a time variant system presented in (3.58). This equation is an approximation, since $g(t,\delta)$ is a truncated series.

$$\dot{x}(t) \approx A(t,\delta)\cdot x(t) + D(t,\delta)\cdot x_{inf} \quad (3.58)$$
$$y(t) = C\cdot x(t)$$

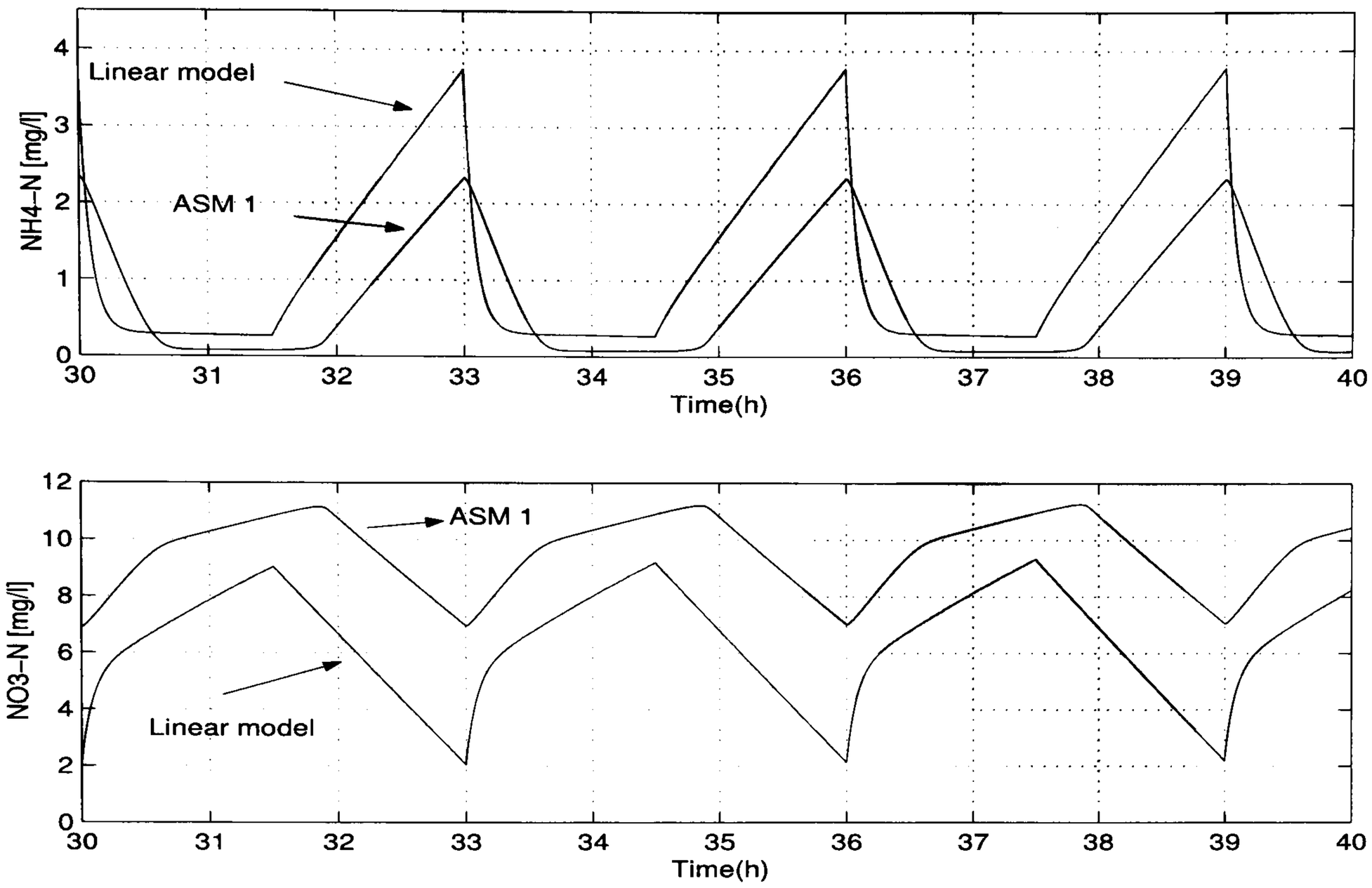

Figure 3.15: Linear Time-Variant model and ASM No.1.

3.3.1.1 Model validation

In order to verify this model, simulations have been performed using the parameter data presented in the original paper of Anderson *et al.* (2000), and comparing it with a simulation of the full ASM No.1 also using the Anderson *et al.* parameter data. The simulations were performed using MATLAB for a switching period of 3 hours and a duty cycle of 50 % (δ=0.5). Results of these two simulations are presented in Figure (3.15). In the linear model, the number of terms used for equation (3.56) was of N=5.

The simulations presented in Figure (3.15) are identical to those presented in Anderson *et al.* (2000). It can be clearly observed that there is a model mismatch, but the trends are the same. The model mismatch could be corrected by using an off-line or on-line parameter estimation as in Jeppsson (1996); however, there are several limitations over the applicability of estimators for ASM No.1 reduced-order models due to absence of enough on-line measurements.

3.3.2 Continuous aeration identification

This section focuses on the identification of a dynamic model for nutrients in an activated sludge wastewater treatment plant. Due to the high cost of nutrient sensors, most plants possess only a reduced number. Therefore, the location of the sensors is of crucial importance, and depends on many factors related to the control strategy, see for example (Ingildsen, 2002) for a brief discussion on nutrient sensor locations and their effects on the control strategy.

Experience shows that in a nitrification-denitrification plant with recirculation it is possible to control the effluent concentrations of ammonia (NH_4), nitrate (NO_3) and total nitrogen (TN) by manipulating the oxygen concentration in the aerated zones. Additionally, the nitrate concentration in the anoxic zone can be controlled using the internal recirculation flow rate, with the purpose of providing sufficient nitrate for the denitrification process. It is also of common experience, that the influent flow and concentration of ammonia (plant load) will have a significant impact over the process. Therefore, to improve the identifiability and controllability of the system, the use of a flow meter and an ammonia meter in the influent is encouraged.

There are two previous studies in identifying black-box models for nutrient control using subspace identification. In (Lindberg, 1997), a state-space model is identified with the purpose of controlling nitrate and ammonia in the effluent by manipulating the dissolved oxygen setpoint, the internal recirculation flow and an external carbon source flow rate. The exercise includes an ammonia and nitrate sensor at the effluent; and ammonia, flow and biodegradable substrate meters at the influent. The simulation plant employed differs in size (capacity) with the COST simulation benchmark employed in this study. A similar identification exercise is reported in (Sotomayor *et al.*, 2003). It differs, from the previous work, in using the model to control nitrate in the anoxic zone and in the effluent by manipulating the recirculation rate and an external carbon source dosing. Similarly, it measures flow, ammonia and biodegradable substrate in the inflow

and uses them as measurable disturbances.

Both of these previous works look at two complementary control problems. Lindberg, identifies the model with the purpose of effluent control of nitrate and ammonia concentrations, while Sotomayor *et al.*, identifies the model with the purpose of nitrate control in the anoxic zone and at the effluent. In both cases, as well, the assumption of an influent substrate measure is unrealistic in practice, since chemical oxygen demand (COD) meters are very unreliable, at the present time, to be used in an on-line control scheme. The fact that both plants use an external carbon source is also very unlikely. An international survey by Ingildsen (2002), revealed that only 1 treatment plant out of 36 plants in 10 different countries employ an external carbon source.

In this study, a model for effluent ammonia, nitrate, total nitrogen and nitrate in the anoxic zone is identified with the primary objective of controlling the effluent discharge of ammonia and total nitrogen. To achieve this purpose, the plant is assumed to have sensors located as in Figure (3.16). It is also assumed that the only control handles are the oxygen setpoints in the three aerated zones and the internal recirculation flow rate. Therefore the results presented in this study are considered to be of much more practical value, since they do not include the unrealistic possibility of having a COD meter in the inlet flow, and an unlikely external carbon dosing. Also, a deterministic model for prediction of influent flow and influent ammonia is also presented at the end of the section.

3.3.2.1 Identification

The measurements of the nutrient sensors located at the end of the aeration tanks and the nitrate sensor at the end of the anoxic zone are employed as outputs for the identification. The inputs are the dissolved oxygen measurements in the aerobic reactors, and the internal recirculation flow rate. Influent flow and ammonia are considered as measurable disturbances. The signals used for the identification procedure are summarised

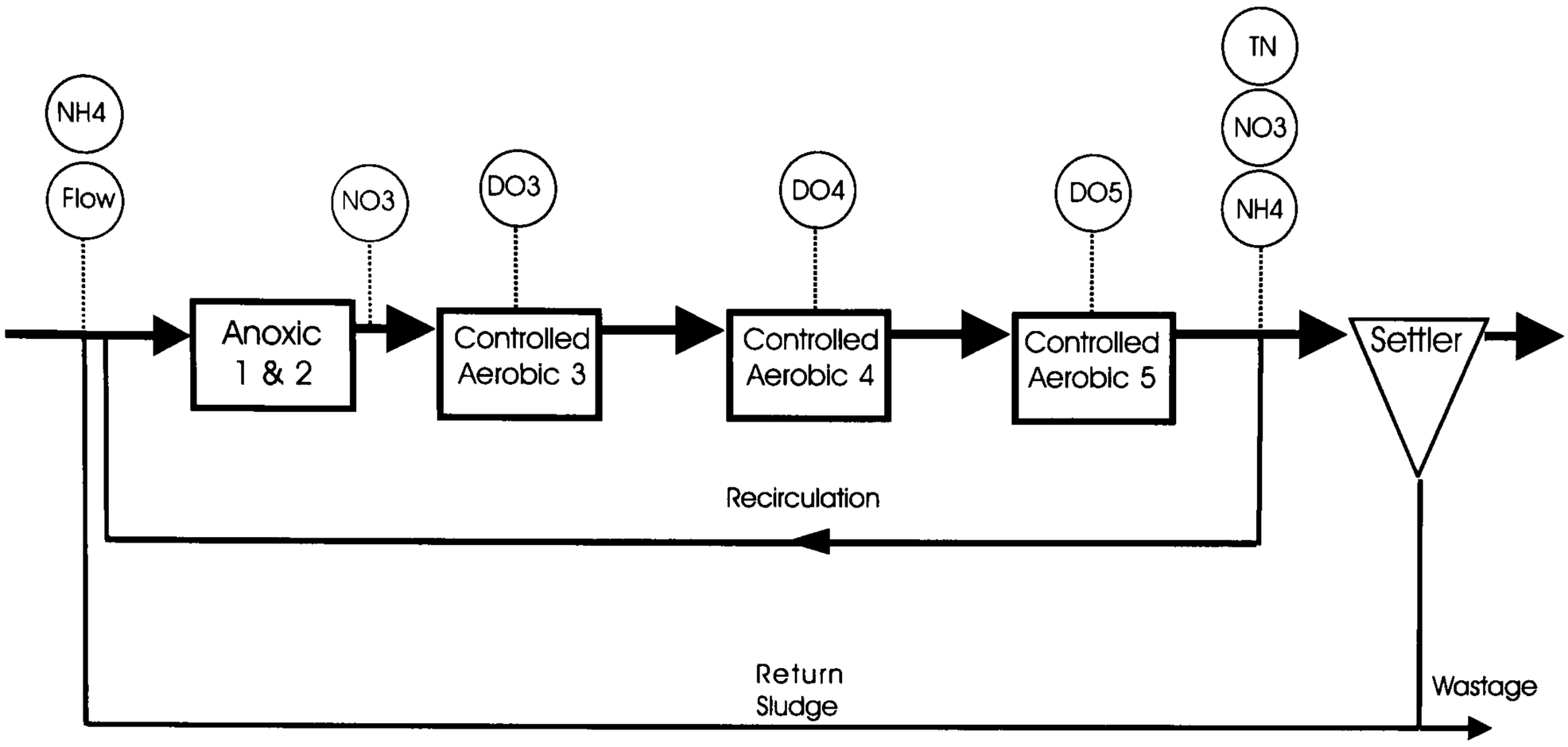

Figure 3.16: Sensor location in the COST Benchmark.

in Figure (3.17).

There are several differences between this approach and the identification of dissolved oxygen (DO). The DO layer is considered to be totally controlled, which implies that there is at least one level of control over the DO regulation. The actual DO control implementation has two layers, first a PID control for each reactor, and then a multi-variable model predictive controller (MPC) guiding all of the three reactors. In addition to the DO controllers, there is the internal recirculation controller, which is assumed to be instantaneous. Therefore, from the point of view of the nutrient control level, all the input signals are controlled or controllable. A second difference is that there is information of influent disturbances, thus providing the capability of implementing a feedforward control strategy.

Since there is actually no possibility of control over the influent flow and influent ammonia, these signals have to be considered as measurable disturbances, and can be included in the system state-space description as in equations (3.59) and (3.60).

$$x(k+1) = A \cdot x(k) + B \cdot u(k) + B_d d_m(k) \tag{3.59}$$

$$y(k) = C \cdot x(k) + D \cdot u(k) + D_d d_m(k) \tag{3.60}$$

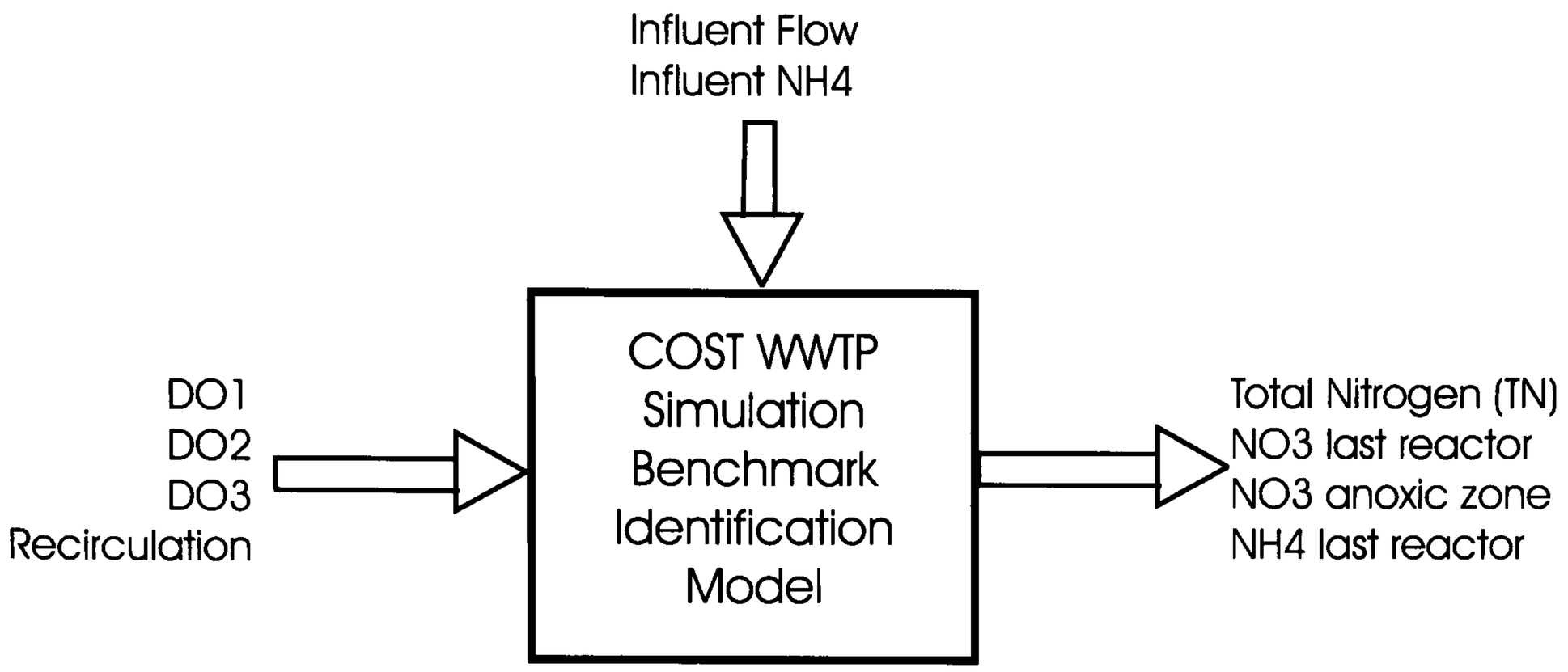

Figure 3.17: Signals for nutrient identification.

where

$$y(k) = \begin{bmatrix} TN & NO_3 & NO_{3_{anoxic}} & NH_4 \end{bmatrix}^T$$

$$u(k) = \begin{bmatrix} DO_{reac3} & DO_{reac4} & DO_{reac5} & Recirculation \end{bmatrix}^T$$

$$d_m(k) = \begin{bmatrix} Flow & NH_{4_{influent}} \end{bmatrix}^T$$

Note, that the flow measurement must be appropriately scaled in order to numerically balance the model. The model can be re-arranged in the following manner:

$$x(k+1) = A \cdot x(k) + \begin{bmatrix} B & B_d \end{bmatrix} \begin{bmatrix} u(k) \\ d_m(k) \end{bmatrix} \tag{3.61}$$

$$y(k) = C \cdot x(k) + \begin{bmatrix} D & D_d \end{bmatrix} \begin{bmatrix} u(k) \\ d_m(k) \end{bmatrix} \tag{3.62}$$

Data collection

Data has been collected for a period of 7 days under a semi-constant flow. The oxygen setpoints in the three reactors were excited with adequate PRBS signals so they could excite the nutrient variables in their time scale. Samples were taken every 15 minutes,

with the sensors modelled as described in Chapter 2.

Three open-loop subspace identification algorithms are employed. These are the Robust N4SID 'CVA', and 'SV' algorithms, and the common N4SID algorithm. Table (3.12) presents the results obtained by using these algorithms. Additionally, Figures (3.18), (3.19), and (3.20) show the response of the models for a period of 7 days compared to the measurement from the plant. Note that, as in the dissolved oxygen identification case, the Robust N4SID algorithms perform considerably better than the conventional N4SID.

Table 3.12: Nutrient identified models

Model	Order	Algorithm	TN vaf %	$NO_{3_{effluent}}$ vaf %	$NO_{3_{anoxic}}$ vaf %	NH_4 vaf %	Overall vaf %
model 1	4	Robust 'SV'	69.66	94.94	95.96	82.83	85.85
model 2	3	Robust 'CVA'	–	50.57	70.45	70.27	–
model 3	4	N4SID	40.49	84.17	93.68	72.98	72.83

Table 3.13: Nutrient observers

Model	Algorithm	TN vaf %	$NO_{3_{effluent}}$ vaf %	$NO_{3_{anoxic}}$ vaf %	NH_4 vaf %	Overall vaf %
Observer 1	Robust 'SV'	71.60	78.77	98.04	94.29	85.68
Observer 2	Robust 'CVA'	73.15	57.25	92.52	82.62	76.39
Observer 3	N4SID	73.01	80.16	98.28	92.46	85.98

3.3.2.2 Observer design

Observers for these three models have been designed by following again the prescription of section 3.2.1.3 and including the unmeasurable disturbance correction as in section 3.2.1.2. Thus, this section will only present the results obtained by simulation of the designed observers. Table (3.13) shows the relative performance of each observer.

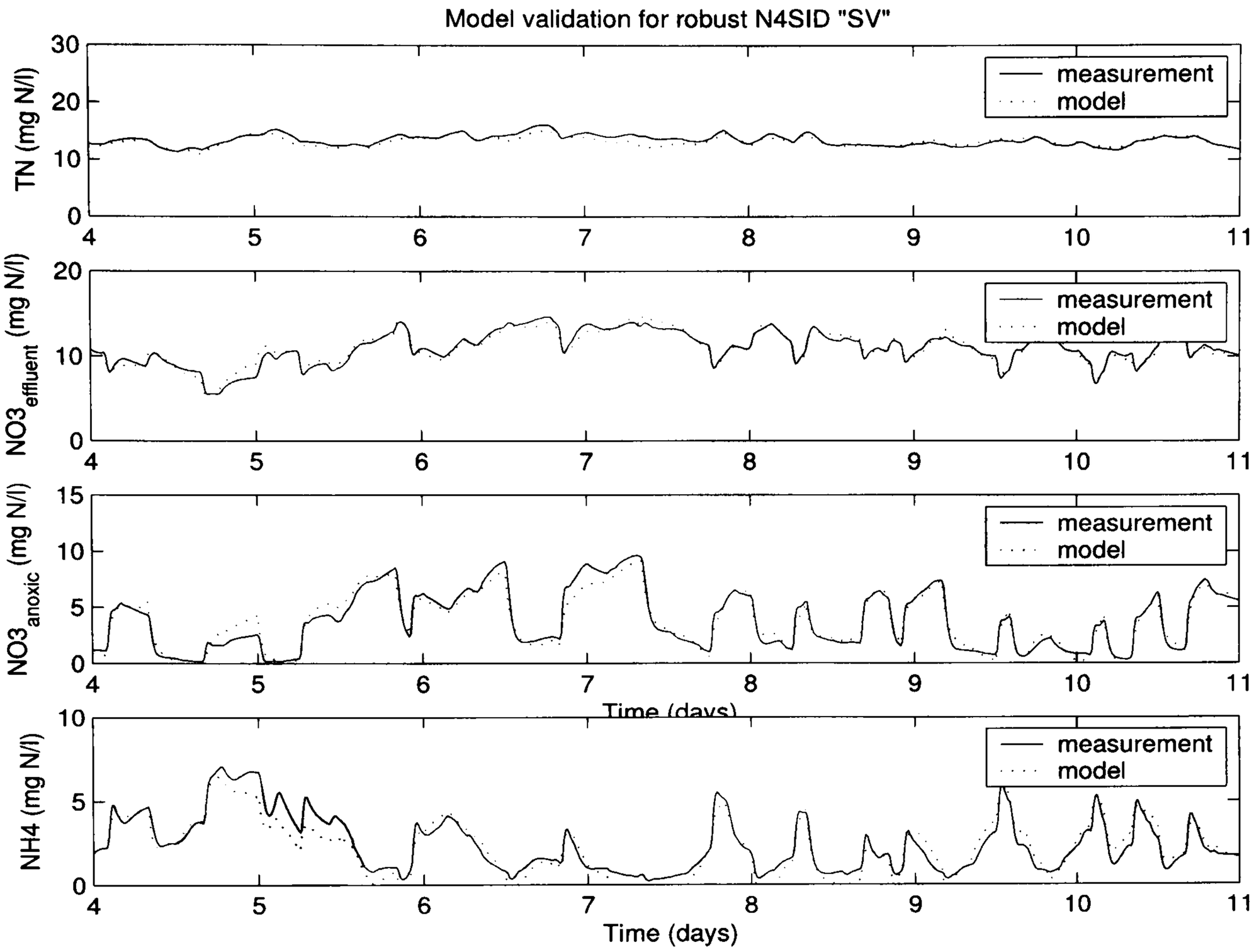

Figure 3.18: Nutrient model identified using Robust N4SID 'SV' algorithm

3.3.2.3 Influent flow and ammonia prediction model

The model developed in this section, has assumed that the influent flow and ammonia concentration are measurable disturbances. Since the purpose of this model is to be used in a MPC controller structure, it might of interest to have some additional information about these disturbances for prediction. In (Nielsen, 2001, 2002), a method to approximate the influent flow by a truncated Fourier series (FS) with three components: a mean value and the first two harmonics, is proposed. However, for simulation, it might be useful to expand this series to include higher frequency components, as follows,

$$\begin{aligned} Flow(k) &= a_o + a_1 cos(w_1 \cdot Ts \cdot k) + b_1 sin(w_1 \cdot Ts \cdot k) + \cdots \\ &\quad + a_n cos(w_n \cdot Ts \cdot k) + b_n sin(w_n \cdot Ts \cdot k) \end{aligned} \tag{3.63}$$

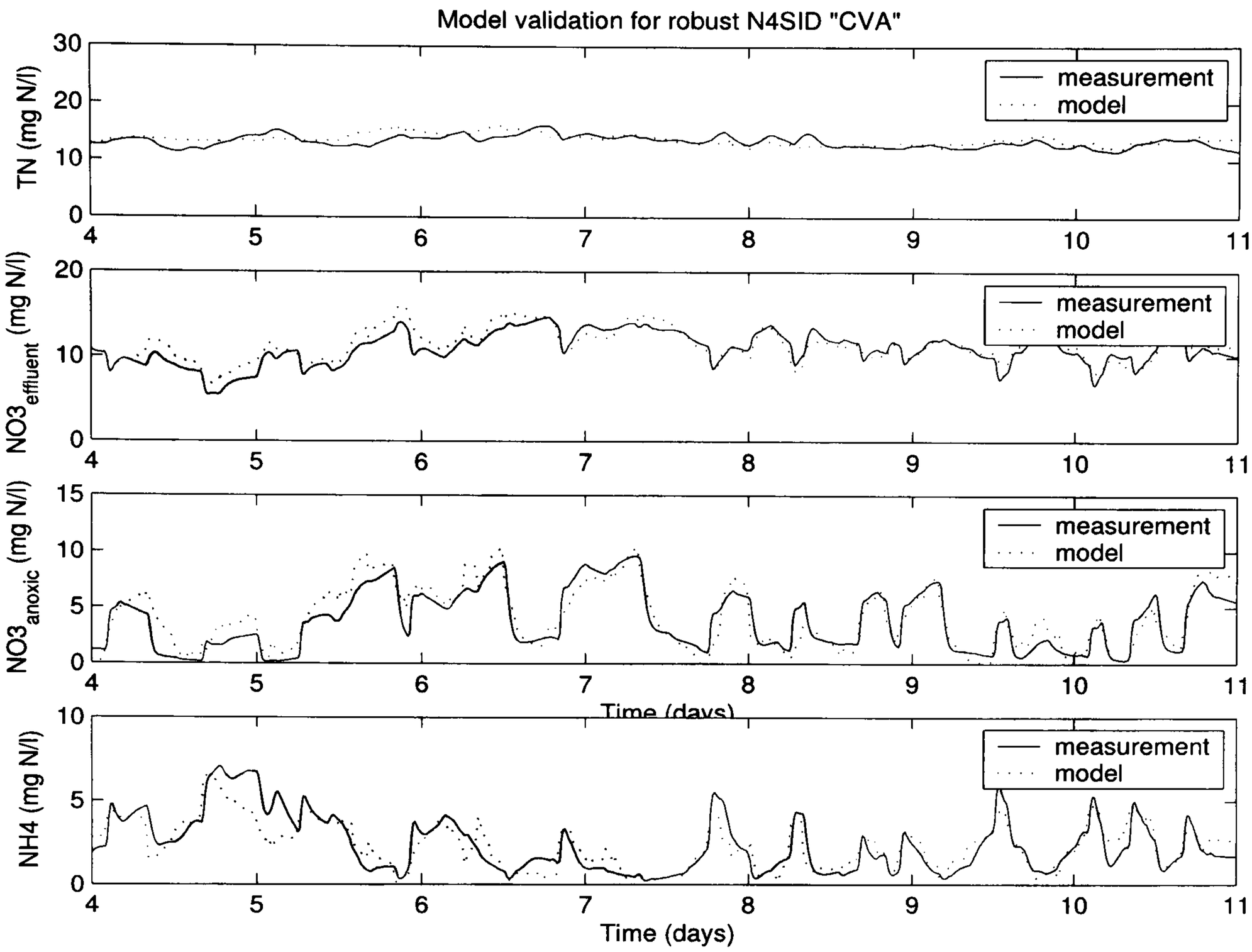

Figure 3.19: Nutrient model identified using Robust N4SID 'CVA' algorithm

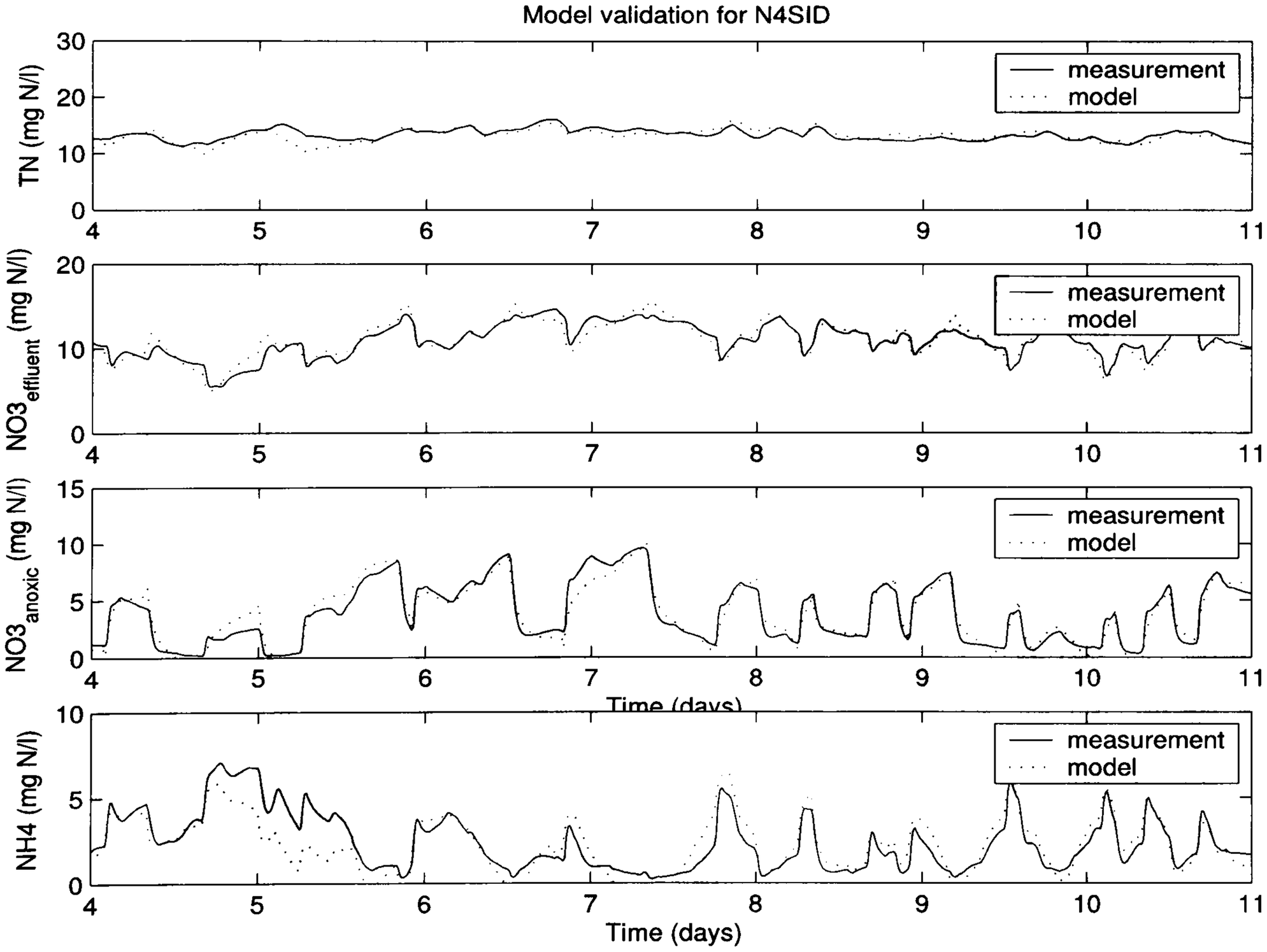

Figure 3.20: Nutrient model identified using N4SID

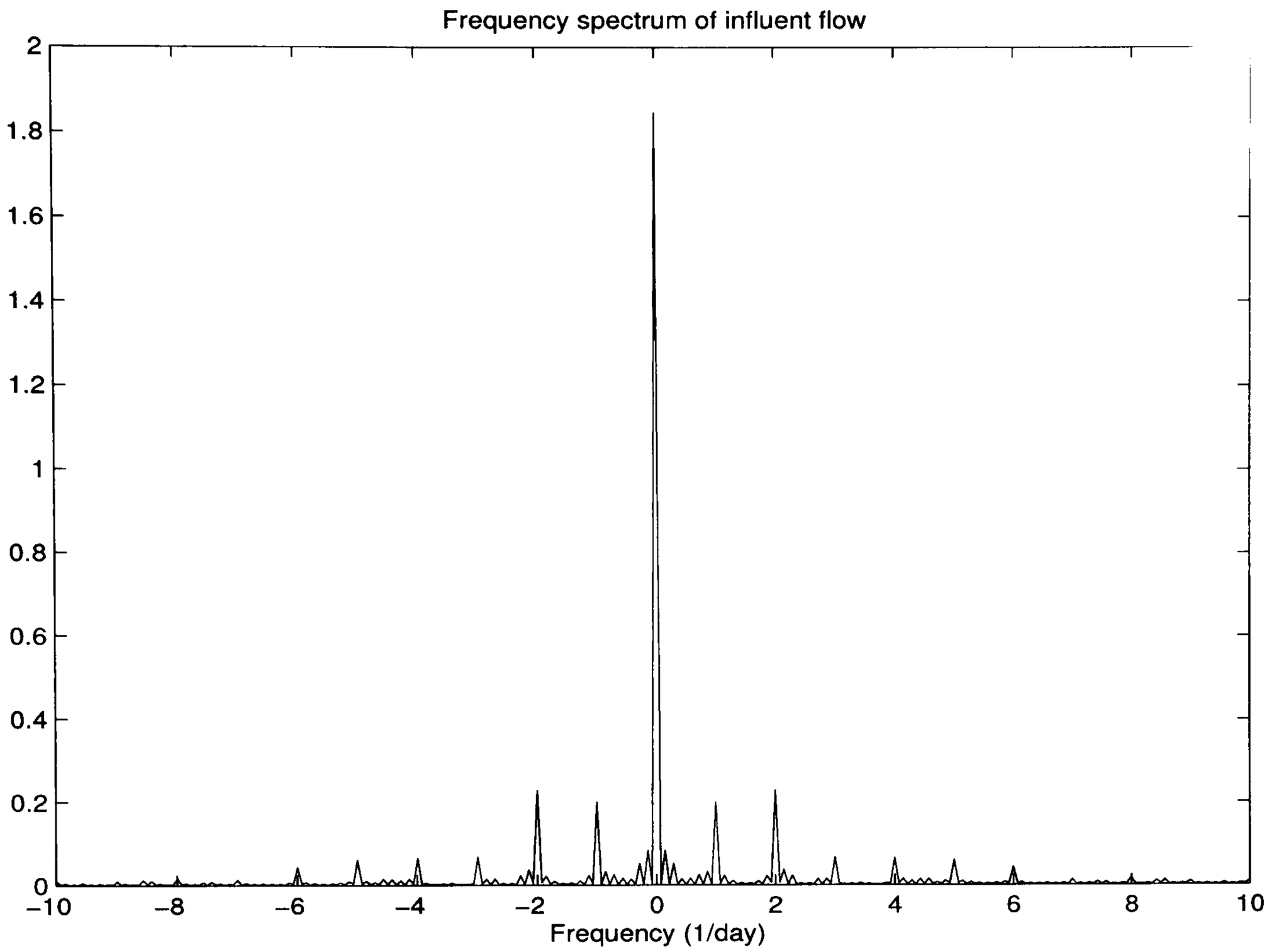

Figure 3.21: Frequency spectrum of influent flow

where, T_s is the sampling period, and equal to 15 minutes for this case. The coefficients and frequency components of the series are determined by performing a Discrete Fourier Transform (DFT) over the historical influent flow, and influent ammonia concentration data. The DFT is defined as in equation (3.64), where the data set is of length N and sampled at a frequency of $f_s = 1/T_s$.

$$F_n = \sum_{k=1}^{N} Flow(k) \cdot e^{-j2\pi \frac{(k-1)(n-1)}{N}} \tag{3.64}$$

Figure (3.21) shows the magnitude spectrum for the influent flow. From this figure it is possible to determine the harmonic frequencies by simple inspection; however, the series coefficients are obtained by employing the following relationships between coefficients of the DFT and the FS,

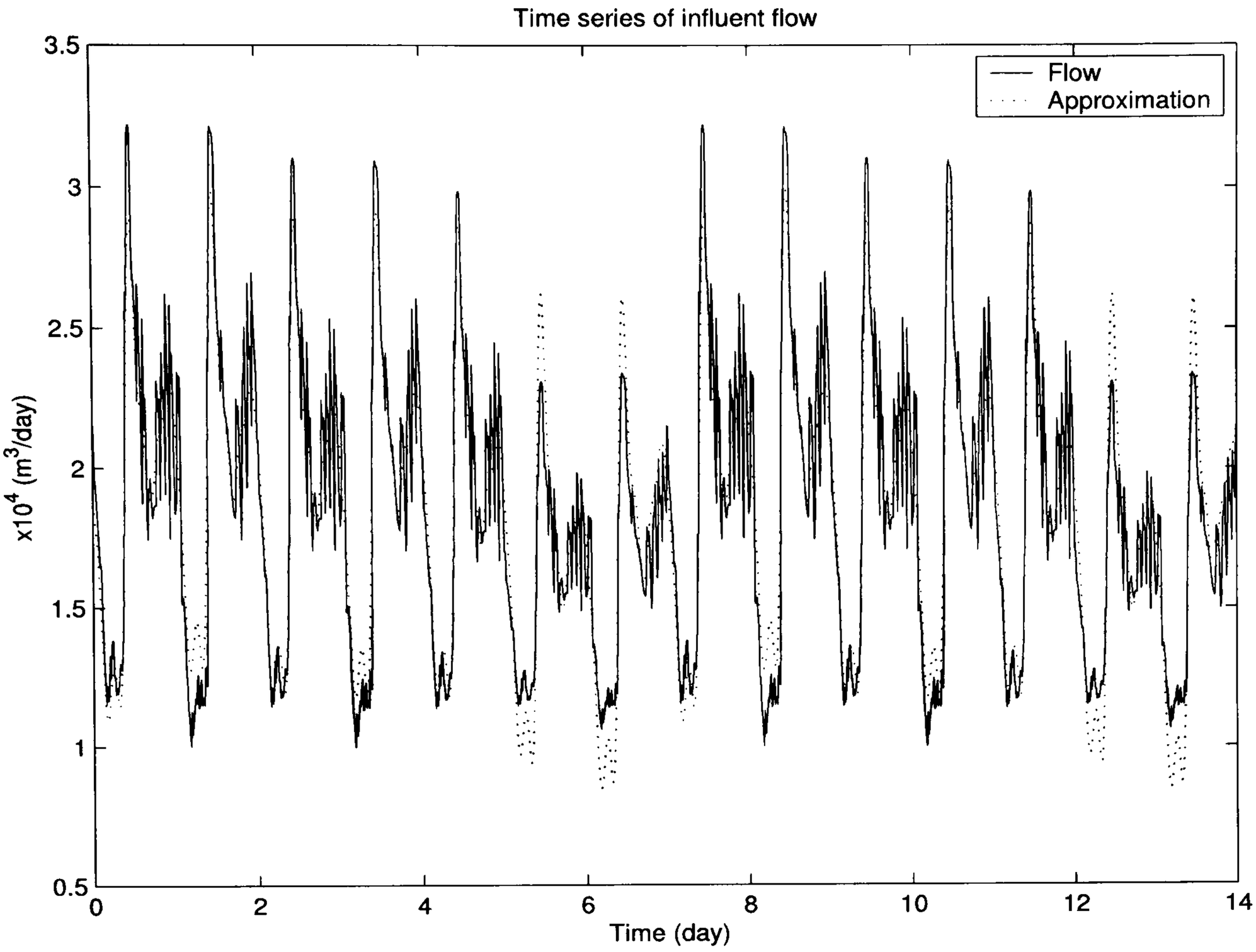

Figure 3.22: Flow time series

$$
\begin{aligned}
a_o &= \frac{F_1}{N} \\
a_n &= \frac{2}{N} Real\{F_{n+1}\} \\
b_n &= -\frac{2}{N} Imag\{F_{n+1}\}
\end{aligned}
$$

For this case, the series has been truncated to eight harmonic components. A similar procedure has been performed for influent NH_4, but with only six harmonic terms and the mean value. Figure (3.23) shows the original time series, and its approximation for influent ammonia.

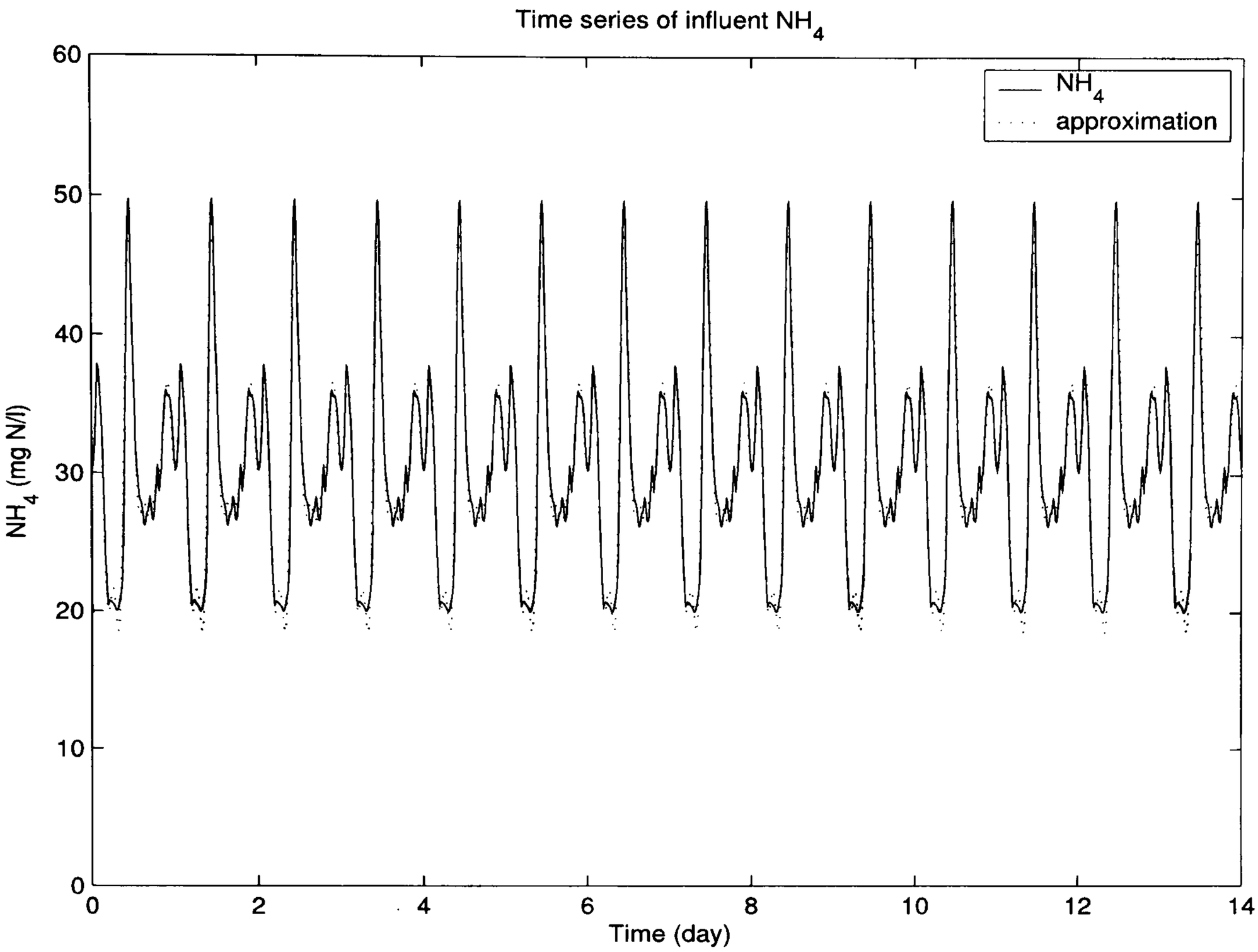

Figure 3.23: Time series of influent NH_4.

3.4 Identification with data from Helsingor WWTP

This section aims at cross-validating the identification results obtained by simulation in the previous sections. To achieve this, historical data from several months from Helsingor WWTP has been carefully selected in order to be representative of a variety of operating conditions of the treatment plant. Later, several identification exercises were performed with the purpose of finding some common properties within the models (i.e. order, poles, and zeros).

3.4.1 Data Selection

Dissolved oxygen data has been collected from Helsingor WWTP. The data is from several months of 2002. The selection of data was based on the following considera-

tions:

1. Representative data set for the period (e.g. low and high hydraulic load).

2. Free of nonlinear effects (e.g. actuator saturation, sensor faults) .

For example, Figure (3.24) shows a data set from March-2002 in which the model can be easily identified since the system is operating in a linear region (i.e. no actuator saturation). Figure (3.25) shows a period in which the actuators are saturated consequently the system is operating in a nonlinear region and therefore the data set is not valid for identification.

Figure (3.26) shows a case in which the plant is operating in a low load condition (e.g. before noon and with low hydraulic load). It is clear in this figure that there are oscillations produced by the aggressiveness of the controller, which continuously runs between saturation levels, especially in the initial overshoot. It is very probable that the control system has entered a limit cycle, since the oscillation frequency is constant (about 20 cycles/day). This effect is very common in mis-tuned control loops with saturation type nonlinearities.

Regarding the influence of the load (i.e. nutrient + hydraulic) in the DO control loops, two clear cases can be distinguished from the online data from the Helsingor plant. The first case is when there is a rain/storm event, in which the hydraulic load increases. Under this condition, it is possible that the inflow could be composed of a high nutrient load.

Figure (3.27) is a good example of this case. The system tries to aerate as much as possible, since NO_3 is decreasing to very low levels, and NH_4 increases dramatically. Also, the return sludge (qret) follows a pattern almost identically as the incoming flow and the recirculation flow increases to try to accelerate the denitrification rate.

A second case, is when the system is subjected to a high nutrient load, but not necessarily a high hydraulic load, as in Figure (3.28). Notice that up to this point, the

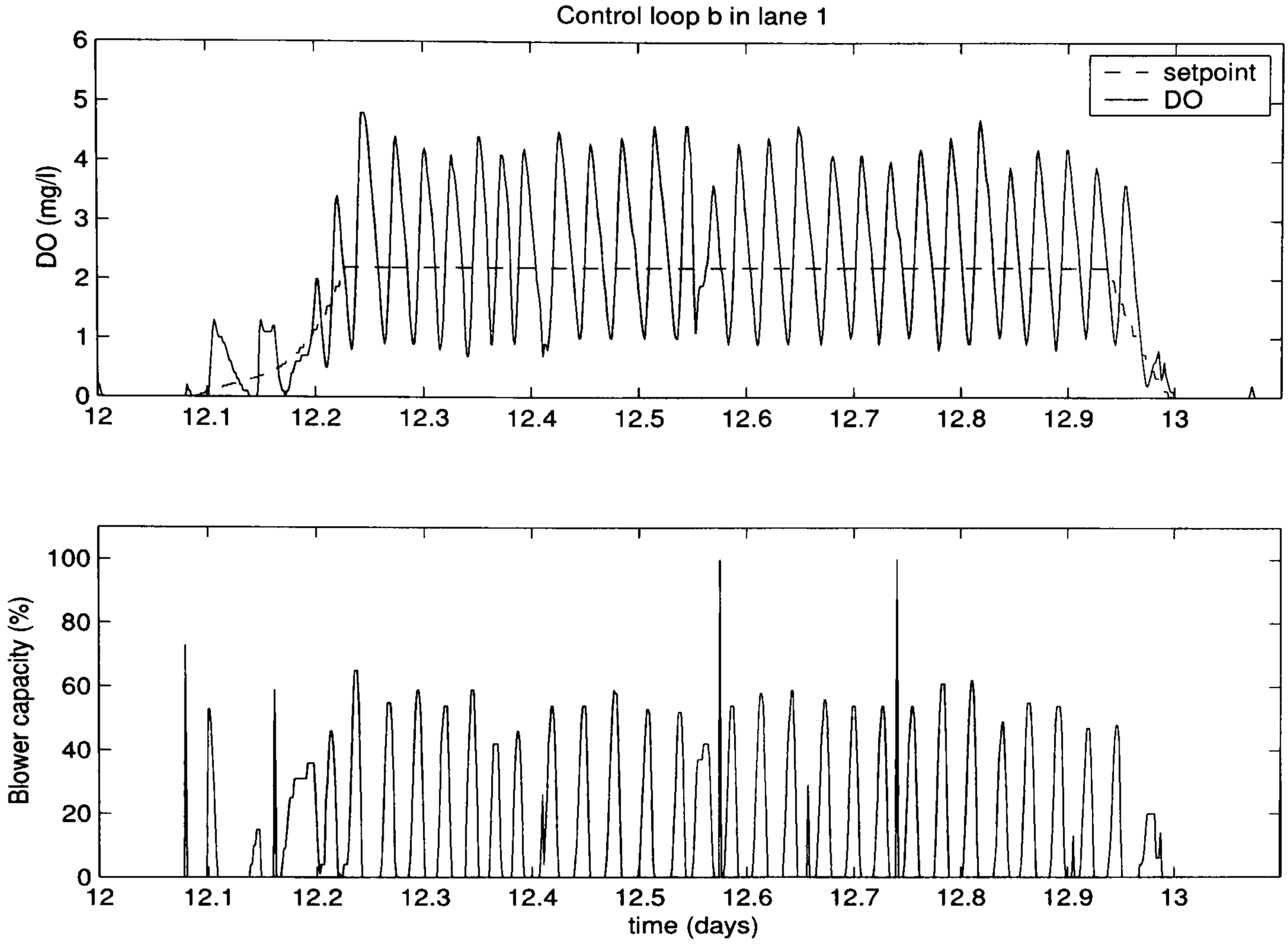

Figure 3.24: DO loop operating in the linear region. There are some periods of time in which the actuator is saturated and therefore producing data which cannot be used for identification

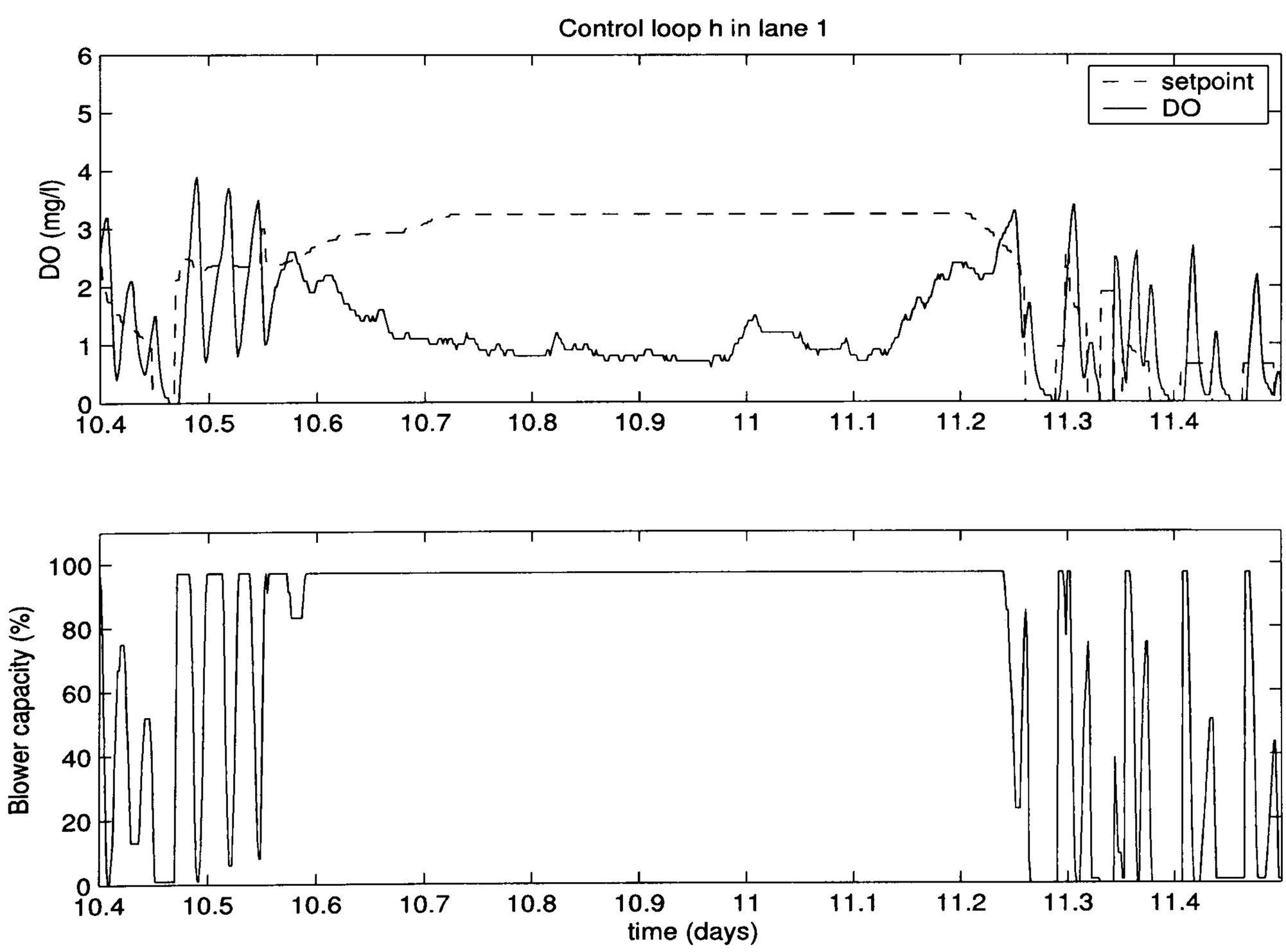

Figure 3.25: DO loop operating in a nonlinear region due to actuator saturation.

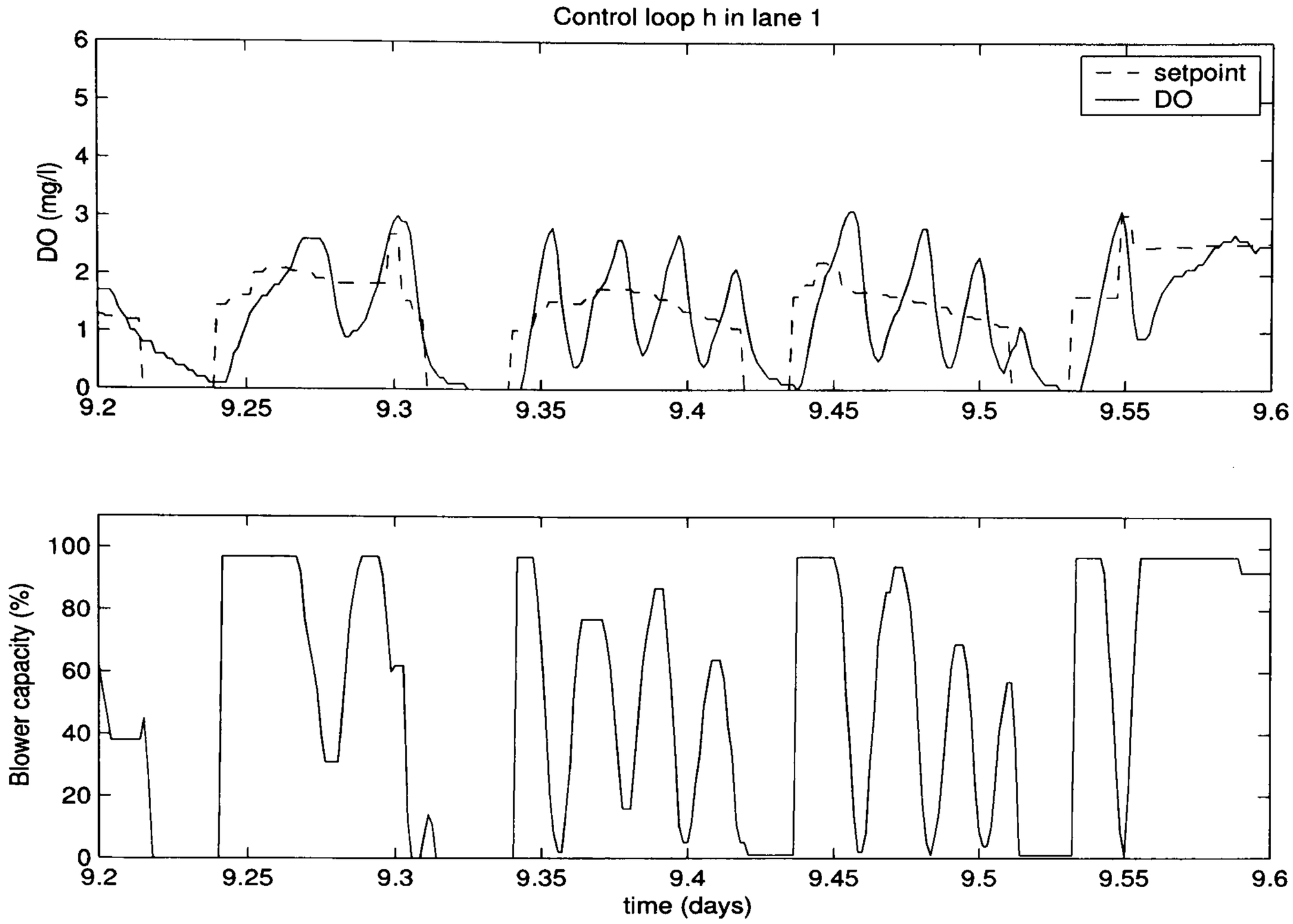

Figure 3.26: DO loop operating in a nonlinear region due to actuator saturation.

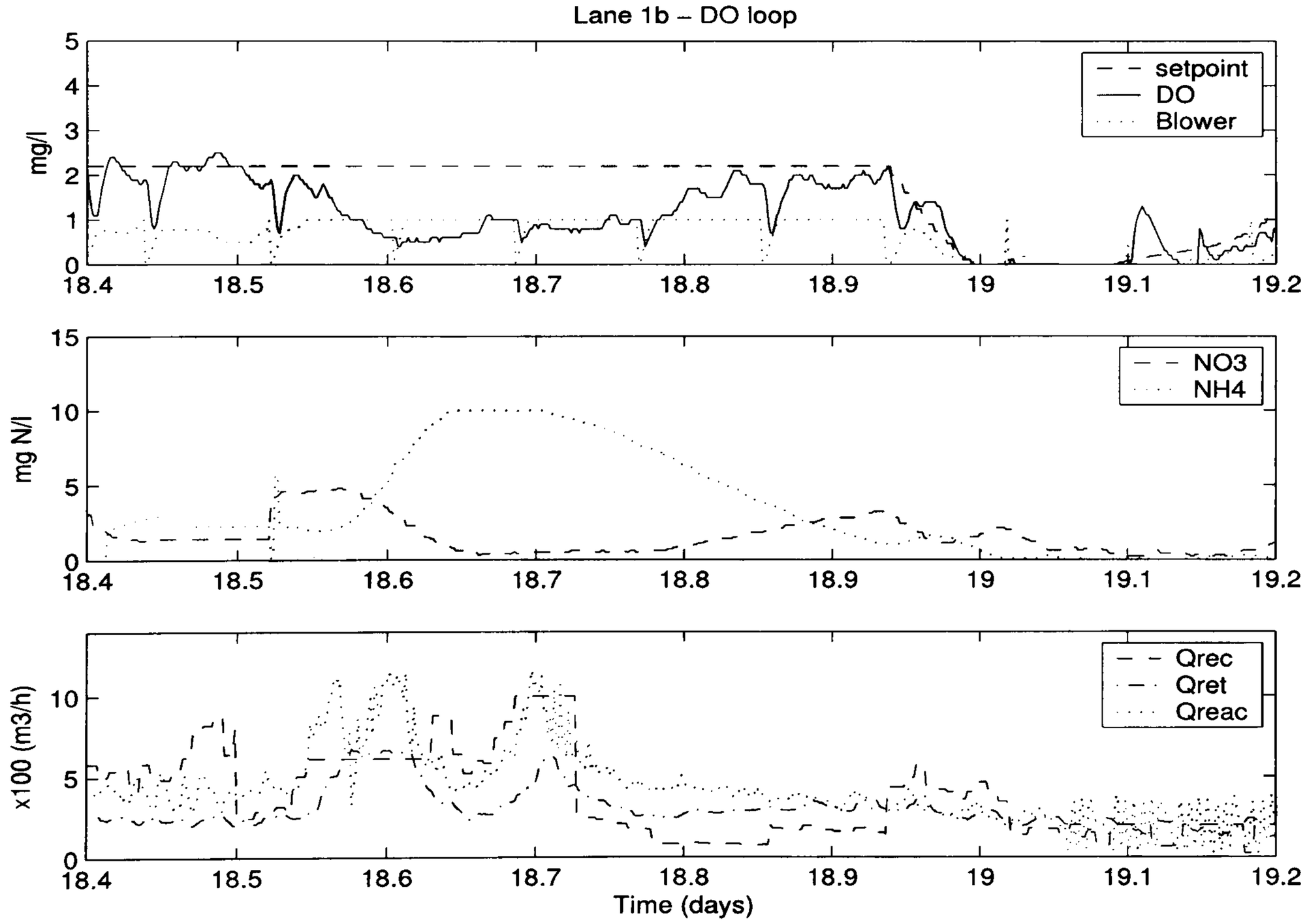

Figure 3.27: Rain/Storm event in the Helsingor WWTP. The DO loop cannot achieve desired setpoint due to actuator limitations. Nutrients (i.e. NH_4) increase due to the composition of the inflow.

nutrient concentration is the major factor that limits the oxygen transfer.

In Figure (3.29), the case is totally the opposite of the ones presented before. The hydraulic load increases due to a short rain, but there is no significant increase in ammonia or nitrate, though the DO loop is still unable to regulate around the setpoint. It is evident, in this example, that rain is the cause of this behaviour.

Given all this case examples of possible scenarios, it is clear that the data necessary to build up the model database has to be chosen carefully, so the models are representative of the process working in the adequate conditions. The following, gives an example of how to build up these data-based models. Some characteristics are exemplified and some conclusions drawn.

3.4.2 Dissolved Oxygen model identification

Due to large amount of data and models identified, this section only provides a summary of six data subsets from one day in May-2002. The final results are however, representative of more than 80 identification exercises with data from different months in 2002. Due to the cyclic aeration of Helsingor WWTP and the delays found in the data (approximately two sample times) the data has been pre-treated in four different ways and four models identified. With these models, simulations were performed and the models evaluated.

For each data subset, the following four cases are considered:

- Full cycle data subset with delay.
- Full cycle data subset without delay.
- Truncated cycle data subset with delay.
- Truncated cycle data subset without delay.

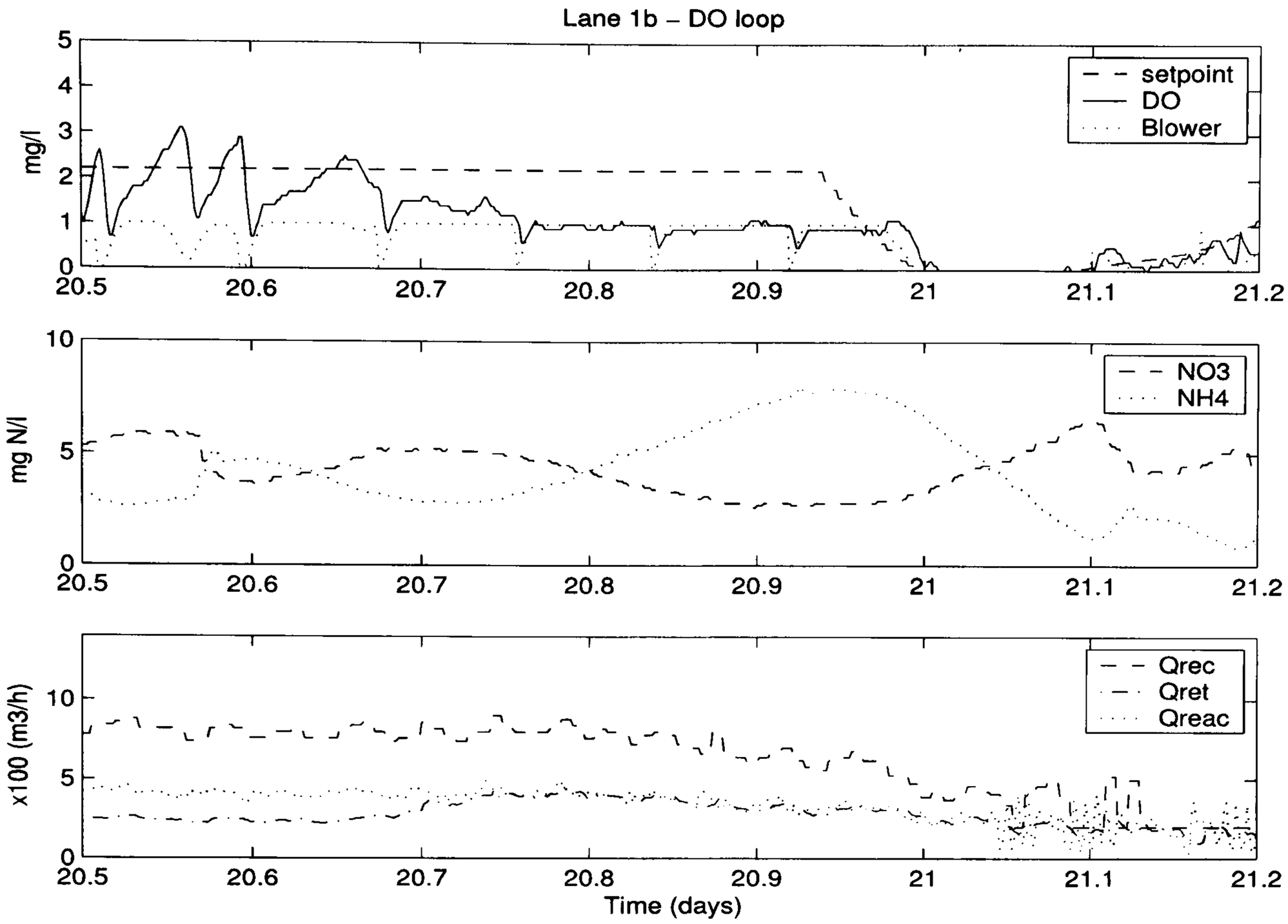

Figure 3.28: Normal hydraulic load with high nutrient composition in the Helsingor WWTP. The DO loop cannot achieve setpoint due to actuator limitations.

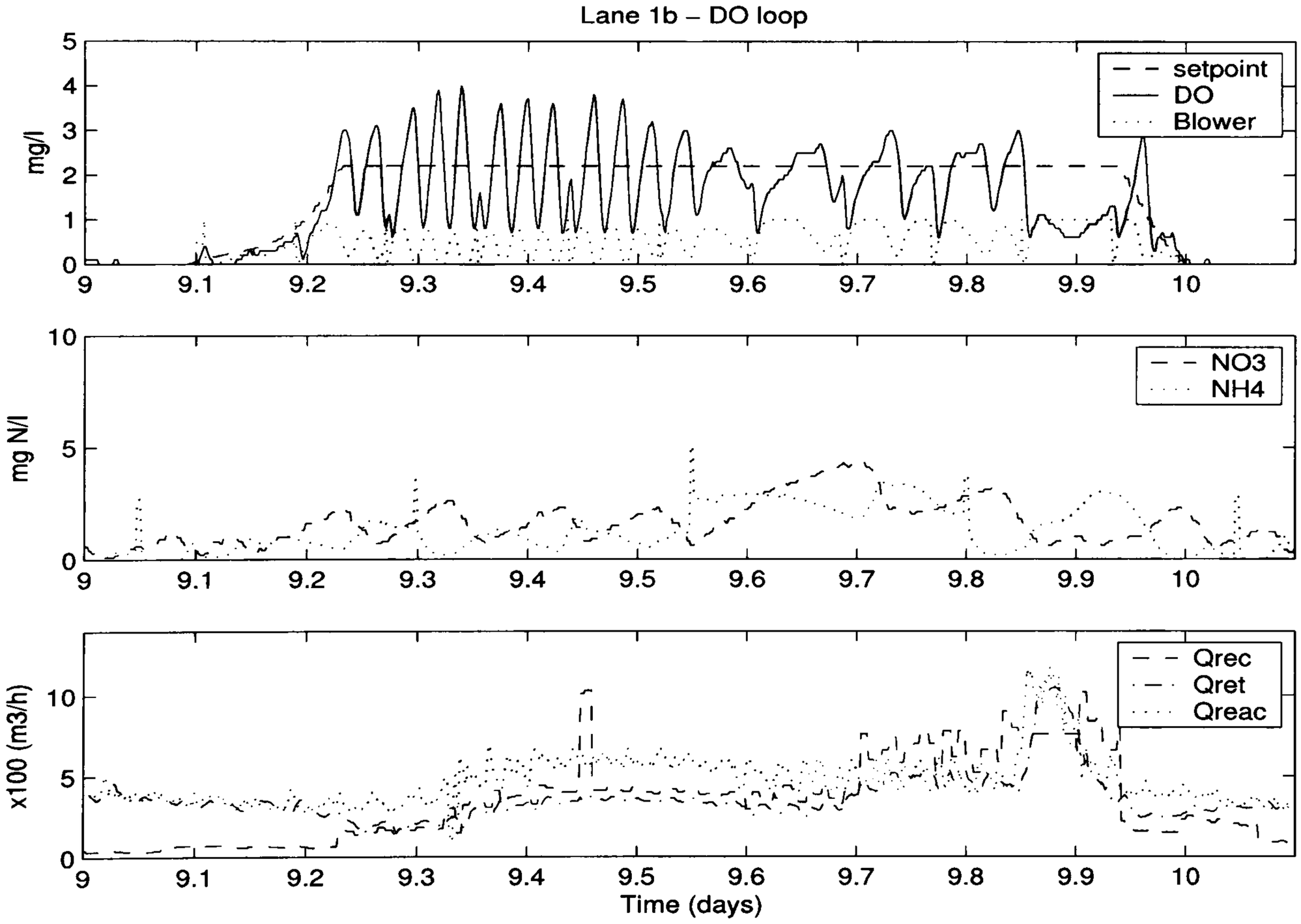

Figure 3.29: High hydraulic load and normal nutrient load. The DO loop cannot achieve setpoint.

Table 3.14: Identified models with data from 21/05/2002

Data set	Transfer function	Zeros	Poles
Full cycle data subset with delay	$\frac{0.02306z^2-0.1295z+0.2054}{z^2-1.304z+0.4705}$	$2.80 \pm j1.0109$	$0.65 \pm j0.21$
Full cycle data subset without delay	$\frac{0.06101z^2-0.142z+0.195}{z^2-1.233z+0.4376}$	$1.16 \pm j1.35$	$0.61 \pm j0.24$
Truncated cycle data subset with delay	$\frac{0.1096z-0.1316}{z-0.6062}$	-1.20	0.60
Truncated cycle data subset without delay	$\frac{0.2643z+0.0823}{z-0.6977}$	0.31	0.69

The term truncated data describes a data set in which only the positive transition in the aeration cycle of the setpoint was considered. The negative transition was dismissed to observe the effect of data truncation in the resulting model.

Identification example

This section describes just one example of the numerous tests performed with data from Helsingor WWTP. The data employed in this example is from the 21 May 2002. Figures (3.30), (3.31), and (3.32) show some identification results. Table (3.14), presents the transfer functions and their respective zero-pole locations.

Identical simulations have been performed over the five remaining data subsets. In general the second order model for the dissolved oxygen closed-loop system achieved the best fitting. Figure (3.33) shows the pole-zero map of the six cycles identified and described before, whereas Figure (3.34) shows the pole-zero map for another six cycles identified in the same month. It is evident that data from the same day displays similar behaviour (i.e. poles are very closed to each other); however, if compared with models from another day, there is a slight variation. Regarding the zeros, in both cases, they seem to vary significantly.

The frequency response of the 12 identified models is plotted in Figure (3.35). The plot shows that the models are contained within a gap 10 dBs at low frequencies, 20 dBs in the medium frequency range, and they are totally uncorrelated for the high

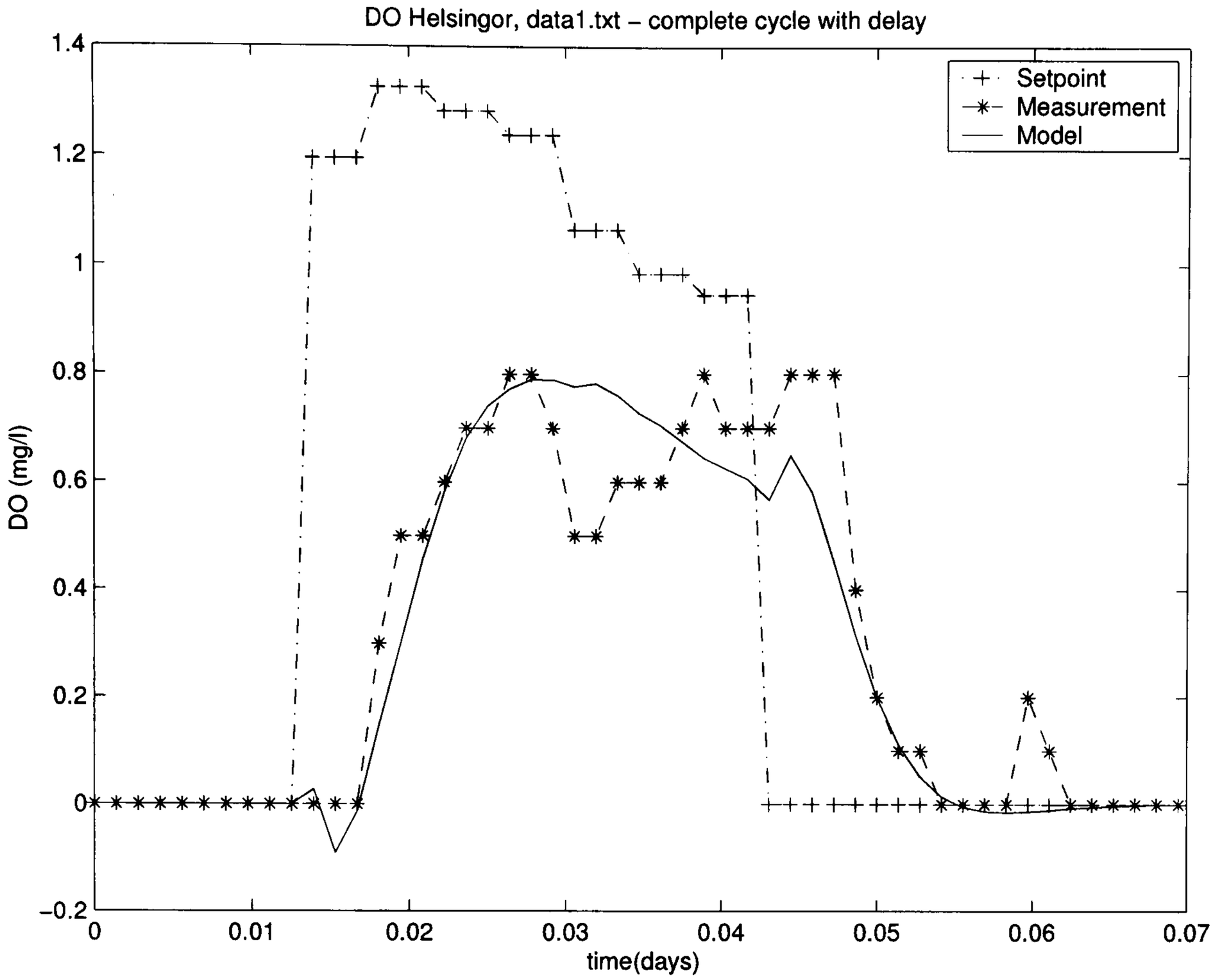

Figure 3.30: DO sensor 1 - complete cycle with delay

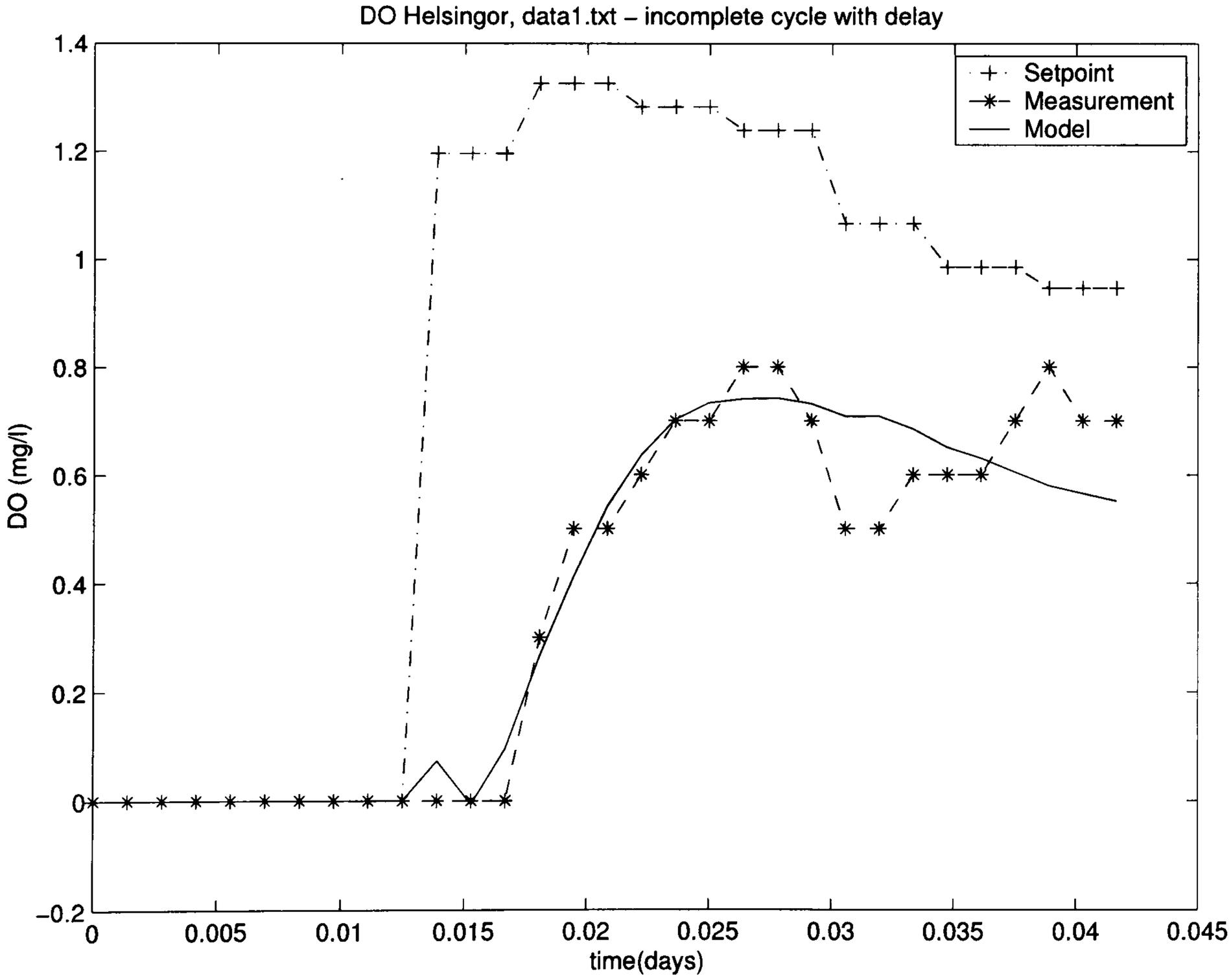

Figure 3.31: DO sensor 1 - incomplete cycle with delay

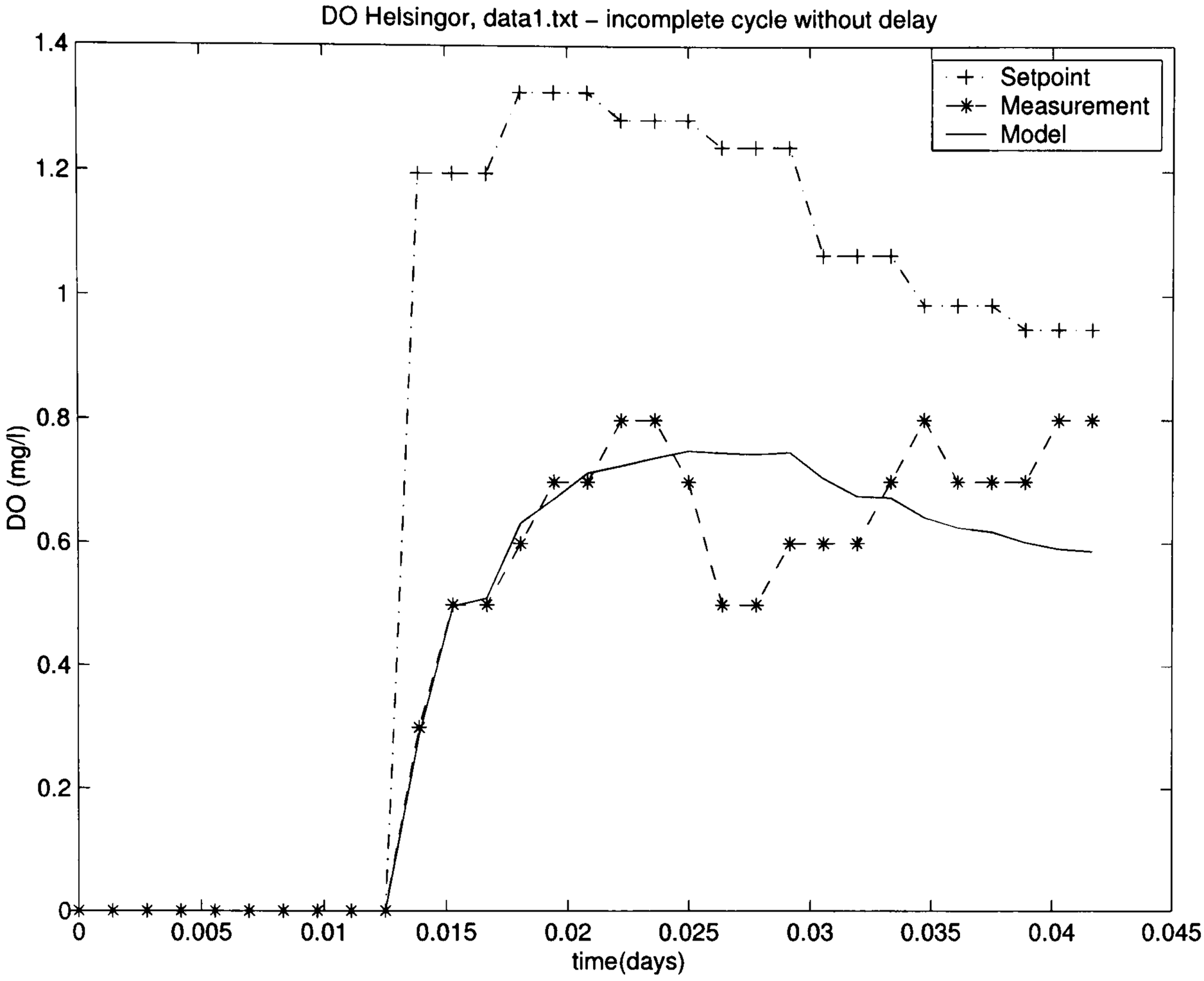

Figure 3.32: DO sensor 1 - incomplete cycle without delay

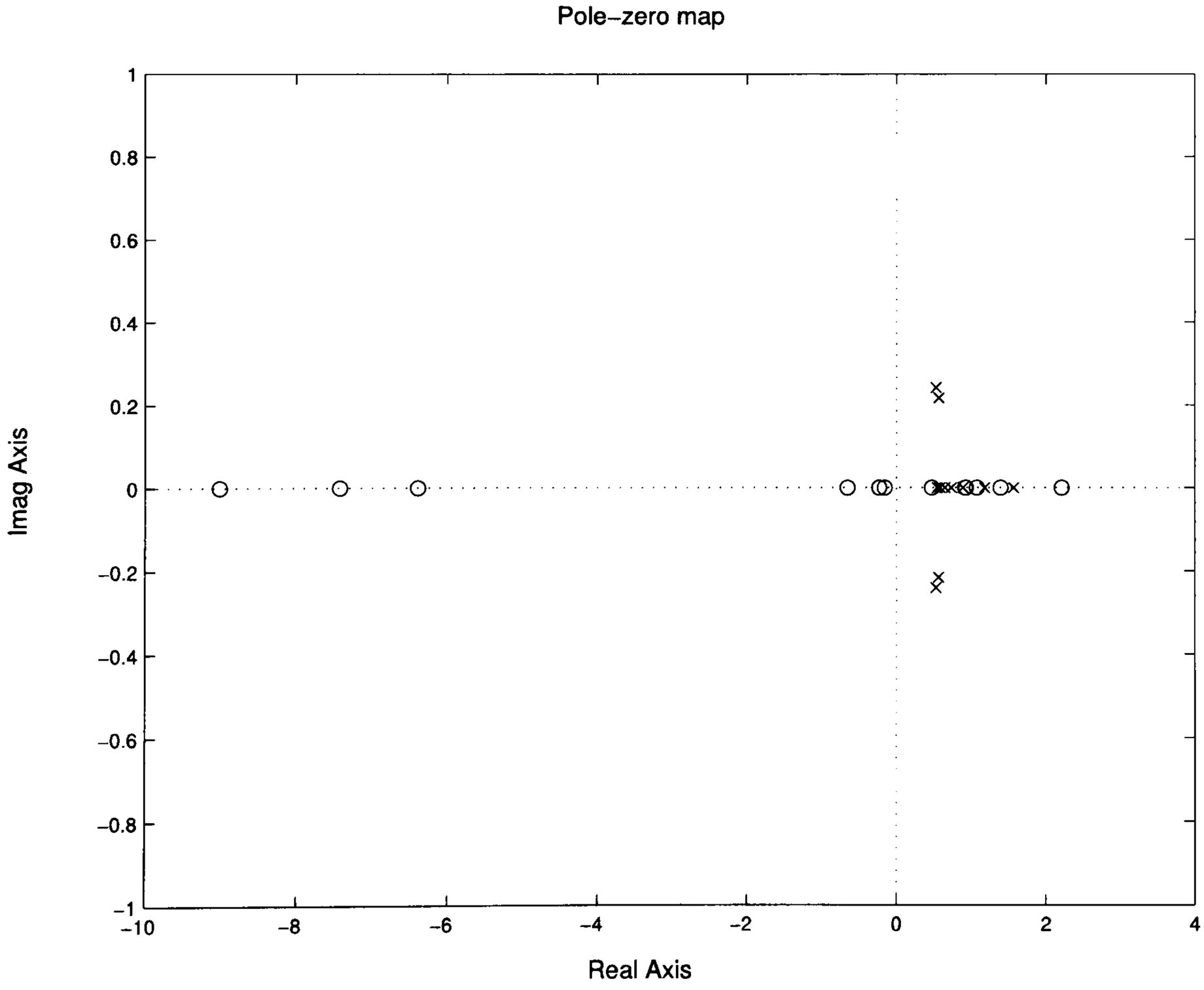

Figure 3.33: Pole-zero map for 6 models identified with data from 21st of May 2002

frequency range. The responses in the low and medium frequency ranges are important to determine the steady-state and stability properties of the systems; whereas the high frequency has to do mainly with the noise rejection and robustness characteristics of the models.

3.4.3 Identification of medium scale variables based on fast variables

In this section the possibility of identifying models for medium scale variables (i.e. NO_3, NH_4, PO_4) is investigated. Medium scale variables are such that their dynamics are in the range of hours. Finding adequate models, which can predict medium scale variable dynamics, can deliver a potential tool which can be useful in special circumstances in which upper control layers fail.

The possible control scheme investigated here, for which the models will be identified, follows the one presented in Figure (3.36).

The knowledge of nitrate dynamics can help in the scheduling of the recirculation rate. Even though results have proven, to some extent, that nitrate dynamics can be identified from data (i.e. DO, Qret, Qrec), this does not mean that this is the best approach to getting a good prediction. So far, this method can only be used within a prediction horizon of only some hours, and when there are no strong nonlinearities (i.e. nitrate is only a function of the inputs).

Models for ammonia can also be identified, but with higher restrictions than for nitrate. The reason for this is that, when the system is subjected to big nutrient loads (i.e. Figure (3.27) and Figure (3.28)), dissolved oxygen actuators saturate and therefore the closed-loop models are no longer valid. An example case when it is possible to identify these models is when the actuators are not saturated as shown in Figure (3.29). In fact, this figure presents a case in which the alternating scheme (i.e. aerobic/anoxic) is

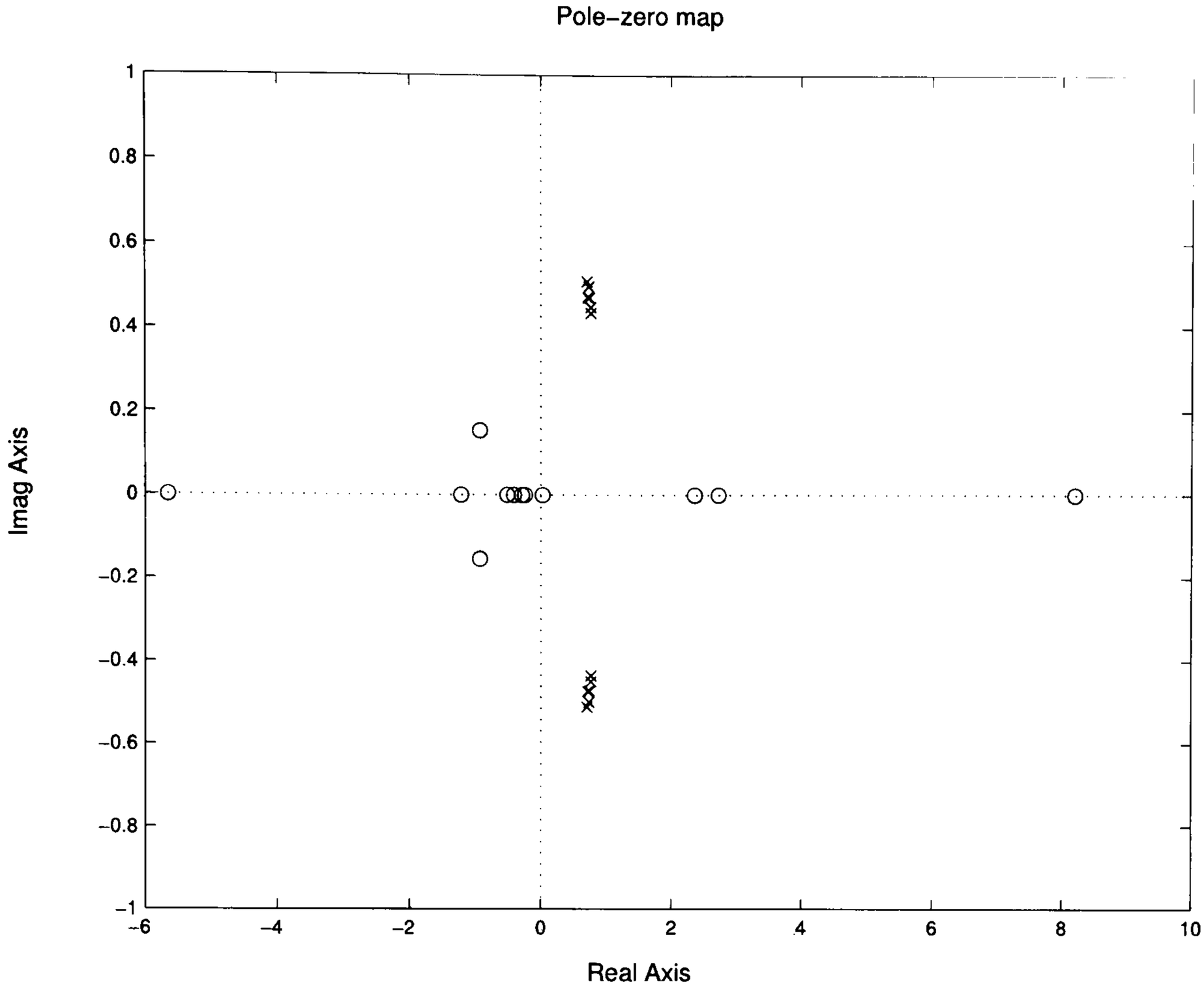

Figure 3.34: Pole-zero map for 6 models identified with data from the 30 of May 2002.

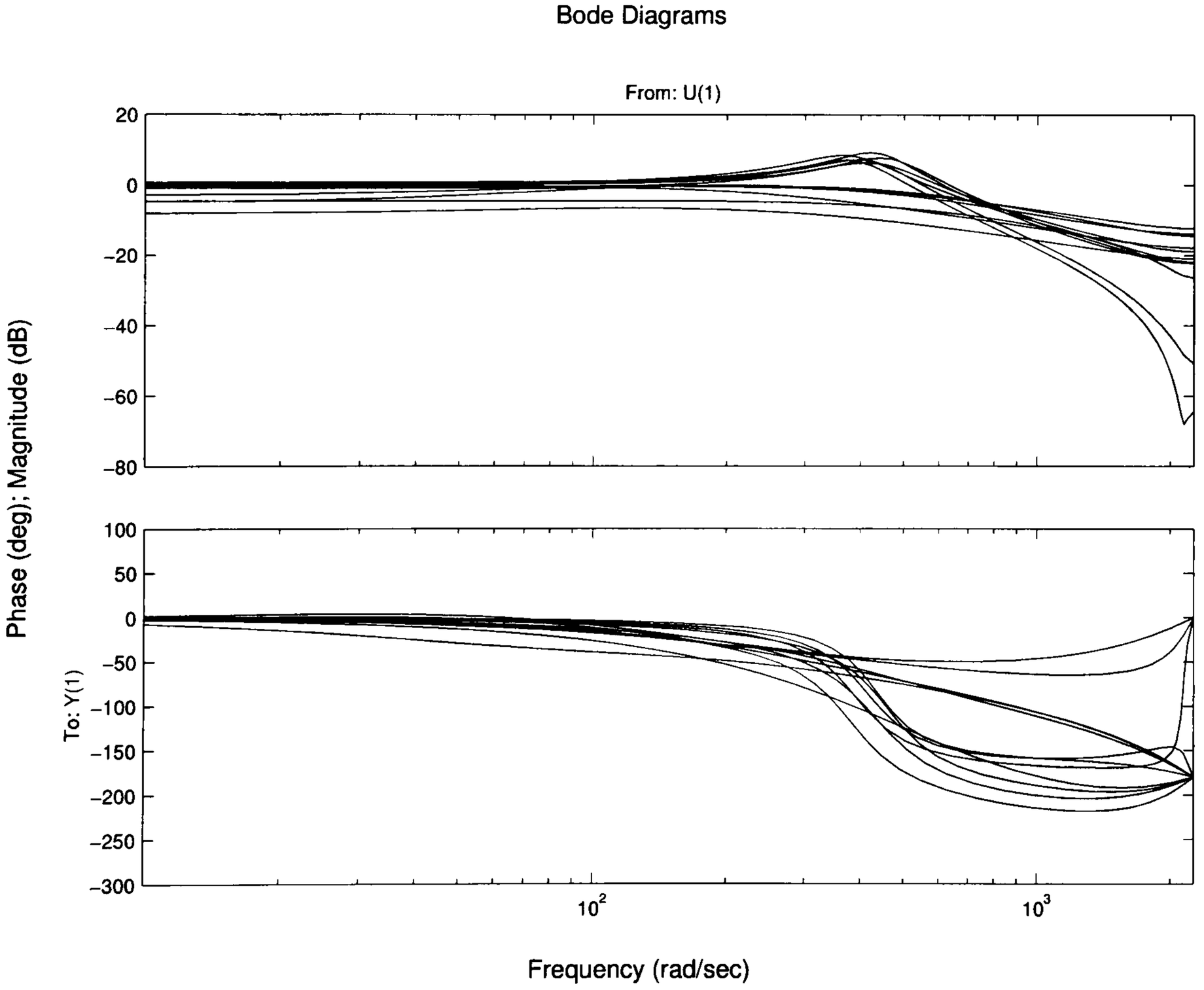

Figure 3.35: Frequency responses of the identified models using data from the 21^{st} and 30^{th} May 2002.

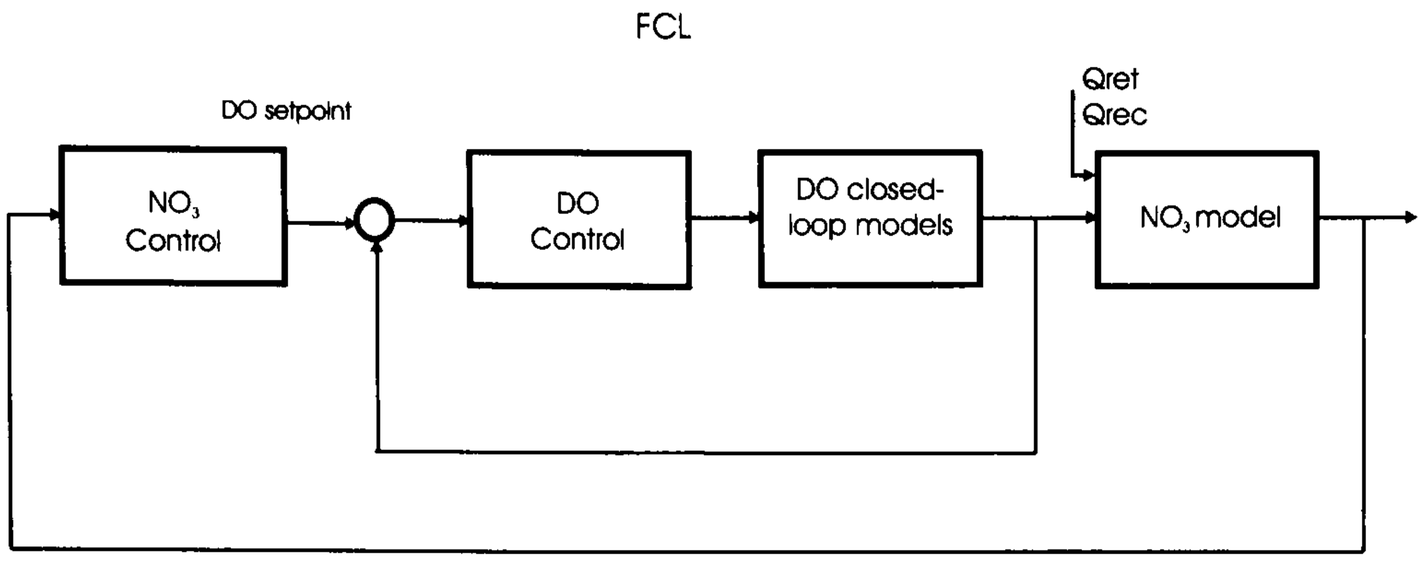

Figure 3.36: Control scheme for nutrient removal control.

evident, due to the increase of nitrate in the aerobic phase and its subsequent decrease during the anoxic phase. Ammonia presents the opposite behaviour.

As an example, using DO in a first case and DO, Qret and Qrec in a second case two models have been identified for nitrate. Figure (3.37) and (3.38) show the responses obtained by these models. The minimum value that concentrations can reach is 0 mg/l; however as seen in Figure (3.38) the model does give a response beyond the physical reality. This physical restriction introduces limitations in the identification, and it is evident that the data should not include nonlinear effects as this.

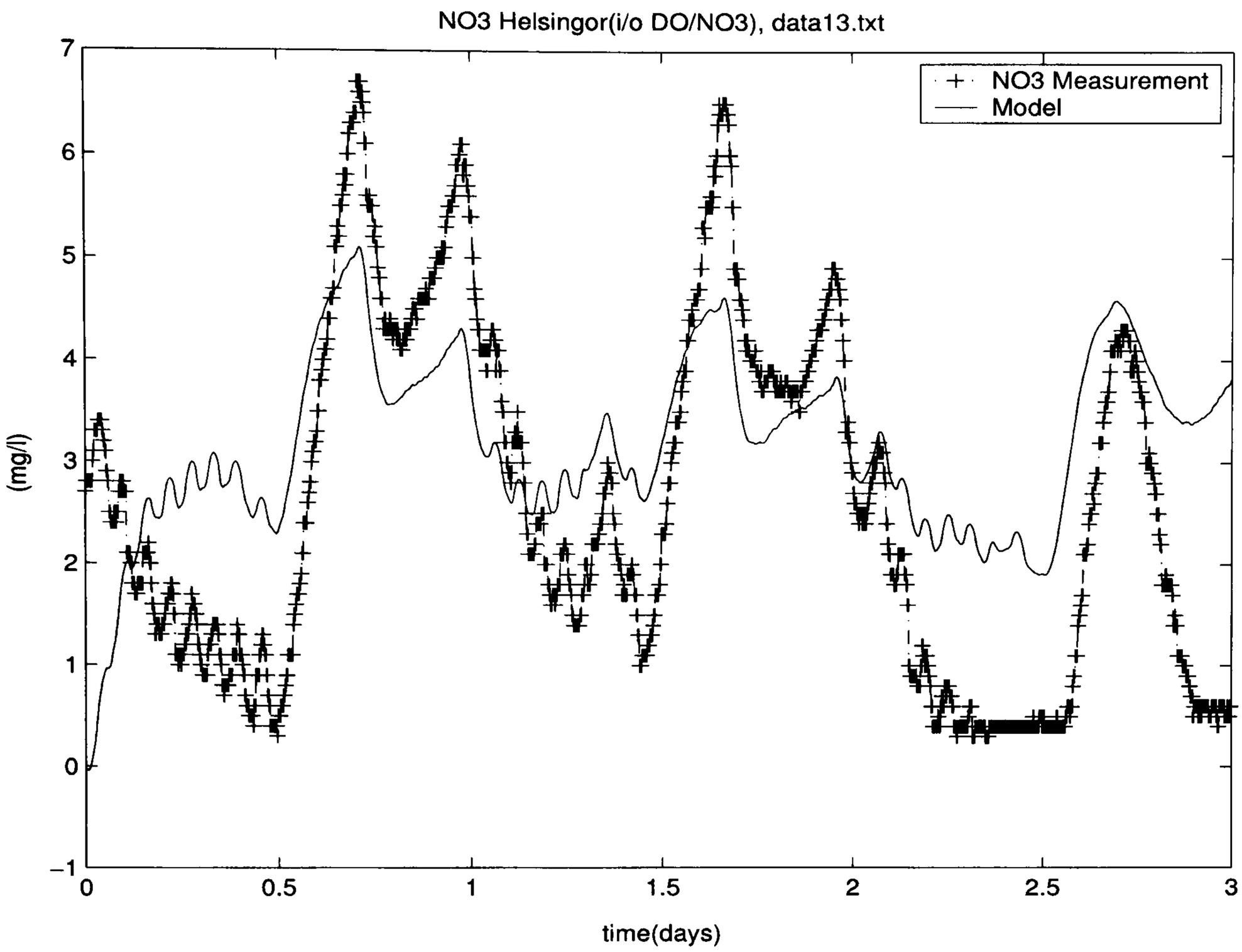

Figure 3.37: Identified model response for NO_3 using DO as input variable.

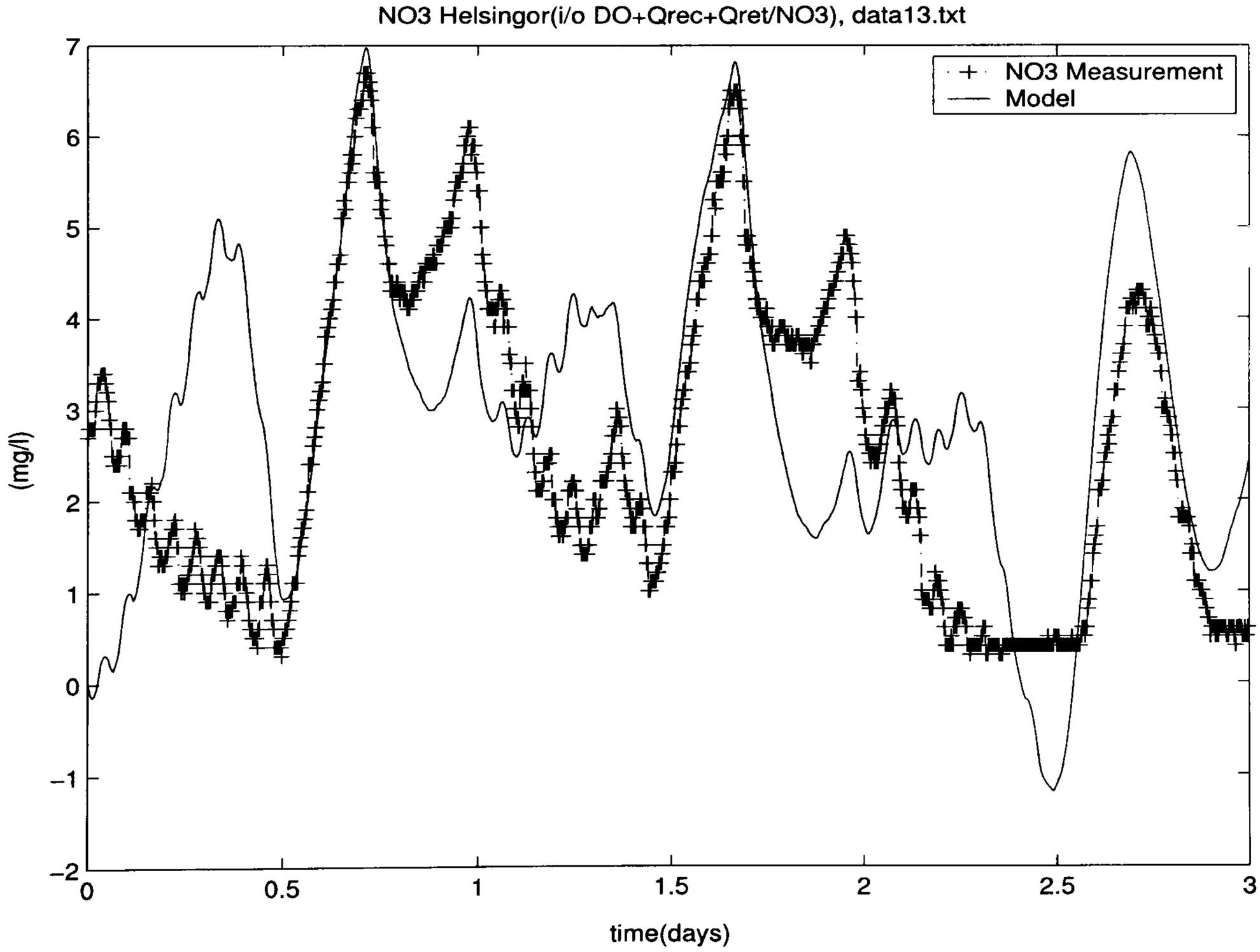

Figure 3.38: Identified model for NO_3 using DO, $Q_{recirculation}$, and Q_{return} as input variables.

3.5 Summary

This chapter has presented a comprehensive study into the identification of models for dissolved oxygen and nitrogen removal for control purposes. Also, an ASM model reduction for an intermittent aeration plant has been developed. Harmonic prediction models for influent flow and ammonia have been briefly discussed as well. The study has employed simulation based data and real-data from Helsingor WWTP.

Subspace algorithms have been used due to their multivariable nature and robust numerical characteristics. A qualitative assessment study of several *open-loop* and *closed-loop* identification algorithms has been performed in order to obtain accurate models. This has unveiled some interesting properties of subspace algorithms; and the necessity of developing new methods to measure and minimise prediction errors within the identification algorithm.

Also, the process of dissolved oxygen identification has followed a systematic procedure throughout the chapter by using the same excitation signals and conditions. Many of the recommendation and experience gained through these identification exercises will be required in Chapter 7, in full-scale plant experimentation.

A section on identification employing real data from several months in 2002 from Helsingor WWTP has also been presented. The section presents different possible loading situations and their effect in the performance of the dissolved-oxygen loops. This study has provided a deeper understanding of some of the most common problems in WWTPs as: (a) un-tuned dissolved-oxygen control loops. (b) Loading effects on the dissolved-oxygen control loops.

In summary, the following models have been obtained by simulation:

1. Three univariate DO models,

2. A multivariable DO model,

3. A nitrogen removal model for a continuous aeration plant,

4. A nitrogen removal model for an intermittent aeration plant,

5. An influent flow prediction model,

6. An influent ammonia prediction model.

All these models will be used in the following chapter for the design of model predictive controllers.

One of the main conclusions arising from the comparison of results between simulation and real-data identification results is that it is possible to model dissolved oxygen as a linear system. Even further, this model might not change significantly in time. This will reduce the need for periodic identification experiments to just sporadic tests, and only when there is a significant reduction in control performance that cannot be compensated by controller re-tuning. Also, a dissolved oxygen control loop can be accurately modelled by 2^{nd} or 3^{rd} order linear models, even though the plant is inherently non-linear.

Chapter 4

Dissolved oxygen and nutrient control

The efficiency and economics of wastewater treatment have become an important issue for water companies in the UK and in the rest of Europe due to new, more stringent EU directives for environmental protection. The most common wastewater treatment process is the activated sludge technology. The costs of wastewater treatment using this technology include chemicals, energy, and human resources for the process and its operation. In order to minimise these costs, the wastewater industry has been led into the development and use of sophisticated strategies for process control. For example the use of intermittent aeration to minimise energy consumption has been reported in several publications Puta *et al.* (1999); Kim *et al.* (2000); Sánchez *et al.* (2002), strategies to increase hydraulic capacity to cope with rain or storm events as reported in Nielsen *et al.* (2000), or improved optimisation by efficient handling of information collected by the control system as reported in (Yuan *et al.*, 2001).

Common problems in wastewater treatment plants have to do with the maintenance and poor effluent quality in many treatment facilities due to poor control approaches. For example, a recent study of four treatment plants in Scotland, Denmark, Germany, and Poland, reported in Sánchez (2002*a*, SMAC project report), concluded, among other, that dissolved oxygen control loops contained P or PI controllers, which were usually mis-tuned and performed poorly. Some side effects of poor tuning are instability and

limit cycles, which in turn lead to blower and valve saturation, wear and tear.

Also, wastewater treatment plants can be located near cities or in remote locations near small towns, thus being either of easy access or very isolated. If a treatment plant is in a remote location, the cost of mobilising human resources and giving maintenance to the plant could represent a significative percentage of the operation budget (estimated at 40% for Scottish Water). Therefore, reducing these operation costs has a high priority in certain situations.

Energy consumption is estimated as about 10% of the total operation budget of a plant. However, depending on the country this could be a higher percentage. In countries like Germany, the cost of energy is totally de-regularised, thus the energy cost can change from period to period. In particular, electricity costs are lower during night, when energy demands are lower. So it might make more sense to optimise the plant to treat the higher loads during the night as in (Puta *et al.*, 1999).

On the other hand, depending on the legislation, operating costs will not be the only source of heavy budgetary burdens. Countries like Denmark, have a taxing scheme over effluent quality. Therefore wastewater companies in Denmark pay a tax per kg. of nitrogen and phosphorus in the effluent.

Given these reasons, it is of prime importance that not only operation costs should be reduced, but also treatment efficiency increased. With the development of advanced process control, many process industries have benefited from a reduction in operating costs without sacrifying plant performance and rather increasing it in most cases. This chapter investigates the design of model predictive controllers (MPC) with the purpose of increasing effluent standards.

The design of MPC controllers for activated sludge WWTP is subject, within other factors, to finding suitable models for prediction and estimation of dissolved oxygen and nutrients. This has been the subject of discussion in Chapter 3, where models have been determined by using subspace identification and deterministic model reduc-

tions and linearisation. This chapter examines the performance of MPC controllers for dissolved oxygen, and nutrient control.

The chapter is organised in the following way: the fist section gives a brief review of model predictive control. In this review only the most necessary equations for the development of this thesis are presented. It is not the objective of this chapter to discuss on the several possible MPC algorithms available, but rather use a simple a reliable MPC controller to examine its advantages for the control of this process. Section 4.2 presents the design and evaluation of univariate and multivariable MPCs for DO control. Finally, section 4.3 presents two approaches to control of nutrients in an activated sludge WWTP. The first employs linear black box models identified using subspace identification techniques for a recirculating plant, while the second approach employs a reduced linear approximation of ASM1 model, for an alternating aeration plug-flow plant. For this last case, a special formulation for MPC has been developed, in order to handle the alternating model control structure. The chapter finalises with a summary of the results produced.

4.1 Review of model predictive control

Consider a discrete-time sampled system described by the state-space model of equations (4.1) and (4.2), where the matrices have the following dimension: $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{l\times n}$, and $D \in \mathbb{R}^{l\times m}$.

$$x(k+1) = Ax(k)+Br(k) \tag{4.1}$$

$$y(k) = Cx(k)+Dr(k) \tag{4.2}$$

Consider as well that the system has η_1 restrictions over the outputs $y(k)$, η_2 restrictions over the inputs $r(k)$, and η_3 restrictions over the change in the inputs $\Delta r(k)$. These

restrictions can be written as in equations (4.3) to (4.5) for each case respectively.

$$\begin{bmatrix} G & | & g \end{bmatrix} \begin{bmatrix} y(k) \\ 1 \end{bmatrix} \leq 0 \tag{4.3}$$

$$\begin{bmatrix} F & | & f \end{bmatrix} \begin{bmatrix} r(k) \\ 1 \end{bmatrix} \leq 0 \tag{4.4}$$

$$\begin{bmatrix} E & | & e \end{bmatrix} \begin{bmatrix} \Delta r(k) \\ 1 \end{bmatrix} \leq 0 \tag{4.5}$$

where the matrices have the following dimensions: $G \in \mathbb{R}^{\eta_1 \times l}$, $g \in \mathbb{R}^{\eta_1 \times 1}$, $F \in \mathbb{R}^{\eta_2 \times m}$, $f \in \mathbb{R}^{\eta_2 \times 1}$, $E \in \mathbb{R}^{\eta_3 \times m}$ and $e \in \mathbb{R}^{\eta_3 \times 1}$. The following sections explain how to calculate the optimal control input $r(k)$ at each sampling instant using model predictive control.

4.1.1 Model predictions

The predictions over an output horizon H_p with control horizon H_u can be calculated using equation (4.6).

$$\mathscr{Y}(k) = \Psi \hat{x}(k) + \Upsilon r(k-1) + \Theta \Delta R(k) \tag{4.6}$$

where:

$$\Psi = \begin{bmatrix} CA \\ \vdots \\ CA^{H_u} \\ \vdots \\ CA^{H_p} \end{bmatrix} \tag{4.7}$$

$$\Upsilon = \begin{bmatrix} CB+D \\ \vdots \\ \sum_{i=0}^{H_u} CA^iB+D \\ \vdots \\ \sum_{i=0}^{H_p-1} CA^iB+D \end{bmatrix} \tag{4.8}$$

$$\Theta = \begin{bmatrix} CB & \cdots & 0 \\ CAB+CB & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{H_u-1} CA^iB & \cdots & CB \\ \sum_{i=0}^{H_u} CA^iB & \cdots & CAB+CB \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{H_u} CA^iB & \cdots & \sum_{i=0}^{H_p-H_u} CA^iB \end{bmatrix} + \begin{bmatrix} D & D & 0 & \cdots & 0 \\ D & D & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D & D & D & \cdots & D \\ D & D & D & \cdots & D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D & D & D & D & D \end{bmatrix} \tag{4.9}$$

The matrices dimensions are $\Psi \in \mathbb{R}^{lH_p \times n}$, $\Upsilon \in \mathbb{R}^{lH_p \times m}$ and $\Theta \in \mathbb{R}^{lH_p \times mH_u}$. Notice that unless direct access to the system states $x(k)$ is available, there will be need for a mechanism to observe or estimate the states denoted as $\hat{x}(k)$.

4.1.2 Constraints

If any restrictions of the type described by equations (4.3) to (4.5) exist, then the optimisation problem needs to be subjected to constraints. To include constraints in the optimisation problem the aforementioned restrictions must be written in terms of $\Delta R(k)$. The constraints mathematical formulation is described next.

4.1.2.1 Constraints on $y(k)$

Considering the prediction horizon of length H_p, the restrictions over the whole horizon are defined by the matrix inequality in equation (4.10).

$$\begin{bmatrix} G & \cdots & 0 & | & g \\ \vdots & \ddots & \vdots & | & \vdots \\ 0 & \cdots & G & | & g \end{bmatrix} \begin{bmatrix} \mathscr{Y}(k) \\ 1 \end{bmatrix} \leq 0 \tag{4.10}$$

Therefore, if V and v are defined as in equations (4.11) and (4.12), the output constraints can be written as in equation (4.13) by using the prediction model defined in (4.6).

$$V = \begin{bmatrix} G & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & G \end{bmatrix} \tag{4.11}$$

$$v = \begin{bmatrix} g \\ \vdots \\ g \end{bmatrix} \tag{4.12}$$

$$V\Theta\Delta R(k) \leq -V\Upsilon r(k-1) - V\Psi\hat{x}(k) - v \tag{4.13}$$

4.1.2.2 Constraints on $r(k)$

The input constraints over the whole control horizon are described by the inequality in equation (4.14).

$$\begin{bmatrix} F & \cdots & 0 & | & f \\ \vdots & \ddots & \vdots & | & \vdots \\ 0 & \cdots & F & | & f \end{bmatrix} \begin{bmatrix} R(k) \\ 1 \end{bmatrix} \leq 0 \tag{4.14}$$

Similarly to the case before, allow the following definitions:

$$N' = \begin{bmatrix} F & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & F \end{bmatrix} \tag{4.15}$$

$$\eta = \begin{bmatrix} f \\ \vdots \\ f \end{bmatrix} \tag{4.16}$$

The matrix N' can then be divided into H_u sub-matrices such that:

$$N' = \begin{bmatrix} N'_1 & \cdots & N'_{H_u} \end{bmatrix} \tag{4.17}$$

Therefore, the inequality constraint of equation (4.14) can be rewritten as in (4.18).

$$\sum_{i=1}^{H_u} N'_i \hat{r}(k+i-1|k) + \eta \leq 0 \tag{4.18}$$

Furthermore, since $\hat{r}(k+i-1|k) = \sum_{j=0}^{i-1} \Delta\hat{r}(k+j|k) + r(k-1)$, (4.18) can be written as:

$$\sum_{j=1}^{H_u} N'_j \Delta\hat{r}(k|k) + \sum_{j=2}^{H_u} N'_j \Delta\hat{r}(k+1|k) + \cdots + \tag{4.19}$$
$$N'_{Hu} \Delta\hat{r}(k+Hu-1|k) + \sum_{j=1}^{H_u} N'_j r(k-1) + \eta \leq 0$$

Finally, let $N_i = \sum_{j=i}^{H_u} N'_j$ and $N = \begin{bmatrix} N_1 & \cdots & N_{H_u} \end{bmatrix}$ so (4.19) can be written as in (4.20).

$$N\Delta R(k) \leq -N_1 r(k-1) - \eta \tag{4.20}$$

4.1.2.3 Constraints on $\Delta r(k)$

Considering the restrictions in (4.5) over the control horizon, it is simple to prove that the constraints can be written as:

$$\mathfrak{E}\Delta R(k) \leq -\mathfrak{e} \tag{4.21}$$

where,

$$\mathfrak{E} = \begin{bmatrix} E & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & E \end{bmatrix} \tag{4.22}$$

$$\mathfrak{e} = \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} \tag{4.23}$$

Finally, all the above mentioned cases can be collected in one inequality constraint as in (4.24).

$$W\Delta R(k) \leq w_1 r(k-1) + w_2 \hat{x}(k) + w_3 \tag{4.24}$$

where,

$$W = \begin{bmatrix} V\Theta \\ N \\ \mathfrak{E} \end{bmatrix} \tag{4.25}$$

$$w_1 = \begin{bmatrix} -V\Upsilon \\ -N_1 \\ 0 \end{bmatrix} \tag{4.26}$$

$$w_2 = \begin{bmatrix} -V\Psi \\ 0 \\ 0 \end{bmatrix} \tag{4.27}$$

$$w_3 = \begin{bmatrix} -v \\ -\eta \\ -\epsilon \end{bmatrix} \tag{4.28}$$

The matrices dimensions are $W \in \mathbb{R}^{(\eta_1 H_p + \eta_2 H_u + \eta_3 H_u) \times (mH_u)}$, $w_1 \in \mathbb{R}^{(\eta_1 H_p + \eta_2 H_u + \eta_3 H_u) \times (m)}$, $w_2 \in \mathbb{R}^{(\eta_1 H_p + \eta_2 H_u + \eta_3 H_u) \times (n)}$ and $w_3 \in \mathbb{R}^{(\eta_1 H_p + \eta_2 H_u + \eta_3 H_u) \times (1)}$.

4.1.3 Cost function

The increment of the control inputs over the prediction horizon are calculated by minimising the quadratic cost function of equation (4.29). In this cost functional, $S(k)$ is a vector of future setpoints to be applied to the process, and is usually considered to be constant over the prediction horizon unless there is information of its future trajectory.

$$J(k) = \|\mathscr{Y}(k) - S(k)\|^2_{\mathscr{Q}} + \|\Delta R(k)\|^2_{\mathscr{R}} \tag{4.29}$$

Define then the following expression:

$$\mathscr{E}(k) = S(k) - \Psi\hat{x}(k) - \Upsilon r(k-1) \tag{4.30}$$

By substituting (4.6) and (4.30) into (4.29), the cost function can be written as:

$$J(k) = \|\Theta\Delta R(k) - \mathscr{E}(k)\|^2_{\mathscr{Q}} + \|\Delta R(k)\|^2_{\mathscr{R}} \tag{4.31}$$

Since $\mathscr{Q}$ and $\mathscr{R}$ are positive definite or at least positive semi-definite then it is possible to decompose them as $\mathscr{Q} = S_{\mathscr{Q}} S^T_{\mathscr{Q}}$ and $\mathscr{R} = S_{\mathscr{R}} S^T_{\mathscr{R}}$. $\Delta R(k)$ can then be calculated as

the least-squares solution to (4.32) subject to the inequality constraint in (4.24).

$$\begin{bmatrix} S_{\mathscr{Q}}\Theta \\ S_{\mathscr{R}} \end{bmatrix} \Delta R(k) = \begin{bmatrix} S_{\mathscr{Q}}\mathscr{E}(k) \\ 0 \end{bmatrix} \tag{4.32}$$

4.1.4 Measurable disturbance and feedforward

There might be some cases in which the effects of disturbances can be anticipated and approximately cancelled out by suitable control actions. An approach of this type is called feedforward control. In order to anticipate the effect of the disturbance it is necessary to have some measurements that will indicate that the effect is about to happen. The effect of the disturbance can only be fully cancelled if there is an exact model of the disturbance to output transfer function.

Feedforward can be easily incorporated into predictive control, it is just a matter of including the effects of the disturbance into the output predictions. The following section presents a brief summary of the basic changes required in the equations introduced in the previous sections.

4.1.4.1 Plant model with measured disturbance

A plant model with l outputs, m inputs, q measured disturbances, and n states can be described by the following set of equations,

$$x(k+1) = Ax(k)+Br(k)+B_d d_m(k) \tag{4.33}$$

$$y(k) = Cx(k)+Dr(k)+D_d d_m(k) \tag{4.34}$$

The model output prediction will now become

$$\mathscr{Y}(k) = \Psi\hat{x}(k)+\Upsilon r(k-1)+\Theta\Delta R(k)+\Xi D_m(k) \tag{4.35}$$

where all the matrices hold their previous definitions and,

$$\Xi = \begin{bmatrix} CB_d & D_d & \cdots & 0 & 0 \\ CAB_d & CB_d & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ CA^{H_p-1}B_d & CA^{H_p-2}B_d & \cdots & CB_d & D_d \end{bmatrix} \tag{4.36}$$

$$w_3 = \begin{bmatrix} -v \\ -\eta \\ -\epsilon \end{bmatrix} \tag{4.37}$$

The matrices have the following dimensions: $\Xi \in \mathbb{R}^{(lH_p)\times q(H_p+1)}$ and $D_m \in \mathbb{R}^{q(H_p+1)\times(1)}$. The state observer has to be designed using the model of equations (4.33) and (4.34). Common practice is to assume that the measured disturbance will remain constant at the last measured value $d_m(k)$. However, if there is a better model of the disturbance this can be incorporated.

4.2 Control of dissolved oxygen

This section presents the design and evaluation of MPC controllers for dissolved oxygen. The section is divided into two parts. The first part presents the univariate case, while the second part considers the multivariable case.

The purpose of the controllers designed and tested in this section is to improve the system performance for setpoint manoeuvring and compensate for external disturbances like load changes due to daily variations in influent composition or weather changes such as rain events. In this context, the proposed controller structure would have a hierarchical architecture such as that presented in Figure (4.1).

This controller structure also reduces implementation costs and provides a high versatility in the tuning of the MPC controller. By placing the MPC controller as a second

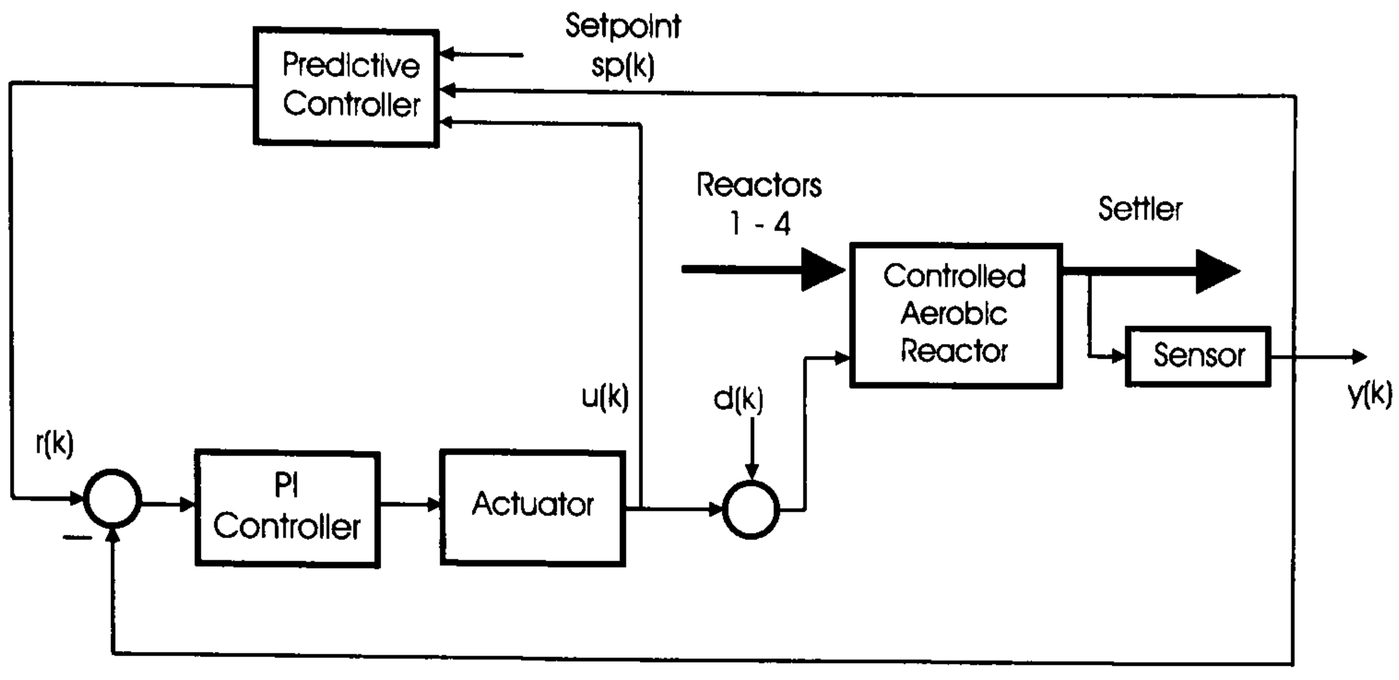

Figure 4.1: Predictive control scheme.

level controller, the implementation could be performed over a more advanced software platform and computationally effective computer. The development of such software platform has been described in chapter 2 and reported in (Sánchez *et al.*, 2003*a*). This structure will also reduce costs, since it can be implemented over an existing controller, usually programmed in a PLC, without need of major modifications. Even more, all the process of implementation can be performed while the plant is operating.

4.2.1 Univariate model predictive control

Several subspace predictive control methods have been developed within the last few years. The technique itself is considered to be fairly new and it has been just recently that some possibilities of implementing model predictive controllers (MPCs) directly from a subspace framework are being explored as for example in (Favoreel and De Moor, 1998; Kadali *et al.*, 2003; Ruscio, 1997*b*). The work of this thesis, however, does not approach the implementation of the MPC controllers in this way, but uses a state space formulation as described in section 4.1.

In this section, three predictive controllers are designed. The predictor is formulated for the composite, composite with disturbance model and augmented with disturbance models which were described in section 3.2.1.2. For the augmented model, only the *CVA* case is considered since results are very similar to the produced by the *SV* algorithm.

4.2.1.1 Controller design

The tuning and design of predictive controllers relies in adjusting parameters like: weights, disturbance models and observer dynamics, reference trajectories, and horizons. Due to the number of degrees of freedom introduced by the quantity of tunable parameters, the design of a predictive controller is mainly subject to a number of rules of thumb, and to several simulations until the required performance is achieved. Only in the special case of having a linear plant, and the MPC controller operating in the linear region (no active constraints), the analysis and design of the controller can be easily integrated into the linear systems theory.

Traditionally, controller performance has been assessed by measuring the time domain responses to several, probing input signals. This, however, might not be always a good indication of the real advantages, or disadvantages, that the controller behaviour might exhibit under certain circumstances. From the practical point of view, simulations are usually not possible; however, the experience gained by these previous exercises might give a more valuable insight than expected. In general, the rules for tuning a MPC can be deduced by observing the effect that each parameter has over the closed-loop performance. A comprehensive review of these effects can be found in (Maciejowski, 2001).

The design of the controllers in this chapter has therefore been performed by trial and error, thus the controller parameters values obtained are the result of number a of simulations. The models employed are the ones described in section 3.2.1. In this section three models were proposed by using different identification algorithms and re-arranging them to include the airflow signal, and a constant unmeasurable disturbance model. These three models were classified as composite, composite w. disturbance, and augmented with disturbance. Only the first of these three models does not include a disturbance compensation, and as will be presented, this will affect its performance under different weather conditions, even though its transient response is better.

Table (4.1) summarises the values for the prediction horizon (H_p), control horizon (H_u) and weightings $\mathscr{Q}$ and $\mathscr{R}$ that produced the best results.

Table 4.1: Controller parameters.

Model	H_p	H_u	$\mathscr{R}$	$\mathscr{Q}$
composite	50	10	10^2	diag[15 10^{-5}]
comp. & dist.	50	10	10^3	diag[15 10^{-5}]
aug. & dist.	50	7	30	diag[15 10^{-5}]

comp. & dist.: composite model with disturbance estimation
aug. & dist.: augmented model with disturbance estimation

4.2.1.2 Constraints

The inclusion of constraints is fundamental in this problem. It is in this way that the physical limitations arising from the actuators (air compressors) are included when solving the optimisation. Constraints also allow the inclusion of operation conditions that are necessary for the process to work. For example, in many WWTPs it is necessary to keep a minimum aeration regardless of the oxygen concentration, just to keep the reactors fully mixed. It is also evident that constraints allow limits to be imposed over variables which in practice cannot go under or over certain limits, as for example the oxygen concentration cannot be less than zero. To implement such restrictions, all the variables must be written as a function of the optimisation variable, that is in this case $\Delta R(k)$. For the optimisation problem, inequality constraints have the form of equation (4.24). For this study, the physical limits are tabulated in Table (4.2).

Table 4.2: Physical Limits

Limit/variable	Q (%)	r (mg/l)
Lower limit	0.001	0.001
Upper limit	90	9

4.2.1.3 Simulation results

This section presents and discusses the simulation results for the three MPC controllers designed for the SISO models identified in the previous chapter. Simulations are also

carried out for the system with only the original PI controller, which lies in the first control level and described in section 2.2, where the COST benchmark is introduced.

The simulation scenarios include constant influent, dry weather influent, rain influent, and storm influent as defined in (Copp, 2002). Within the simulations, the constant influent is utilised to assess the transient response to changes in setpoint and disturbance rejection, while the dynamic influent files are used to provide a statistical evaluation of the performance in the long term. Tables (4.3-4.4) show the results for setpoint tracking and disturbance rejection, while Tables (4.5-4.7) show the statistics for dynamic performance under the specified weather conditions.

Results show that even though the performance of the composite model with disturbance estimation is acceptable in the transient analysis, its performance is significant lower when the simulation is run for dynamic influent. Also, notice the large settling time of the composite model. This excessively large number indicates that there is a sustained offset in the model, give by the absence of a correction for the unmeasurable disturbance at the output. These results also show the benefit of including a second level of control over a PI control loop, since the PI performance is poorer than when combined with a MPC.

4.2.2 Multivariable model predictive control

The use of a multivariable controller for dissolved oxygen can provide several advantages over using single decoupled PI controllers for each aeration basin. When using a decoupled controller, the control system will try to follow the setpoint by having only information from its own basin DO measurement, thus unable to preview the effects of a setpoint change in the other basins. By introducing a multivariable controller on top of the existing PI controllers, it is possible to reject much easier any condition change that propagates through the system. A multivariable MPC controller for dissolved oxygen will also introduce the advantage of predicting the future behaviour of the system

Table 4.3: Dynamic performance

Case	Overshoot (%)	Settling Time (min)
composite	0	>360
composite w. disturbance	2.11	55
augmented w. disturbance	1.277	27
PI	0.99	74

Table 4.4: Disturbance rejection

Case	Peak (%)	Rejection Time (min)
composite	-43.78	>360
composite w. disturbance	-34.72	55
augmented w. disturbance	-35.58	44
PI	-37.92	98

Table 4.5: Dry weather statistics

Case	Max	Min	Mean	Var	St.Dev.
composite	2.71	1.37	2.00	0.087	0.295
comp. & dist.	2.04	1.91	1.99	$3.2\text{x}10^{-4}$	0.018
aug. & dist.	2.24	1.67	1.99	0.007	0.086
PI	2.49	1.40	1.99	0.030	0.174

comp. & dist.: composite model with disturbance estimation
aug. & dist.: augmented model with disturbance estimation

Table 4.6: Rainy weather statistics

Case	Max	Min	Mean	Var	St.Dev.
composite	2.70	1.38	2.02	0.070	0.265
comp. & dist.	2.04	1.89	1.99	$2.4\text{x}10^{-4}$	0.015
aug. & dist.	2.24	1.68	1.99	0.006	0.075
PI	2.49	1.42	2.00	0.024	0.153

comp. & dist.: composite model with disturbance estimation
aug. & dist.: augmented model with disturbance estimation

Table 4.7: Storm weather statistics

Case	Max	Min	Mean	Var	St.Dev.
composite	2.70	1.38	1.99	0.080	0.282
comp. & dist.	2.05	1.91	1.99	$3.0\text{x}10^{-4}$	0.017
aug. & dist.	2.24	1.67	1.99	0.007	0.086
PI	2.49	1.41	1.99	0.027	0.165

comp. & dist.: composite model with disturbance estimation
aug. & dist.: augmented model with disturbance estimation

by having information of the cross coupling dynamics. Also, by being able to control all three basin at the same time, global control will be coordinated to achieve the desired setpoints or performance.

This section presents the design and evaluation of such a controller. The design procedure is similar to the univariate case, but with some obvious differences. However, the performance assessment will not be limited to the last aeration tank, but to the global interaction between basins.

As explained before, the system is composed of three aerated and controllable basins, commanded by three identical decoupled PI controllers as described in section 2.2 and previously shown in Figure (3.7).

4.2.2.1 Controller design

The design of a multivariable MPC controller is very similar to the univariate case. The difference relies in weight selection, which are square, positive semi-definite matrices of appropriate sizes, depending on the number of inputs, outputs and horizons. The easiest approach to tune the weights is to begin with all of them with the same value. After a cycle of simulations, each should be adjusted to eliminate undesirable features in the system response.

In section 3.2.2, it was concluded that the models, which best predicted the system behaviour, were identified using the robust N4SID SV and CVA algorithms. In particular, the models with unmeasurable output disturbance of 9^{th} order (i.e. aug1_9d and aug2_9d). Both models responses are almost identical, so there is no need to perform simulations with both models since any of them will give similar results. Further, a close analysis into their pole locations, reveals that their corresponding poles are very close to each other as in Figure (4.2). Therefore, the controller has been designed using the model identified with the robust N4SID 'SV' algorithm. Table (4.8) summarises the controller parameters which revealed a sufficiently good performance after several simulations.

Table 4.8: Multivariable MPC controller parameters.

Model	H_p	H_u	$\mathscr{R}$	$\mathscr{Q}$
aug1_9d	50	7	diag[1 1 1]	diag[1 1 1 10^{-3} 10^{-3} 10^{-3}]

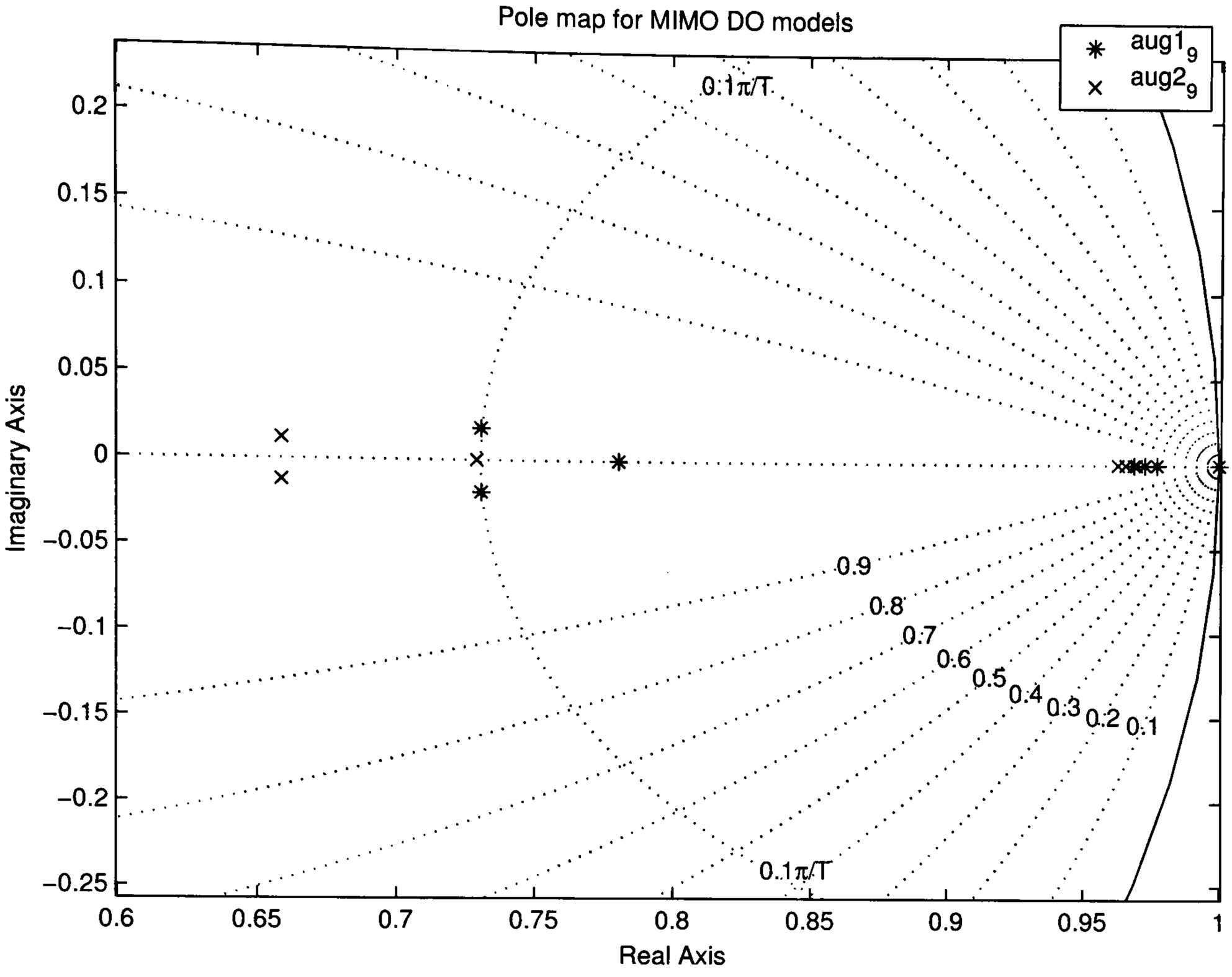

Figure 4.2: Pole map for MIMO DO augmented models.

4.2.2.2 Constraints

The constraints for the multivariable case are the same for the univariate case, as in Table (4.2), but repeated for each input and output.

4.2.2.3 Simulation results

The evaluation of the performance of the multivariable MPC controller follows that used in the univariate case, with the addition of observations over the couplings between the basins. A simple experiment of setpoint changes and output disturbances is presented in Figure (4.3). The setpoint change in each basin occurs at 0.5, 1.5 and 2.5 days respectively. Additionally, an external constant output disturbance of 10% of the blower capacity occurs at 1, 2 and 3 days respectively for each aeration tank. The figure also shows the airflow into the basin measured in terms of the total capacity

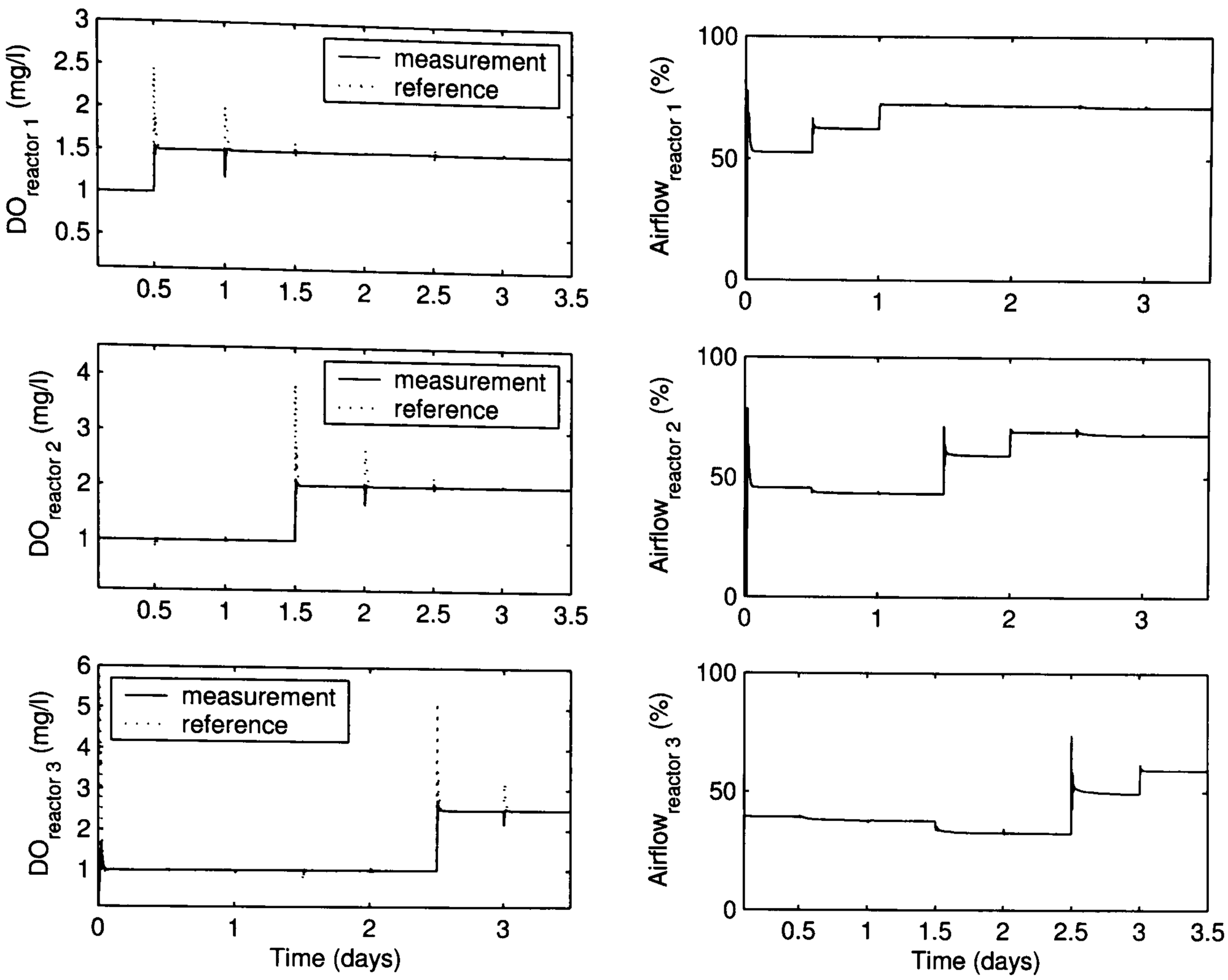

Figure 4.3: Controller dynamic performance: setpoint change and disturbance rejection

of the blower. Notice how the controller tries to compensate for the setpoint changes and disturbances propagated through the system. These effects are quantified in Tables (4.9) and (4.10), while Tables (4.11) to (4.13) present the evaluation of the statistical performance of the controller for different weather conditions. A close examination of these tables for the last aeration basin, reveal that the multivariable controller is much more aggressive in its dynamical performance in a setpoint change, but it will also reduce the effect of a disturbance propagating through the system much faster.

4.3 Nitrogen removal control

One of the main objectives of wastewater treatment is the removal of nutrients in the form of nitrogen and phosphorus from the wastewater. The bio-chemical processes

Table 4.9: Dynamic performance

DO measurement	Overshoot (%)	Settling Time (min)
Aeration tank 1	2.54	18
Aeration tank 2	6.09	25
Aeration tank 3	9.49	35

Table 4.10: Disturbance rejection

DO measurement	Peak (%)	Rejection Time (min)
Aeration tank 1	-21.51	21.99
Aeration tank 2	-17.26	14
Aeration tank 3	-14.56	13

Table 4.11: Dry weather statistics

DO measurement	Max	Min	Mean	Var	St.Dev.
Aeration tank 1	1.52	1.44	1.49	0.0001	0.012
Aeration tank 2.	1.54	1.42	1.49	0.0003	0.016
Aeration tank 3	2.07	1.88	1.99	0.0007	0.025

Table 4.12: Rainy weather statistics

DO measurement	Max	Min	Mean	Var	St.Dev.
Aeration tank 1	1.52	1.44	1.49	0.0001	0.011
Aeration tank 2	1.54	1.43	1.49	0.0001	0.013
Aeration tank 3	2.07	1.89	1.99	0.0004	0.022

Table 4.13: Storm weather statistics

DO measurement	Max	Min	Mean	Var	St.Dev.
Aeration tank 1	1.54	1.44	1.49	0.0001	0.012
Aeration tank 2	1.54	1.42	1.49	0.0002	0.015
Aeration tank 3	2.06	1.88	1.99	0.0006	0.024

involved in nutrient removal are complex, thus their mathematical description. Chapter 2 briefly discussed the complexity of the ASM models (Henze *et al.*, 1987, 1995, 1999; Gujer *et al.*, 1999), which are the most accepted models for activated sludge in the scientific community.

Several publications have reported the use of reduced order modifications of these models for different purposes. For example the model developed by Jeppsson (1995) with the purpose of control, or Huang and Hao (1996) for alternating aerobic-anoxic process evaluation, are all derived from the original ASM 1.

The main problem of using reduced order models directly derived from the full ASM 1

is that, in general, they have poor parameter identifiability characteristics; which in turn requires the use of sophisticated estimation techniques as Kalman filtering (Jeppsson, 1996; Arnold and Dietze, 2001), or H_∞(Katebi, 2001). The use of these techniques however will introduce a higher degree of complexity to the control of the system.

At the moment, most plants rely on classical control methods as PI, or simple operation rules gained from past experience and some process knowledge. Very few publications report the use of advanced process control methods for nitrogen removal, as predictive control, basically due to the highly nonlinear behaviour and constant changing conditions.

A close search within the literature reveals that wastewater process engineers, control engineers, and scientist have been unable to successfully provide a clear understanding of which is the best '*road*' to follow when trying to improve nitrogen removal. Apparently, the only concensus appears to be for the need of a feedforward-feedback approach. For example, Vrecko *et al.* (2003) proposes the use of a feedforward-feedback PI controller achieving reasonably good results.

Sometime earlier, Lindberg (1997) employed an optimal LQ feedforward-feedback approach to control nitrogen removal using a linear model identified using subspace identification. It is interesting to find that one of the author's main conclusions is that the controller performance will highly depend on the quality of the model, thus making the identification procedure of critical importance.

More recently, Alex *et al.* (2002) and Sotomayor and Garcia (2002), have reported on the design of MPC controllers. Both control structures employ a feedforward-feedback control strategy; however, Alex *et al.* (2002) employs a non-linear model, while Sotomayor and Garcia (2002) employ a linear approach. Vrecko *et al.* (2003) and Alex *et al.* (2002) have employed the same simulation benchmark; therefore making it possible to compare the performance of both controllers. Even though the non-linear MPC strategy is more sophisticated than the feedforward-feedback PI controller, the latter

achieves a higher nutrient removal rate with less effluent standards violations. Unfortunately, the linear MPC of Sotomayor and Garcia (2002) did not employ the COST simulation benchmark; therefore, making it impossible to compare with the other two strategies.

Another important concern in nutrient removal is the optimisation of the resources employed. Most plants work with fixed DO setpoints and continuous aeration, probably only slightly varying the setpoint in accordance with the seasons. This approach is however rather inefficient in energy terms. Due to the cyclic nature of the influent load, there are periods of time in which it is even possible to turn off the aeration, or decrease just to provide adequate mixing, without violating the effluent standards. It is estimated that it is possible to save around 10% of energy costs by using intermittent aeration. The optimal use of this type control approach is the subject of major research at the moment. Several authors have reported different approaches to implement this type of control structure achieving some degree of success. Kim *et al.* (2000) presents a predictive approach used in an experimental test-bench. The method determines the aeration time (phase length) by minimising the monetary cost subject to constraints representing the effluent standards. Puta *et al.* (1999), minimises a similar cost function, but using the full ASM1 model.

This section presents the development and design of two predictive control strategies for nitrogen removal. The first approach considers an alternating aeration control scheme. The switching model developed in section 3.3.1 has been used, and the predictive controller designed to operate over the mean prediction of the effluent (zero frequency component), thus being more consistent with ASM1 limitations. The second part of the section, explores the design of a MPC with continuous aeration using the model identified in section 3.3.2. The MPC controller incorporates an influent prediction model as a feedforward component to improve control actions.

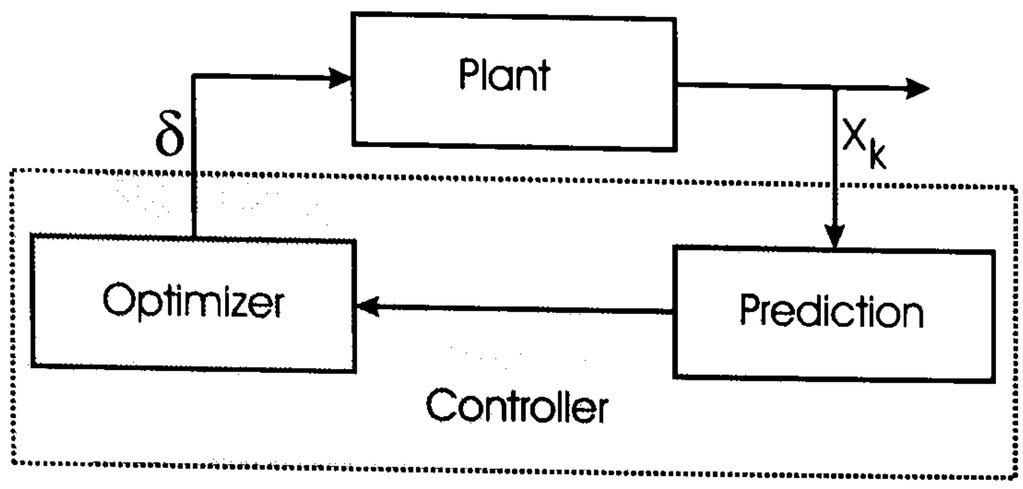

Figure 4.4: Controller Architecture.

4.3.1 Alternating aeration predictive control

One of the most recent control structures under research is the alternating aerobic-anoxic approach. Under this control scheme, switching the aeration system ON and OFF indirectly controls the effluent characteristics.

Some theoretical and experimental studies developed by Kim *et al.* (2000) and Puta *et al.* (1999) find the optimal switching times in the sense of minimising a monetary cost function using a mathematical model of the plant. In particular in (Kim *et al.*, 2000) the optimisation is performed once per day assuming constant influent conditions. The results indicate, that the level of prediction is very limited, and is therefore giving a poor control performance. Also, the controller kept the aeration ON most of the time without considering periods of time in which aeration was unnecessary.

The control approach discussed in this section, uses a linear time-variant modification of the model developed by Anderson *et al.* (2000) and finds the optimal switching time under zero frequency signal tracking. This is approach is considered to be more consistent with the limitations originally drafted in the development of ASM1 regarding average operating conditions.

4.3.1.1 Controller architecture

The controller architecture follows the line of traditional MPC, with some modifications due to the cyclic nature of the embedded model. A block diagram of the controller is presented in Figure (4.4).

4.3.1.2 Prediction.

The alternating aerobic-anoxic (AAA) wastewater time variant system, described by equation (3.58), can be sampled at a specific frequency, and therefore be transformed into a discrete system as presented in equation (4.38).

$$\begin{aligned} x[k+1] &= A[k,\delta]\cdot x[k] + D[k,\delta]\cdot x_{inf} \\ y[k] &= C\cdot x[k] \end{aligned} \tag{4.38}$$

Using a recursive approach to calculate the predicted state of the system based on measurements of the state at sampling time k, it can be easily shown that the predictions at any future sampling instant $k+n$ and up to the prediction horizon H_p can be calculated as in equation (4.39). It should be clear, that full state measurement is assumed. This is an assumption which is difficult to overcome in practice. Possible solutions to this problem are the use of soft-sensors using estimation algorithms. Some research work has been carried out in this field, which usually concludes that the main limitation is the identifiability of the process (model). Some examples of proposed estimators can be found in Katebi (2001); Jeppsson (1995); Arnold and Dietze (2001); Lindberg (1997).

$$\begin{aligned} \hat{x}[k+1\mid k] &= A[k,\delta]\cdot x[k] + D[k,\delta]\cdot x_{inf} \\ \hat{x}[k+2\mid k] &= A[k+1,\delta]\cdot \hat{x}[k+1\mid k] + D[k+1,\delta]\cdot x_{inf} \\ &= A[k+1,\delta]\cdot A[k,\delta]\cdot x[k] + (A[k+1,\delta]\cdot D[k] + D[k+1])\cdot x_{inf} \\ &\vdots \\ \hat{x}[k+n\mid k] &= \prod_{i=1}^{n} A[k+n-i]\cdot x[k] + \sum_{j=1}^{n}\Big(\prod_{i=1}^{n-j} A[k+n-i]\cdot D[k-1+j]\Big)\cdot x_{inf} \\ &\vdots \\ \hat{x}[k+Hp\mid k] &= \prod_{i=1}^{Hp} A[k+Hp-i]\cdot x[k] + \sum_{j=1}^{Hp}\Big(\prod_{i=1}^{Hp-j} A[k+Hp-i]\cdot D[k-1+j]\Big)\cdot x_{inf} \end{aligned} \tag{4.39}$$

Equation (4.39) can be arranged into a matrix representation, and output predictions can be calculated as presented in equations (4.40-4.41).

$$\begin{bmatrix} \hat{x}[k+1 \mid k] \\ \vdots \\ \hat{x}[k+n] \mid k] \\ \vdots \\ \hat{x}[k+Hp \mid k] \end{bmatrix} = \begin{bmatrix} A[k,\delta] \\ \vdots \\ \prod_{i=1}^{n} A[k+n-i] \\ \vdots \\ \prod_{i=1}^{Hp} A[k+Hp-i] \end{bmatrix} \cdot x[k] + \tag{4.40}$$

$$\begin{bmatrix} D[k,\delta] \\ \vdots \\ \sum_{j=1}^{n} \left[\prod_{i=1}^{n-j} A[k+n-i] \cdot D[k-1+j] \right] \\ \vdots \\ \sum_{j=1}^{Hp} \left[\prod_{i=1}^{Hp-j} A[k+Hp-i] \cdot D[k-1+j] \right] \end{bmatrix} \cdot x_{inf}$$

$$\begin{bmatrix} \hat{y}[k+1 \mid k] \\ \vdots \\ \hat{y}[k+n \mid k] \\ \vdots \\ \hat{y}[k+Hp \mid k] \end{bmatrix} = \begin{bmatrix} C & 0 & \cdots & 0 \\ 0 & C & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \end{bmatrix} \cdot \begin{bmatrix} \hat{x}[k+1 \mid k] \\ \vdots \\ \hat{x}[k+n \mid k] \\ \vdots \\ \hat{x}[k+Hp \mid k] \end{bmatrix} \tag{4.41}$$

This, in an abbreviated notation, can be written as in equation (4.42), where n varies between 1 and Hp, and denotes the row element numbering.

$$\begin{aligned} \hat{X} &= A[k,n] \cdot x[k] + D[k,n] \cdot x_{inf} \\ \hat{Y} &= C \cdot \hat{X} \end{aligned} \tag{4.42}$$

4.3.1.3 Cost Function and Optimisation

The selection of an appropriate cost function depends on many factors. However, the approach developed in this paper makes use of the quadratic error of the average value (zero frequency component) of the predictions over a complete aerobic-anoxic cycle and the set point or a reference trajectory to approach the set point. Also, only the unconstrained case is analysed.

In order to calculate the output predictions of the AAA system the algorithm described by equation (4.40) is used. The average value of each state variable arranged in the vector representation presented in equation (3.53) is calculated by the average of the predictions over the horizon Hp. Equation (4.43) shows how the average (zero frequency component) of a discrete vector signal is calculated.

$$\bar{y} = \frac{1}{H_p} \sum_{i=1}^{Hp} \hat{y}[k+i \mid k] \tag{4.43}$$

where $\hat{y}$[k+i] and $\bar{y}\ \varepsilon\ \Re^8$. Therefore, using equation (4.43) the cost function is defined as follows.

$$J \;=\; (\bar{y}-\gamma)^T \cdot Q \cdot (\bar{y}-\gamma) \tag{4.44}$$

where γ could be a reference trajectory to approach the set point, which is updated on each prediction cycle, or the set point; and Q is a weights matrix of adequate dimensions which can be time dependent and used to include penalising functions. For this case Q has been considered to be the identity matrix I.

An additional term can be added to equation (4.44), to penalise steep changes in the control input, and represent a minimum energy consumption approach. The final cost function is as presented in equation (4.45). In this equation δ is the parameter to be optimised at prediction cycle k+1, while δ_k is the optimal δ value found at prediction cycle k.

$$J = (\bar{y}-\gamma)^T \cdot (\bar{y}-\gamma) + (\delta - \delta_k)^2 \quad (4.45)$$

Finally, the search for the optimum value of the duty cycle δ that minimises equation (4.45) can be done using several numerical methods, since it is difficult to find a closed analytical representation for the gradient.

Additionally the model obtained contains state-dependent nonlinearities for which other predictive control formulations have been proposed as in (Grimble and Ordys, 2001; Ordys and Grimble, 2001).

4.3.1.4 Simulation results

Two types of simulations have been performed. The first one considers a reference trajectory to approach the set point, and the second uses the set point directly. For the reference trajectory case, a time constant of 12 hours with a sampling time of 3 hours has been chosen. The prediction horizon is of one cycle (3 hours) for both cases. The reference trajectory at cycle n can then be calculated as in Maciejowski (2001):

$$\gamma_{n+1} = s - e^{-(n+1)\cdot\frac{T_s}{T_{ref}}} \cdot e_k \quad (4.46)$$

where e_k is the error between the plant output and the set point s at prediction instant k. The set point for S_{NH} and S_{NO} are presented in Table 4.14. The system initial conditions are the influent characteristics, presented in Anderson *et al.* (2000).

Simulation results, for the case in which a reference trajectory is used are presented in

Table 4.14: Controller setpoints

S_{NH}	S_{NO}
1.26 [mg/l]	5.31 [mg/l]

Figure (4.5), and Figure (4.6) shows simulations for the case in which the set point is used directly.

It is interesting to observe that in the case in which a reference trajectory is used, the optimal control input δ begins with a lower value than in the case of the use of the set point directly. The cost function also seems to converge to the minimal value faster when using the reference trajectory. The presented simulations also show that once the system is near the setpoint, the control input begins to oscillate around the optimal value. A possible explanation for this behaviour is that the control system (optimiser) is not able to keep both controlled variables at the same time in the exact setpoints, but in a near neighbourhood.

An important limitation of this method is that the maximum output magnitude cannot be controlled directly. This means that at certain periods of time the output concentrations are higher than the permissible. Some possible formulations to solve this problem can be the use of constraints to limit the effluent concentrations and the calculation of an adequate set point.

There are several ways in which constraints can be included, but probably the use of the weight matrix Q, which in this case has been assumed to be the identity matrix (I), is the easiest. This matrix could include time-dependent penalisation functions, which include the constraints in the quadratic cost function.

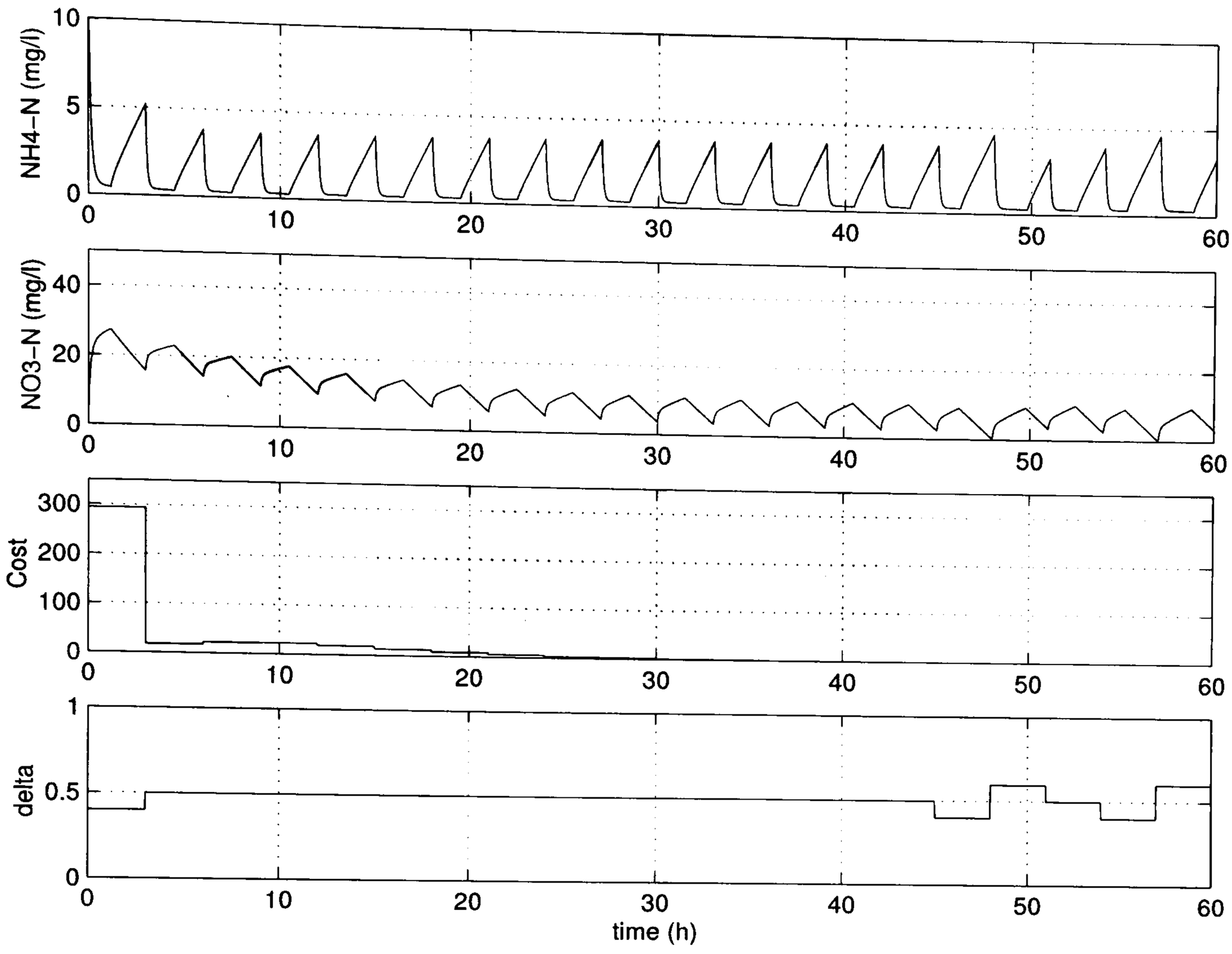

Figure 4.5: System response with reference trajectory.

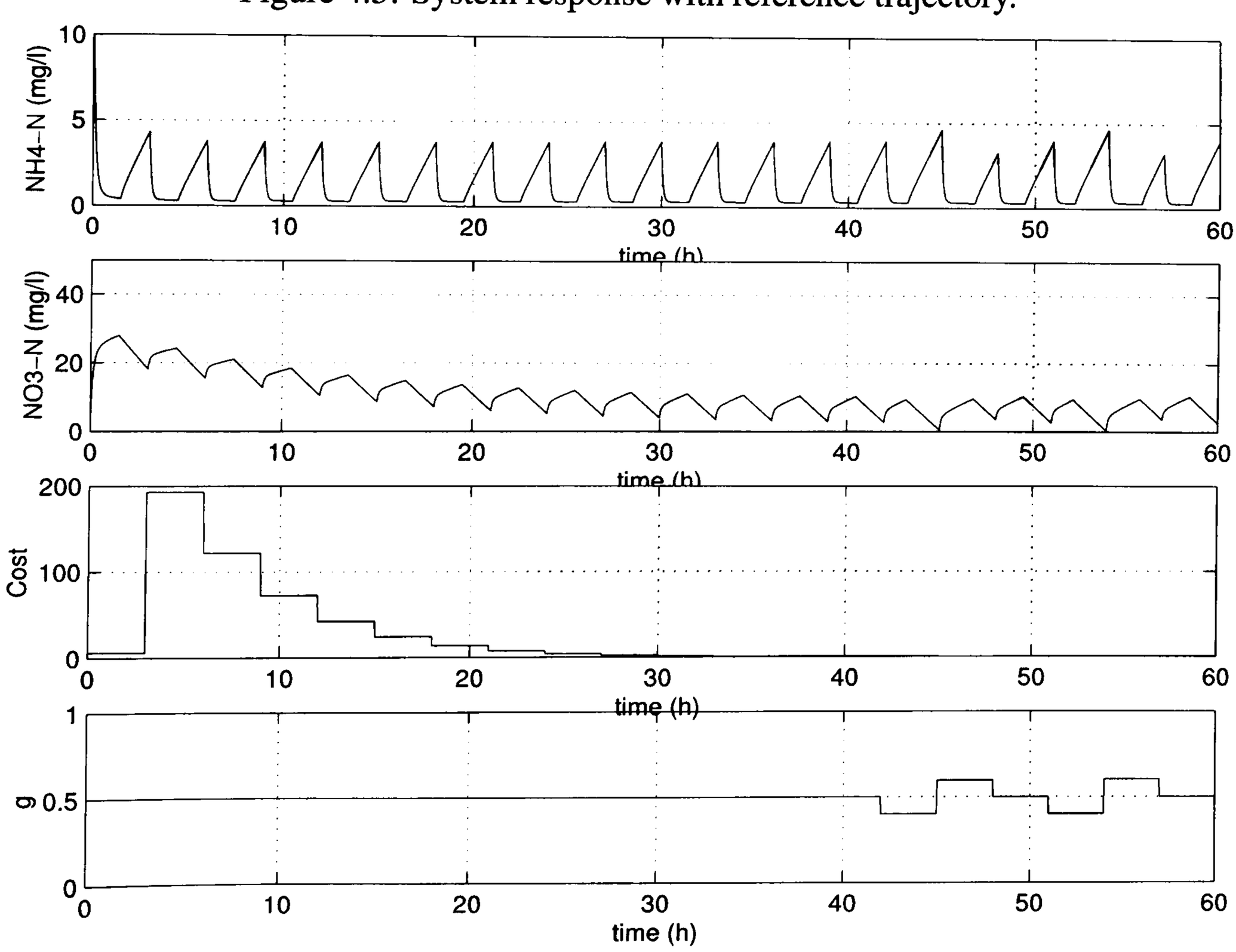

Figure 4.6: System response with no reference trajectory.

4.3.2 Continuous aeration model predictive control

The use of model predictive control for nutrient control has been very limited due to the complexity of the ASM models. Even though ASM models represent the state of the art in activated sludge understanding, they are far from being acceptable for control purposes due to their complexity and un-identifiability. Several authors have suggested simplifications these models, in particular ASM1 for control purposes. However, they still exhibit some rather peculiar nonlinear behaviours. As remarked in the original ASM1 report, the model could produce invalid calculations compared to the actual process, due to the heavy uncertainty in its parameters and unmodelled dynamics.

Therefore some other ways of obtaining more numerically efficient models which provide reliable predictions of the actual state of the plant would be of great help. Further, if this model is linear, then it would be possible to simplify even more the control problem. This is, however, not always possible and actually rather difficult. Section 3.3.2, presented a possibility in which a 4^{th} order model was identified for the prediction of total nitrogen (TN), nitrate (NO_3), ammonia (NH_4) in the effluent and nitrate ($NO_{3_{anoxic}}$) in the anoxic zone, using the internal recirculation and three dissolved oxygen (DO) measurements in the three aerated basins of the COST simulation benchmark. The model considers as well, the influent flow (F) and ammonia ($NH_{4_{influent}}$) are measured, and incorporated into the prediction by using the harmonic model approximation presented in the same section. This is a new strategy, which has been reported in Nielsen (2001), and has been under research to estimate weather conditions (influent flow) for Aeration Tank Settling (ATS). The modelling of the processes in ATS conveys a wide scope of process knowledge, involving a redistribution of the suspended solids within the plant. This is however a subject beyond the scope of this thesis.

The section is organised in a similar way to the MPC design for DO control. The controller design is the first part to be discussed, followed by the formulation of constraints based on the effluent standards and control system limitations. Finally simulations re-

sults are presented and discussed.

4.3.2.1 Controller design

The design of the controller is considerably more complicated than for the case of dissolved oxygen. However, most of the effort has been demanded in obtaining the model. Much effort has also been placed in calibrating the controller; thus the design presented in this section is the result of a number of simulations.

The model obtained is composed by four control inputs: the dissolved oxygen concentrations in the las three aeration basins, and the recirculation flow rate. The outputs of the model are: total nitrogen in the effluent, nitrate in the effluent, ammonia in the effluent, and nitrate at the end of the anoxic zone. Additionally, two disturbances are measured: influent flow and influent ammonia concentration. The controller will then be a feedforward-feedback type; where the influent flow and ammonia are the feedforward signals.

One of the most limiting factors in the design of the controller is the determination of the constraints for the optimisation problem. Given the effluent standards, one approach would be to impose hard constraints on the effluent concentrations. This is however impractical, since there will be many times in which the optimisation problem will be un-feasible. The simplest approach would just be of using a setpoint and weighting the most important signals higher. This will not assure that the effluent standards will not be violated; however, it will provide a feasible solution for most conditions.

Also, the important factor to analyse in this situation is not the number of violation; but the total amount of nutrients discharged over a period of time. Therefore, even if there is an effluent violation at a certain time, the might be other times in which the effluent concentration is also very low. Thus, the total amount of nutrients discharged will not be as high as supposed.

Table 4.15: Nutrient MPC controller parameters.

H_p	H_u	$\mathscr{R}$	$\mathscr{Q}$
40	40	diag[2 1.5 1 0.8]	diag[1.5 0.1 0.1 1.5]

Table 4.16: Controller setpoints

TN	$NO3_{effluent}$	$NO3_{anoxic}$	$NH4_{effluent}$
15 [mg N/l]	10 [mg N/l]	3 [mg N/l]	2.5 [mg N/l]

Given these reasons, the nutrient MPC controller parameter are presented in Table (4.15), and the setpoints for the different concentrations are presented in Table (4.16).

4.3.2.2 Simulation results

As described previously, simulations are only run for the case of dry weather flow. Figures (4.7) and (4.8) show the manipulated variables over a half day period and the effluent nutrient concentrations over a 7 day period. Notice, that the effluent concentration of total nitrogen (TN) is most of the time kept under an 18 (mg/l) and only with a limited amount of violations. Similarly ammonia (NH_4), is also kept at a low level most of the time, except for sporadic periods of time. In general, there will be seven ammonia peaks per week in dry weather conditions, due to the cyclic behaviour of the influent.

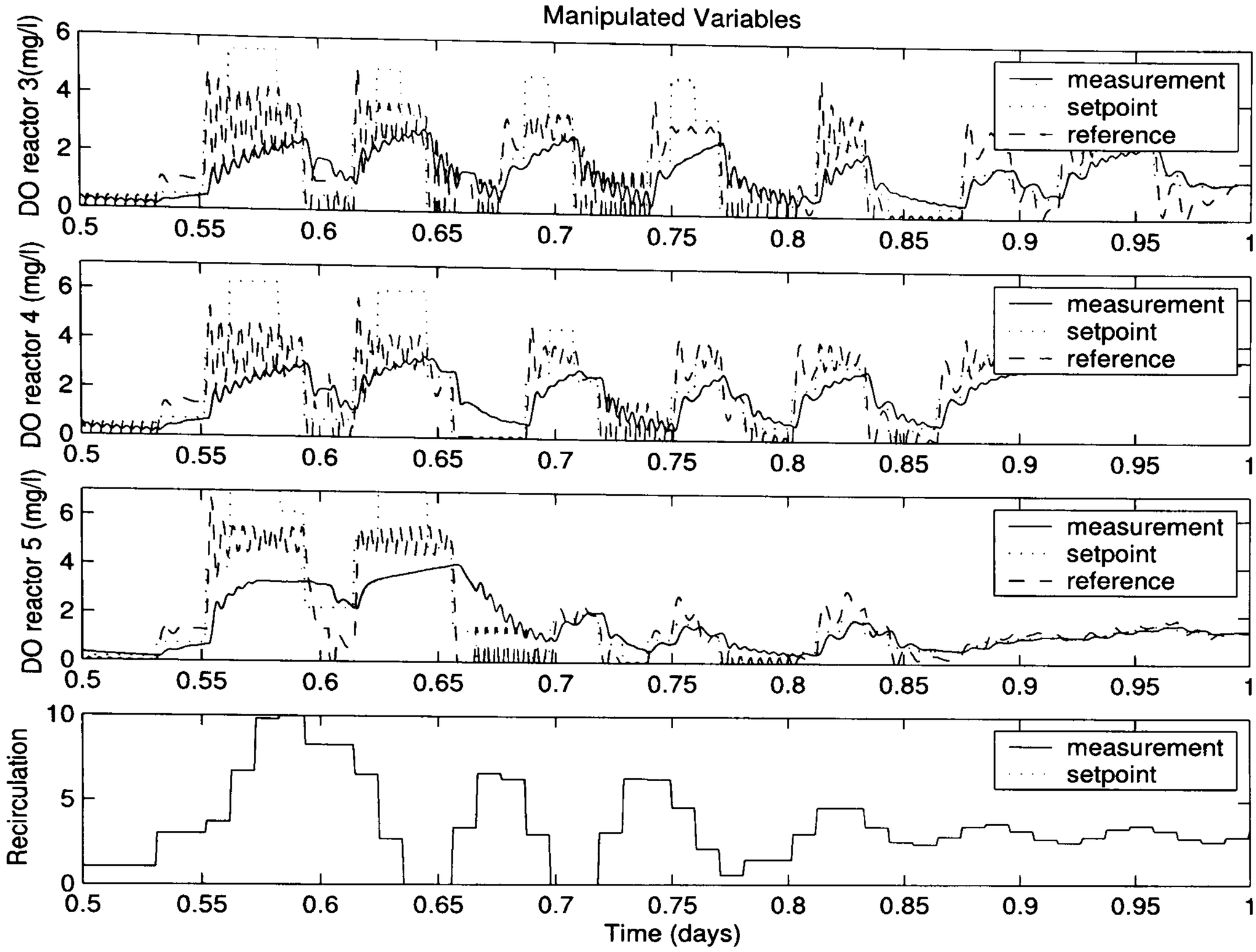

Figure 4.7: Manipulated variables

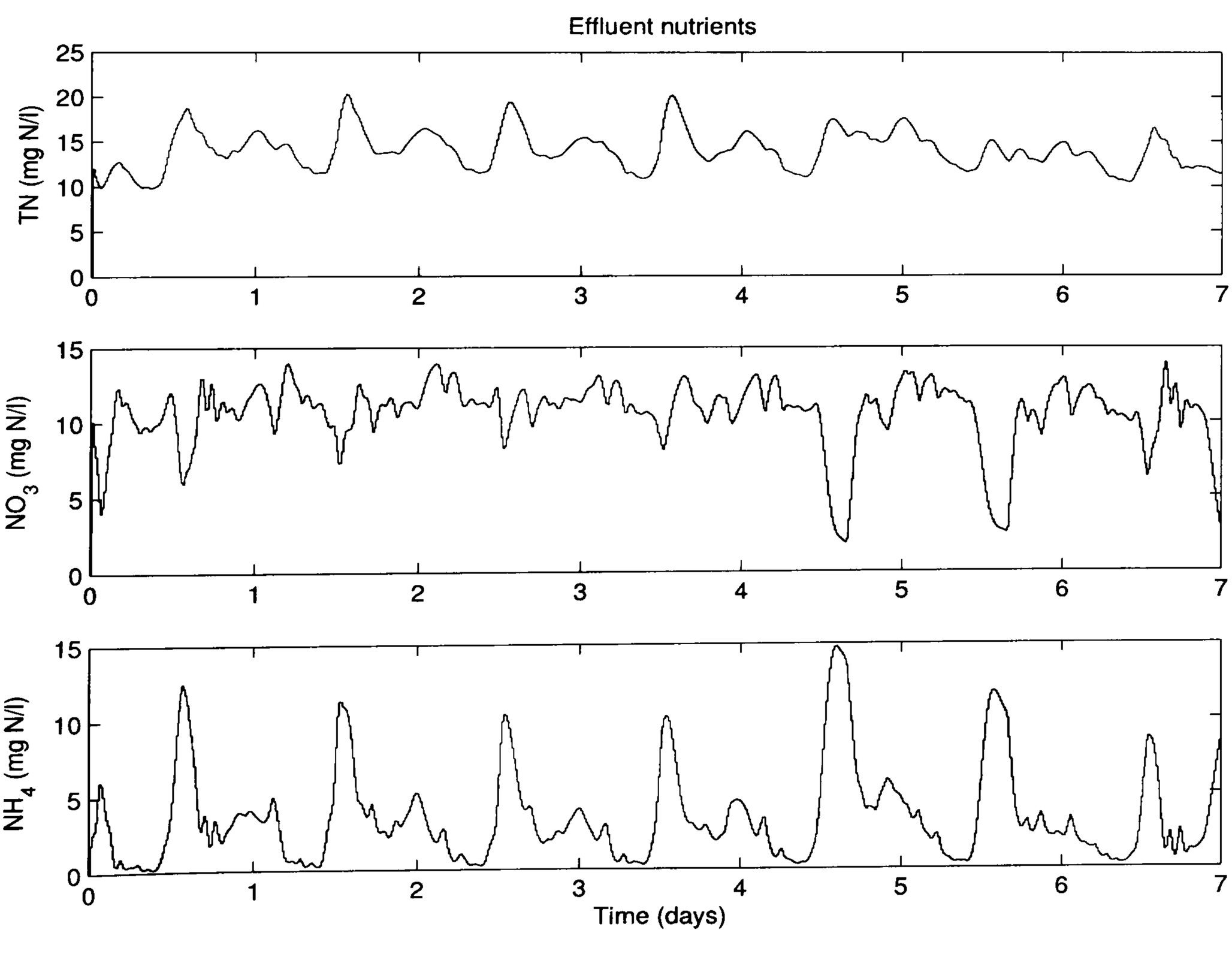

Figure 4.8: Effluent concentrations for total nitrogen (TN), nitrate (NO_3) and ammonia (NH_4).

4.4 Summary

This chapter has presented the design of model predictive controllers for dissolved oxygen and nutrient removal. The controllers have been designed employing linear models obtained in Chapter 3. The following is a summary of the simulation study results presented in this chapter:

Dissolved oxygen control

For dissolved oxygen, two cases are examined: (a) a univariate case and (b) a multivariable case. The designed controllers have been assessed by using their dynamical response to setpoint changes and their ability to reject disturbances. Also, a statistical assessment is performed by using different weather influent conditions: dry, rain, and storm.

The use of augmented models, as defined in Chapter 3, has proven to be the most adequate type of model by the results obtained. Also, a close comparison between the results obtained in the univariate case and in the multivariable case, demonstrates that a multivariable approach can be more beneficial to the system, since loop coordination helps to improve setpoint achievement in a more efficient and fast way. Disturbance rejection also presents a significant improvement.

Nutrient removal control

For the nutrient case two types of controller have been designed: one for continuous aeration and a second for an alternating plant. The performance of the controller is assessed only for the dry weather case due to the highly nonlinear behaviour.

For the continuous aeration nutrient controller, the major difficulty presented is in obtaining a reasonably good model. The design of a model predictive controller, as

formulated in this chapter, requires that the model should give good predictions; otherwise the performance of the controller is severely deteriorated. Simulations indicate nutrient control using a linear model is possible under dry weather flow. Other type of disturbances have not been simulated due to (a) the high non-linearity of the system (b) the control objective in a plant under these disturbances is not to reduce only pollutant load, but to maintain sludge inventories in order to preserve the treatment capability of the plant. Therefore, any simulation under these conditions would be unrealistic for the operation of a plant.

In the intermittent aeration case, the formulation of a model predictive control algorithm using a switching model structure has been the main aim. The controller calculates a duty factor which indicates the relationship between the ON and the OFF aeration period. By controlling the aeration cycles, it is possible to indirectly control the level of nitrate and ammonia in the effluent of the reactor.

Chapter 5

IFT and LQG tuning and process loop monitoring

Much of the process industry's tuning paradigm originated with the two Ziegler-Nichols design methods (Ziegler and Nichols, 1942), which were developed for simple and effective PID controller tuning. The step-response method assumes a first order plus time delay (FOPTD) system model, which is a common type of response found in the process industries. The step response is used to identify implicitly this parametric model from which design rules are deduced. The second of the Ziegler-Nichols methods is the sustained oscillation method. In this method the critical stability point on the process's Nyquist frequency response is found and the PID design rules then use the ultimate period and ultimate gain as data. This second method is an example of a non-parametric model based design.

Some of the more recent parametric model based design techniques use the idea of trying to make the PID response as close as possible to that obtainable from more sophisticated methods based on models (Katebi and Moradi, 2001; Moradi *et al.*, 2002; Uduehi *et al.*, 2002). Other parametric methods approach the design of PID controllers by using the method of optimal restricted-structure controllers. In this route a linear quadratic gaussian (LQG) controller design is restricted to one of limited parameters

as for example as in (Johnson and Sánchez, 2003), which is the method presented in the second part of this chapter. The approach considers some prior modelling or identification experience, and uses the key assumption that the process could be described using a simple first order stochastic model. This assumption is shown to lead to explicit formula for optimal LQG controllers and cost function values used in benchmarking. Furthermore, the resulting controllers are shown to be of PID-type.

By way of complete contrast perhaps the only genuinely known model-free method currently available is that of iterative feedback tuning(IFT) (Hjalmarsson *et al.*, 1994, 1995, 1998; Hjalmarsson, 1999; Mahathanakiet *et al.*, 2002; Ho *et al.*, 2003). This method uses the system directly to generate all the data responses needed to perform an online optimisation of the restricted structure controller parameters. Recently, controller parameter cycling (CPC) has been presented as a new way of implementing the model-free optimisation scheme of IFT (Crowe *et al.*, 2003).

As exemplified in previous chapters, the control problems arising in modern activated sludge wastewater treatment plants include unknown process models, noisy measurements, highly unpredictable process disturbance inputs and a corrosive measurement environment leading to slow sensor degradation and failure. The control loops, are usually of PID-type. Consequently, the prudent control approach is to encourage the development and use of reliable equipment and to support control loops with tuning and monitoring software algorithms able to indicate when controller settings and measurement devices should be examined for possible readjustment.

This chapter presents two methods with direct application to wastewater treatment plants. The first section considers a deterministic, continuous-time formulation for IFT and the development of adequate algorithms for simulation and implementation. The second introduces a LQG approach for tuning PID controllers based in optimal polynomial theory. Both cases assume a single input - single output process model. Finally a process loop monitoring algorithm, based on an LQG criterion, is developed. This algorithm allows the implementation of a monitoring scheme to assess the perfor-

mance of running loops in real-time in a wastewater treatment plant.

5.1 Iterative feedback tuning (IFT)

Much recent research in the control community has focused on finding viable approaches for using nonlinear control in industrial applications. Most of this work assumes that some form of mathematical model will be available, and therein is the real applications difficulty; nonlinear models are often hard to derive and can be expensive to develop. If a model is available, the use of nonlinear model based predictive control is one method in particular vogue at present (Mayne *et al.*, 2000; Mayne, 2001).

A different approach to generating routine controllers for industrial plant is to use only a little model information and rely on the robustness of the controller for success. The online relay experiment method of Åström and Hägglund (1985) was particularly successful in industrial applications. At the heart of the relay method is a non-parametric identification principle. It was the simplicity of this approach which inspired others to develop extensions (Yu, 1999) and alternative non-parametric methods for automated industrial three term controller design (Crowe and Johnson, 1998, 1999, 2000).

In 1994, Hjalmarsson *et al.* published the first in a series of papers on a model-free approach to restricted complexity controller design (Hjalmarsson *et al.*, 1994). The method was termed the Iterative Feedback Tuning (IFT) method of controller design. The principle behind this approach is to use a set of specific process experiments to produce data which could be used in a stochastic optimisation routine to optimise a simple loop.

The seminal papers for the Iterative Feedback Tuning method due to Hjalmarsson *et al.* (1994, 1998) adopt a fairly general formulation that incorporates the following features:

1. A system description involving a stochastic process output disturbance.

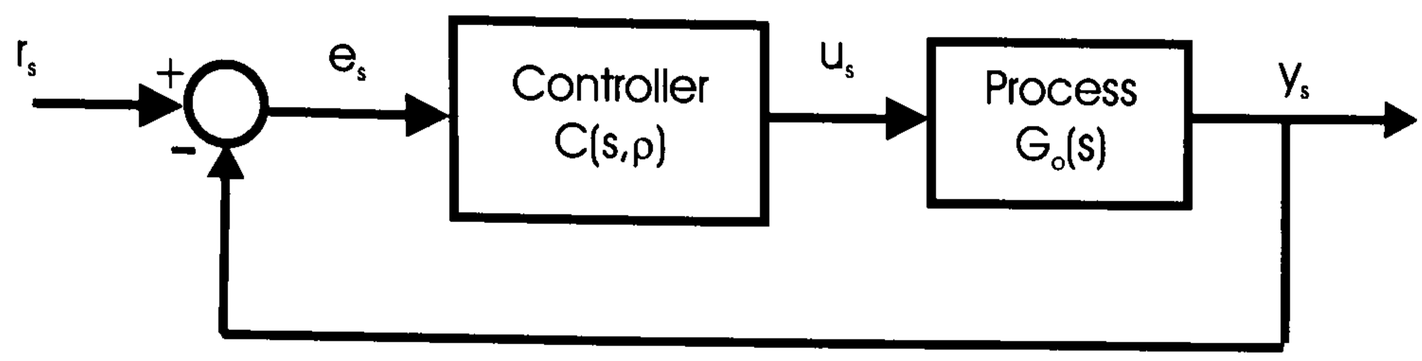

Figure 5.1: Control system with unity feedback and parametrised controller.

2. A two degrees of freedom control law.

3. Use of a stochastic optimisation approach (Robbins and Munro, 1951).

4. A restricted structure control law.

Consequently this level of generality obscures the simplicity of the method, and makes it difficult to investigate some of the issues relevant to industrial or practical implementation. A much easier approach to the method is to formulate a deterministic version of Iterative Feedback Tuning for the case of simple PI control. This section focuses on this case, and provides an algorithm which can be used for iterative tuning of a PI controller. The section also investigates the use and implementation of this algorithm via simulation for the case of a PI controller.

This section is composed of two parts. The first part presents a deterministic formulation for a SISO system. The section includes the iterative optimisation problem and the development of the main IFT algorithm for the calculation of the gradient. Also, a variant to the recursive update of the parameters by using the Hessian is discussed. The second part, develops the theory for a MIMO deterministic IFT tuning algorithm, and contains similar topics as the first part extended to the multivariable case. Finally, simulations results for several case studies are presented in section 5.1.3.

5.1.1 IFT formulation for SISO systems

Consider a single input, single output system driven by a single degree of freedom control law, as presented in Figure (5.1), where the reference, reference error, controller output and the system output signals are denoted by r_s, e_s, u_s, and y_s respectively.

Based on this system description, the following lemmas develop the theory necessary for a deterministic continuous-time IFT algorithm for single input, single output systems.

Notation

The notation for the time domain signals is r_t, e_t, u_t and y_t respectively. The subscript t denotes the time dependence of these signals. The notation for the Laplace transforms of these signals is $r_s = \mathscr{L}\{r_t\}$, $e_s = \mathscr{L}\{e_t\}$, $u_s = \mathscr{L}\{u_t\}$ and $y_s = \mathscr{L}\{y_t\}$.

Lemma 5.1.1. Closed-loop relationships

Given the previous definitions for the system depicted in Figure (5.1), the following relations between the system signals hold:

$$y_S(\rho) = T_o(s) r_s \tag{5.1}$$

$$e_s(\rho) = S_o(s) r_s \tag{5.2}$$

$$u_s(\rho) = S_o(s) C(s,\rho) r_s \tag{5.3}$$

where,

$$T_o(s) = \left[\frac{G_o(s)C(s,\rho)}{1+G_o(s)C(s,\rho)}\right] \tag{5.4}$$

$$S_o(s) = \left[\frac{1}{1+G_o(s)C(s,\rho)}\right] \tag{5.5}$$

Proof. From the diagram, the vector of outputs ($y_s(\rho)$) can be calculated from the vector of errors ($e_s(\rho)$) as in equation (5.6), and the error signal is defined as in equation (5.7).

$$y_s(\rho) = G_o(s)C(s,\rho)e_s(\rho) \tag{5.6}$$

$$e_s(\rho) = r_s(s) - y_s(\rho) \tag{5.7}$$

Therefore, by replacing the output signal into the error definition and reorganising gives equation (5.2). Similarly, replacing the error signal into the output signal and reorganising gives equation (5.1). Likewise, the control output can be calculated from the error signal as in equation (5.8), and using equation (5.2) gives equation (5.3).

$$u_s(\rho) \;=\; C(s,\rho)e_s(\rho) \tag{5.8}$$

□

Remarks:

- The transfer functions S_o and T_o are known as the sensitivity and complementary sensitivity respectively.

- A PI controller, $C(s,\rho)$ can be parameterised as follows,

$$C(s,\rho) \;=\; K_p + \frac{K_i}{s} \tag{5.9}$$

and $\rho_1 = K_p$, $\rho_2 = K_i$, thus $\rho \in \mathbb{R}^2$.

5.1.1.1 Optimisation problem

Consider the cost functional of equation (5.10), where L_e and L_u are weighting filters and λ is a scalar weight.

$$J(\rho) \;=\; \frac{1}{2T_f}\int_0^{T_f}\left\{(L_e e_t)^2 + \lambda\,(L_u u_t)^2\right\}dt \tag{5.10}$$

The IFT optimisation problem can then be formulated as,

$$\min_{\rho} J(\rho) \tag{5.11}$$

subject to $\rho > 0$ and $C(\rho)$ stabilising the closed loop. This problem is known to be the fixed structure (restricted structure) LQ optimal control problem with weighted error and control signals. Incorporating a limit process $T_f \to \infty$ will yield the steady state version of the optimisation problem. The condition $\rho > 0$ is to ensure that the PI controller parameters are positive. To simplify the subsequent analysis, the weighting filters are set to $L_e = I$ and $L_u = I$.

Lemma 5.1.2. Cost gradient

The gradient of the cost functional of equation (5.10) with respect to the controller parameter vector ρ is given by,

$$\frac{\partial J(\rho)}{\partial \rho} = \frac{1}{T_f}\int_0^{T_f}\left\{e_t\frac{\partial e_t}{\partial \rho} + \lambda u_t \frac{\partial u_t}{\partial \rho}\right\}dt \tag{5.12}$$

Proof. The gradient of the cost functional is found by using Leibniz's Theorem for the differentiation of an integral (Abramowitz and Stegun, 1972),

$$\frac{d}{dc}\int_{a(c)}^{b(c)} f(x,c)dx = \int_{a(c)}^{b(c)}\frac{\partial}{\partial c}f(x,c)dx + f(b,c)\frac{db}{dc} - f(a,c)\frac{da}{dc} \tag{5.13}$$

Using this theorem in equation (5.12) proves the lemma. □

The optimisation of the cost functional follows an iterative gradient algorithm, which can be summarised as follows:

Algorithm 5.1.1. IFT numerical optimisation

1. Initialisation

   Choose weighting λ.

   Choose costing time interval T_f.

   Choose convergence tolerance ε.

   Set loop counter $k = 0$.

   Choose initial controller parameter vector $\rho(0) = \rho_o$.

2. Gradient calculation

$$\frac{\partial J}{\partial \rho}(k) = \frac{1}{T_f}\int_0^{T_f}\left\{e_t\frac{\partial e_t}{\partial \rho}(k) + \lambda u_t\frac{\partial u_t}{\partial \rho}(k)\right\}dt$$

3. Check convergence

   If $\left\|\frac{\partial J}{\partial \rho}(k)\right\| < \varepsilon$ then stop.

4. Update parameter vector

   Select or calculate the update parameters γ_k and R_k.

   Compute $\rho(k+1) = \rho(k) - \gamma_k R^{-1}\frac{\partial J}{\partial \rho}(k)$

Some important remarks about this algorithm are:

1. A selection of γ_k is necessary. This can be fixed step or line search step selection.

2. Setting $R = I$ gives an algorithm from the steepest descent family of optimisation routines.

3. Setting $R = H(\rho(k))$ where H is the Hessian matrix, produces a Newton iteration for the optimisation; in this case $H_{ij} = \frac{\partial^2 J}{\rho_i \rho_j}$.

5.1.1.2 IFT implementation

The challenge of implementing IFT relies in calculating the necessary signals for the gradient calculation, directly from the closed-loop system. These signals are e_t, $\frac{\partial e_t}{\partial \rho}$, u_t, and $\frac{\partial u_t}{\partial \rho}$. The necessary steps to achieve this are provided through the following Lemmas.

Lemma 5.1.3. Sub-gradient calculation

The partial derivatives of e_s and u_s are,

$$\frac{\partial e_s}{\partial \rho} = -\left(\frac{1}{C(\rho)}\right)\left(\frac{\partial C(\rho)}{\partial \rho}\right)T_o e_s \tag{5.14}$$

$$\frac{\partial u_s}{\partial \rho} = \left(\frac{\partial C(\rho)}{\partial \rho}\right) S_o e_s \tag{5.15}$$

Proof. Equations (5.14) and (5.15) are found by differentiating equations (5.2) and (5.3). □

By using this lemma, the error signal e_s can be used in a closed-loop identification step to generate the gradient expressions as follows.

Lemma 5.1.4. Signal recording

Let the error signal e_s be used as a reference signal to the closed-loop system, then two response signals can be defined as, $y^{(1)}$ and $u^{(1)}$. Thus, the gradient signals (5.14) and (5.15) can be calculated as,

$$\frac{\partial e_s}{\partial \rho} = -G_{grad}(s,\rho) y^{(1)} \tag{5.16}$$

$$\frac{\partial u_s}{\partial \rho} = G_{grad}(s,\rho) u^{(1)} \tag{5.17}$$

where,

$$G_{grad}(s,\rho) = \left[\left(\frac{1}{C(\rho)}\right)\left(\frac{\partial C(\rho)}{\partial \rho}\right)\right] \tag{5.18}$$

Proof. Consider the block diagram in Figure (5.1). If the error signal e_s is used as a reference signal to the closed-loop system, then two response signals can be defined as $y^{(1)} = T_o e_s$ and $u^{(1)} = S_o C(s,\rho) e_s$. Replacing this two responses into equations (5.14) and (5.15) proves the lemma. □

This lemma can be interpreted as in Figure (5.2) and used for the purpose of simulation.

If the controller is a PI with a parametrisation as in equation (5.9) , then equation (5.18) can be written as,

$$G_{grad}(s,\rho) = \begin{bmatrix} \frac{s}{K_p s + K_i} & \frac{1}{K_p s + K_i} \end{bmatrix} \tag{5.19}$$

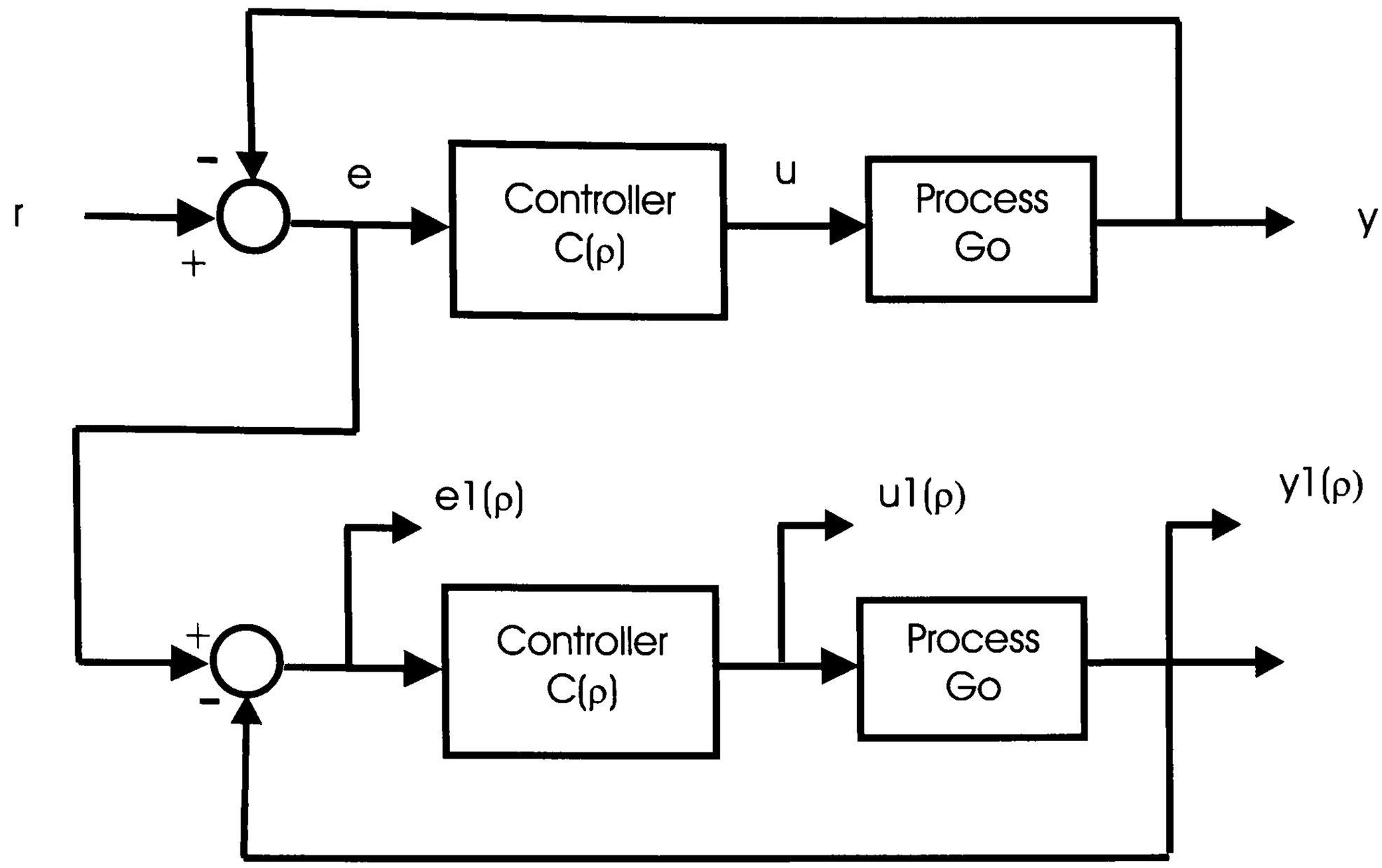

Figure 5.2: Signal generation for gradient calculation.

Lemmas 5.1.3 and 5.1.4 have been derived in the Laplace domain, but they indicate how the gradient of equation (5.12) is to be calculated. This requires one set of system responses due to the reference r, and a second set of system responses due to the input e. Therefore, the algorithm for the gradient computation at iteration step k^{th} is as follows.

Algorithm 5.1.2. Gradient computation, k^{th} step

1. Setup controller using $\rho(k)$

2. Responses

   Run closed-loop system with reference input r.

   Record signals, e_t and u_t.

   Run closed-loop system with reference input e_t

   Record signals $y_t^{(1)}$ and $u_t^{(1)}$.

3. Processing responses

Process the recorded signals using,

$$\begin{aligned}\frac{\partial e_s}{\partial \rho} &= -G_{grad}(s,\rho)y_s^{(1)} \\ \frac{\partial u_s}{\partial \rho} &= G_{grad}(s,\rho)u_s^{(1)}\end{aligned}$$

4. Use time domain formula to compute gradient,

$$\frac{\partial J}{\partial \rho}(k) = \frac{1}{T_f}\int_0^{T_f}\left\{e_t\frac{\partial e_t}{\partial \rho}(k) + \lambda u_t\frac{\partial u_t}{\partial \rho}(k)\right\}dt$$

In practice, careful data collection procedures can automate the processing required to generate the gradient. This requires routine running in parallel with the data collection. In general, a set of two consecutive experiments are required.

5.1.1.3 Hessian calculation

The Hessian can be used during the parameter update to obtain an optimisation algorithm of the Newton type. The Hessian involves the calculation of second order partial derivatives of the cost function and is defined in equation (5.20). The following lemma provides a method to calculate the Hessian.

$$H(\rho) = \begin{bmatrix} \frac{\partial^2 J}{\partial \rho_1 \partial \rho_1} & \frac{\partial^2 J}{\partial \rho_1 \partial \rho_2} & \cdots & \frac{\partial^2 J}{\partial \rho_1 \partial \rho_n} \\ \frac{\partial^2 J}{\partial \rho_2 \partial \rho_1} & \frac{\partial^2 J}{\partial \rho_2 \partial \rho_2} & \cdots & \frac{\partial^2 J}{\partial \rho_2 \partial \rho_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J}{\partial \rho_n \partial \rho_1} & \frac{\partial^2 J}{\partial \rho_n \partial \rho_2} & \cdots & \frac{\partial^2 J}{\partial \rho_n \partial \rho_n} \end{bmatrix} \tag{5.20}$$

Lemma 5.1.5. Hessian calculation

Let the Hessian be noted as $H(\rho)$ and defined as in equation (5.20). Each element of the Hessian is calculated by using equation (5.21).

$$\frac{\partial}{\partial \rho_i}\left(\frac{\partial J}{\partial \rho_j}\right) = \frac{1}{T_f}\int_0^{T_f}\left\{\left[e_t\frac{\partial}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right)+\frac{\partial e_t}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right)\right]\right. \tag{5.21}$$
$$\left.+\lambda\left[u_t\frac{\partial}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right)+\frac{\partial u_t}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right)\right]\right\}dt$$

where,

$$\frac{\partial}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right) = -\frac{\partial\left(G_{grad}\right)_j}{\partial \rho_i}y^{(1)} - \left(G_{grad}\right)_j\frac{\partial y^{(1)}}{\partial \rho_i} \tag{5.22}$$
$$\frac{\partial}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right) = \frac{\partial\left(G_{grad}\right)_j}{\partial \rho_i}u^{(1)} + \left(G_{grad}\right)_j\frac{\partial u^{(1)}}{\partial \rho_i} \tag{5.23}$$
$$\frac{\partial y^{(1)}}{\partial \rho_i} = \left(G_{grad}\right)_i\left[y^{(1)} - 2y^{(2)}\right] \tag{5.24}$$
$$\frac{\partial u^{(1)}}{\partial \rho_i} = \left(G_{grad}\right)_i\left[u^{(1)} - 2u^{(2)}\right] \tag{5.25}$$

and $\left(G_{grad}\right)_j$ is the j^{th} element of G_{grad} as defined previously in equation (5.18).

Proof. Differentiating the cost gradient with respect to a controller parameter ρ_i gives,

$$\frac{\partial}{\partial \rho_i}\left(\frac{\partial J}{\partial \rho_j}\right) = \frac{1}{T_f}\int_0^{T_f}\left\{\left[e_t\frac{\partial}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right)+\frac{\partial e_t}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right)\right]\right. \tag{5.26}$$
$$\left.+\lambda\left[u_t\frac{\partial}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right)+\frac{\partial u_t}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right)\right]\right\}dt$$

To calculate the second partial derivatives coefficients, differentiate equations (5.16) and (5.17), thus giving,

$$\frac{\partial}{\partial \rho_i}\left(\frac{\partial e_s}{\partial \rho_j}\right) = -\frac{\partial\left(G_{grad}\right)_j}{\partial \rho_i}y^{(1)} - \left(G_{grad}\right)_j\frac{\partial y^{(1)}}{\partial \rho_i} \tag{5.27}$$
$$\frac{\partial}{\partial \rho_i}\left(\frac{\partial u_s}{\partial \rho_j}\right) = \frac{\partial\left(G_{grad}\right)_j}{\partial \rho_i}u^{(1)} + \left(G_{grad}\right)_j\frac{\partial u^{(1)}}{\partial \rho_i} \tag{5.28}$$

To calculate $\frac{\partial y_s^{(1)}}{\partial \rho_i}$ recall that $y_s^{(1)} = T_o e_s$, thus

$$\begin{aligned}
\frac{\partial y_s^{(1)}}{\partial \rho_i} &= \frac{\partial T_o}{\partial \rho_i} e_s + T_o \frac{\partial e_s}{\partial \rho_i} \\
&= \left\{ G_o \frac{\partial C}{\partial \rho_i} (1 + G_o C)^{-1} - G_o C (1 + G_o C)^{-2} G_o \frac{\partial C}{\partial \rho_i} \right\} e_s + T_o \frac{\partial e_s}{\partial \rho_i} \\
&= \left\{ \left[\frac{1}{C} \frac{\partial C}{\partial \rho_i} \right] \left[T_o - T_o^2 \right] \right\} e_S - T_o \left(G_{grad} \right)_i T_o e_s \\
&= \left(G_{grad} \right)_i T_o (1 - 2T_o) e_s \\
&= \left(G_{grad} \right)_i T_o e_s - 2 \left(G_{grad} \right)_i T_o y_s^{(1)} \\
&= \left(G_{grad} \right)_i \left[y^{(1)} - 2y^{(2)} \right]
\end{aligned}$$

Similarly for $\frac{\partial u_s^{(1)}}{\partial \rho_i}$ recall that $u_s^{(1)} = S_o C e_s$, thus

$$\begin{aligned}
\frac{\partial u_s^{(1)}}{\partial \rho_i} &= \frac{\partial S_o}{\partial \rho_i} C e_s + S_o \frac{\partial C}{\partial \rho_i} e_S + S_o C \frac{\partial e_s}{\partial \rho_i} \\
&= -(1 + G_o C)^{-2} G_o \frac{\partial C}{\partial \rho_i} C e_s + S_o \frac{\partial C}{\partial \rho_i} e_s - S_o C \left(G_{grad} \right)_i y_s^{(1)} \\
&= -\left(G_{grad} \right)_i T_o S_o C e_s + \left(G_{grad} \right)_i S_o C e_s - S_o C \left(G_{grad} \right)_i y_s^{(1)} \\
&= \left(G_{grad} \right)_i \left[u^{(1)} - 2u^{(2)} \right]
\end{aligned}$$

Clearly, $y^{(2)} = T_o(s) y^{(1)}$ and $u^{(2)} = S_o(s) C(s, \rho) y^{(1)}$, thus proving the lemma. □

Remark: The Hessian is symmetric, therefore only half of its elements need to be calculated.

Notice that the result of the last lemma, implies that there is need for a third experiment in which the input reference is the output of the second experiment. The following algorithm summarises the IFT implementation when calculating the Hessian.

Algorithm 5.1.3. Gradient and Hessian computation, k^{th} step

1. Setup controller using $\rho(k)$

2. Responses

Run closed-loop system with reference input r.

Record signals, e_t and u_t.

Run closed-loop system with reference input e_t

Record signals $y_t^{(1)}$ and $u_t^{(1)}$.

Run closed-loop system with reference input $y_t^{(1)}$

Record signals $u_t^{(2)}$ and $y_t^{(2)}$.

3. Processing responses

Process the recorded signals using,

$$\begin{aligned}
\frac{\partial e_s}{\partial \rho} &= -G_{grad}(s,\rho)y_s^{(1)} \\
\frac{\partial u_s}{\partial \rho} &= G_{grad}(s,\rho)u_s^{(1)} \\
\frac{\partial^2 e_s}{\partial \rho_i \partial \rho_j} &= -\frac{\partial}{\partial \rho_i}\left(G_{grad}\right)_j y_s^{(1)} - \left(G_{grad}\right)_j \frac{\partial y_s^{(1)}}{\partial \rho_i} \\
\frac{\partial^2 u_s}{\partial \rho_i \partial \rho_j} &= \frac{\partial}{\partial \rho_i}\left(G_{grad}\right)_j u_s^{(1)} + \left(G_{grad}\right)_j \frac{\partial u_s^{(1)}}{\partial \rho_i}
\end{aligned}$$

4. Calculate gradient and Hessian

Use time domain formula to compute gradient,

$$\frac{\partial J}{\partial \rho}(k) = \frac{1}{T_f}\int_0^{T_f}\left\{e_t\frac{\partial e_t}{\partial \rho}(k) + \lambda u_t \frac{\partial u_t}{\partial \rho}(k)\right\}dt$$

Use time domain formula to compute Hessian components,

$$\begin{aligned}
\frac{\partial}{\partial \rho_i}\left(\frac{\partial J}{\partial \rho_j}\right) = \frac{1}{T_f}\int_0^{T_f}\Bigg\{&\left[e_t\frac{\partial}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right) + \frac{\partial e_t}{\partial \rho_i}\left(\frac{\partial e_t}{\partial \rho_j}\right)\right] \\
&+\lambda\left[u_t\frac{\partial}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right) + \frac{\partial u_t}{\partial \rho_i}\left(\frac{\partial u_t}{\partial \rho_j}\right)\right]\Bigg\}dt
\end{aligned}$$

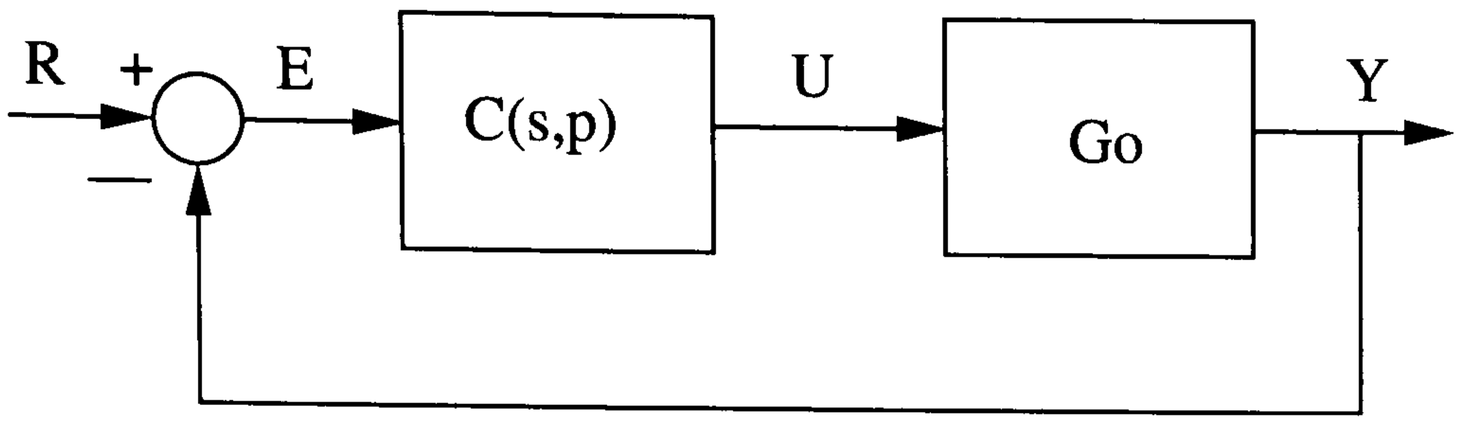

Figure 5.3: MIMO IFT system

5.1.2 IFT formulation for MIMO systems

The formulation presented in section 5.1.1 is applicable to SISO systems only. For the case of MIMO systems, the procedure will follow a similar pattern with some differences. This section presents the development of the deterministic formulation of IFT for MIMO continuous systems. As will be demonstrated, the algorithm shows a high similarity to the discrete stochastic case developed by Hjalmarsson (1999), in the way in which the experiments over the plant are conducted.

Let the system in Figure (5.3) be a MIMO square system, which means that the system has same number of inputs and outputs. Then, the system vector signals will have the following dimensions: $y_t(\rho) \in \mathbb{R}^l$, $e_t(\rho) \in \mathbb{R}^l$, $u_t(\rho) \in \mathbb{R}^m$, and the systems: $G_o \in \mathbb{R}^{l\times m}$, and $C(s,\rho) \in \mathbb{R}^{m\times l}$; so $G_oC(s,\rho) \in \mathbb{R}^{l\times l}$.

The Laplace transforms of the vector signals will have the same dimensions as their time counterparts. Additionally the parameter vector will be of the following dimension $\rho \in \mathbb{R}^{n_p}$.

Lemma 5.1.6. Closed-loop relationships.

Given the previous definitions for the system depicted in Figure (5.3), the following relations between the system signals hold:

$$Y_S(\rho) = S(\rho)G_oC(s,\rho)R_s \tag{5.29}$$

$$E_s(\rho) = S(\rho)R_s \tag{5.30}$$

$$U_s(\rho) = C(s,\rho)S(\rho)R_s \tag{5.31}$$

where,

$$S(\rho) \quad = \quad [I + G_o C(s,\rho)]^{-1} \tag{5.32}$$

Proof. From the diagram the vector of outputs ($Y_s(\rho)$) can be calculated from the vector of errors ($E_s(\rho)$) as in equation (5.33), and the error signal is defined as in equation (5.34).

$$Y_S(\rho) \quad = \quad G_o C(s,\rho) E_s(\rho) \tag{5.33}$$

$$E_S(\rho) \quad = \quad R_s - Y_s(\rho) \tag{5.34}$$

Therefore, replacing the output signal into the error definition and reorganising gives equation (5.30). Similarly, replacing the error signal into the output signal and reorganising gives equation (5.29). Likewise, the control output can be calculated from the error signal as in equation (5.35), and using equation (5.30) gives equation (5.31).

$$U_s(\rho) \quad = \quad C(s,\rho) E_s(\rho) \tag{5.35}$$

□

5.1.2.1 Cost function and cost gradient

The cost function will be scalar function to be minimised with respect to controller parameter vector ρ. Due to the multivariable structure, the internal signals are now vectors, however the Leibniz differentiation theorem still holds, and the gradient can be found using Lemma 5.1.7. The gradient however, requires the calculation of sub-gradients of the error and control signal. Even though the procedure seems very similar to the scalar case, the calculation of these sub-gradients will provide different results. Lemma 5.1.8 provides these results and is considered to be the core to the implemen-

tation of the algorithm.

Lemma 5.1.7. The gradient of the scalar cost functional of equation (5.36) can be calculated with respect to ρ_i using equation (5.37), where ρ_i is the i^{th} controller parameter.

$$J(\rho) = \frac{1}{2T_f}\int_0^{T_f}\left\{e_t^T e_t + \lambda u_t^T u_t\right\}dt \tag{5.36}$$

$$\frac{\partial J(\rho)}{\partial \rho_i} = \frac{1}{T_f}\int_0^{T_f}\left\{e_t^T\frac{\partial e_t}{\partial \rho_i} + \lambda u_t^T\frac{\partial u_t}{\partial \rho_i}\right\}dt \tag{5.37}$$

Proof. Applying Leibniz theorem to equation (5.36) gives,

$$\frac{\partial J(\rho)}{\partial \rho_i} = \frac{1}{2T_f}\int_0^{T_f}\left\{\frac{\partial}{\partial \rho_i}\left[e_t^T e_t\right] + \lambda\frac{\partial}{\partial \rho_i}\left[u_t^T u_t\right]\right\}dt \tag{5.38}$$

For the partial differentiation of the multiplication of two vectors the following property holds: $\frac{\partial}{\partial x}\left[x^T Q x\right] = 2x^T Q$. Thus using the last property and the chain rule gives equation (5.37). □

Lemma 5.1.8. The sub-gradients for e_t and u_t with respect to controller parameter ρ_i are calculated in the Laplace domain, and are equal to:

$$\frac{\partial E_s(\rho)}{\partial \rho_i} = -S(\rho)G_o\tilde{E}_i \tag{5.39}$$

$$\frac{\partial U_s(\rho)}{\partial \rho_i} = \tilde{E}_i - C(\rho)S(\rho)G_o\tilde{E}_i \tag{5.40}$$

where,

$$\tilde{E}_i = \frac{\partial C(\rho)}{\partial \rho_i}E_s(\rho) \tag{5.41}$$

Proof. The partial derivative of $e_t(\rho)$ with respect to any controller parameter ρ_i can be calculated in the Laplace domain using equation (5.30) as follows,

$$\frac{\partial E_s(\rho)}{\partial \rho_i} = \frac{\partial}{\partial \rho_i}\{S(\rho)R_s\} \tag{5.42}$$

$$= \frac{\partial}{\partial \rho_i}\left\{[I + G_o C(\rho)]^{-1} R_s\right\} \tag{5.43}$$

The derivative of the inverse of a matrix with respect to a scalar parameter is calculated using the following property,

$$\frac{\partial}{\partial x}\left[Y^{-1}(x)\right] = -Y^{-1}(x)\frac{\partial Y(x)}{\partial x}Y^{-1}(x) \tag{5.44}$$

Thus, differentiating the inner part of equation (5.43) gives,

$$\frac{\partial E_s(\rho)}{\partial \rho_i} = -S(\rho)G_o\frac{\partial C(\rho)}{\partial \rho_i}S(\rho)R_s \tag{5.45}$$

Then, using equation (5.30) again and defining $\tilde{E}_i$ as in equation (5.41) gives equation (5.39). Similarly, to obtain the partial derivative of $U(s,\rho)$,

$$\frac{\partial U_s(\rho)}{\partial \rho_i} = \frac{\partial}{\partial \rho_i}\{C(\rho)S(\rho)R_s\} \tag{5.46}$$

$$= \frac{\partial C(\rho)}{\partial \rho_i}S(\rho)R_s - C(\rho)S(\rho)G_o \cdot \tag{5.47}$$

$$\frac{\partial C(\rho)}{\partial \rho_i}S(\rho)R_s$$

$$= \tilde{E}_i - C(\rho)S(\rho)G_o\tilde{E}_i \tag{5.48}$$

□

5.1.2.2 Signal recording

The problem now is how to calculate the sub-gradients using the system closed-loop signals. In the SISO case, the sub-gradients were calculated by using the error signal, of a step response experiment, as new reference for the system. Notice however, that in

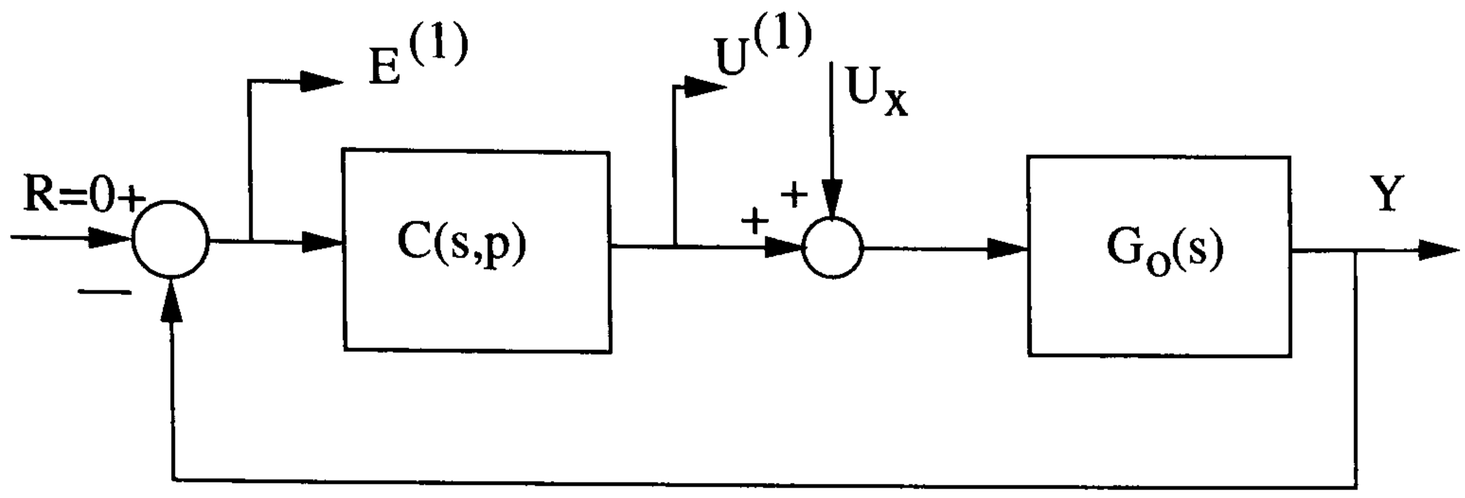

Figure 5.4: MIMO IFT signal recording.

the multivariable case (equations (5.39) and (5.40)), the sub-gradients are a function of a known signal ($\tilde{E}_i$) passing through a particular system filter. The following lemma, describes how to calculate the sub-gradients by performing a new experiment as in the SISO case, but in a different manner.

Lemma 5.1.9. Considering the system configuration in Figure (5.4), the partial derivatives of $E_s(\rho)$ and $U_s(\rho)$ with respect to a parameter ρ_i are equal to,

$$\frac{\partial E_s(\rho)}{\partial \rho_i} = E_i^{(1)} \tag{5.49}$$

$$\frac{\partial U_s(\rho)}{\partial \rho_i} = \tilde{E}_i + U_i^{(1)} \tag{5.50}$$

Proof. From the block diagram in Figure (5.4), the error signal $E_s^{(1)}$ is equal to,

$$E_s^{(1)} = 0 - G_o\left[U_x + C(\rho)E_s^{(1)}\right] \tag{5.51}$$

$$= -\left[I + G_oC(\rho)\right]^{-1} G_oU_x \tag{5.52}$$

$$= -S(\rho)G_oU_x \tag{5.53}$$

and,

$$U_s^{(1)} = C(\rho)E_s^{(1)} \tag{5.54}$$

$$= -C(\rho)S(\rho)G_oU_x \tag{5.55}$$

If $U_x = \tilde{E}_i$, then equations (5.53) and (5.39) are equivalent, thus proving equation

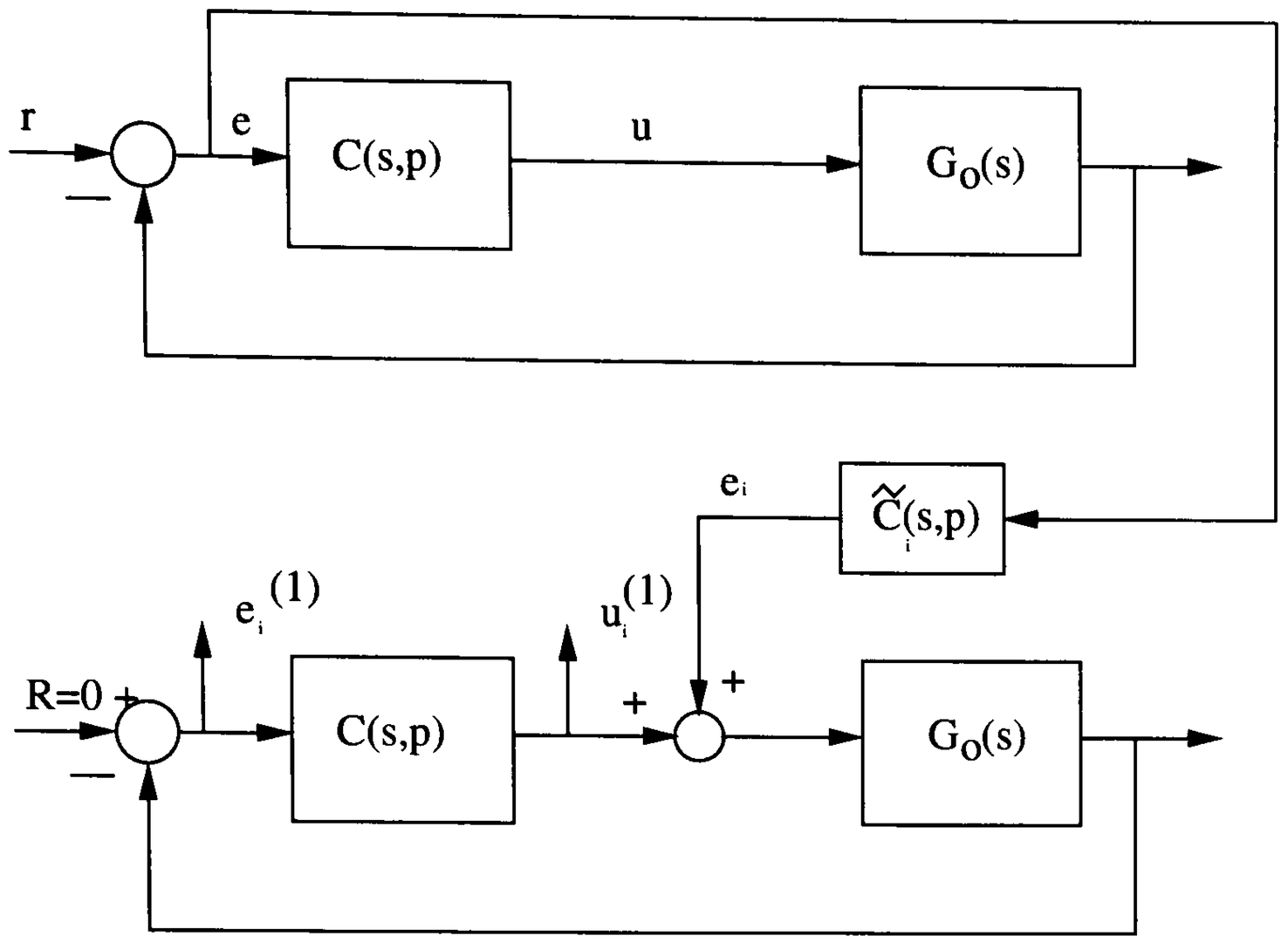

Figure 5.5: MIMO IFT simulation implementation.

(5.49). Similarly, using equation (5.55) in equation (5.40) proves equation (5.50). □

5.1.2.3 Algorithm implementation

Finally, the tuning algorithm follows the same numerical implementation as in the SISO case (Algorithm 5.1.1), but the gradient computation is replaced with the following algorithm for which an equivalent simulation diagram is presented in Figure (5.5).

Algorithm 5.1.4. Gradient computation, k^{th} step

1. Setup controller using $\rho(k)$

2. Responses

   Run closed-loop system with reference input r.

   Record signal e_t.

   Calculate $\tilde{E}_i = \frac{\partial C(s,\rho)}{\partial \rho_i} E(s,\rho)$

   Run closed-loop system with reference input $r = 0$ and $u_x = \tilde{e}_i$

   Record signals $e_i^{(1)}$ and $u_i^{(1)}$.

3. Process the recorded signals using,

$$\begin{aligned}\frac{\partial E(s,\rho)}{\partial \rho_i} &= E_i^{(1)}(s)\\ \frac{\partial U(s,\rho)}{\partial \rho_i} &= \tilde{E}_i(s)+U_i^{(1)}(s)\end{aligned}$$

4. Use time domain formula to compute gradient,

$$\frac{\partial J}{\partial \rho_i}(k)=\frac{1}{T_f}\int_0^{T_f}\left\{e_t^T\frac{\partial e_t}{\partial \rho_i}(k)+\lambda u_t^T\frac{\partial u_t}{\partial \rho_i}(k)\right\}dt$$

5.1.3 Simulation case studies

This section presents three case studies of PI tuning in an activated sludge wastewater treatment plant. The COST Simulation Benchmark (Copp, 2002) is used to illustrate the use of the IFT algorithm. The first two case studies consider that the DO loop in the last aerobic reactor is to be tuned, assuming that the plant will be receive a constant influent flow with constant component concentrations. The DO loop in reactor 5 is depicted in Figure (3.1). The last example considers the tuning of the fourth and the fifth reactor employing the multivariable formulation of the algorithm.

5.1.3.1 Simulation case study No.1

Consider that the PI controller driving the aeration in the last aerobic reactor of the COST simulation benchmark is to be tuned. Table (5.1) summarises the algorithm set up.

The simulations were performed by using a step change in the setpoint input from 2 (mg/l) to 3 (mg/l). Two values of λ have been used to demonstrate its effect in the algorithm. The step size was kept constant, however it might be possible to use a line search to update this parameter at each step; however, since the system is non-linear,

the optimisation problem might not be convex and therefore the optimisation algorithm might not always reach the global minimum.

Figures (5.6) and (5.7) show the step responses of the system for the various settings, and Table (5.2) shows the obtained parameters. Figure (5.8) and (5.9) show contour level plots of the PI controller parameters approach to the minimum of the cost function.

Table 5.1: IFT algorithm and simulation setup

Parameter	Value	Notes
Time period, T_f	1	Achieve steady state
convergence tol. ε	machine accuracy	
initial controller parameter vector, $\rho(0)$	$\begin{bmatrix} 1.00025 & 50.0125 \end{bmatrix}^T$	
R	I	steepest descent
λ	10^{-3} 10^{-5}	
γ_k	Fixed size	

Table 5.2: IFT results

λ	Iterations	K_p	K_i
10^{-5}	3	4.5580	150.0686
10^{-3}	2	4.8148	50.21813

5.1.3.2 Simulation case study No.2

Consider the same problem as in the previous case; however in this case the Hessian will be calculated and use in the parameter update step. Table (5.3) resumes the algorithm set-up and Table (5.4) the obtained results. Finally, Figure (5.10) presents the simulation results.

The use of the Hessian requires that the step size γ and the weighting λ be reduced and augmented respectively. Also, the effect of the change of weight is considerable in the obtained responses. Thus, the use of the Hessian in the parameter makes the algorithm more sensitive, and therefore special care should be taken in their calibration.

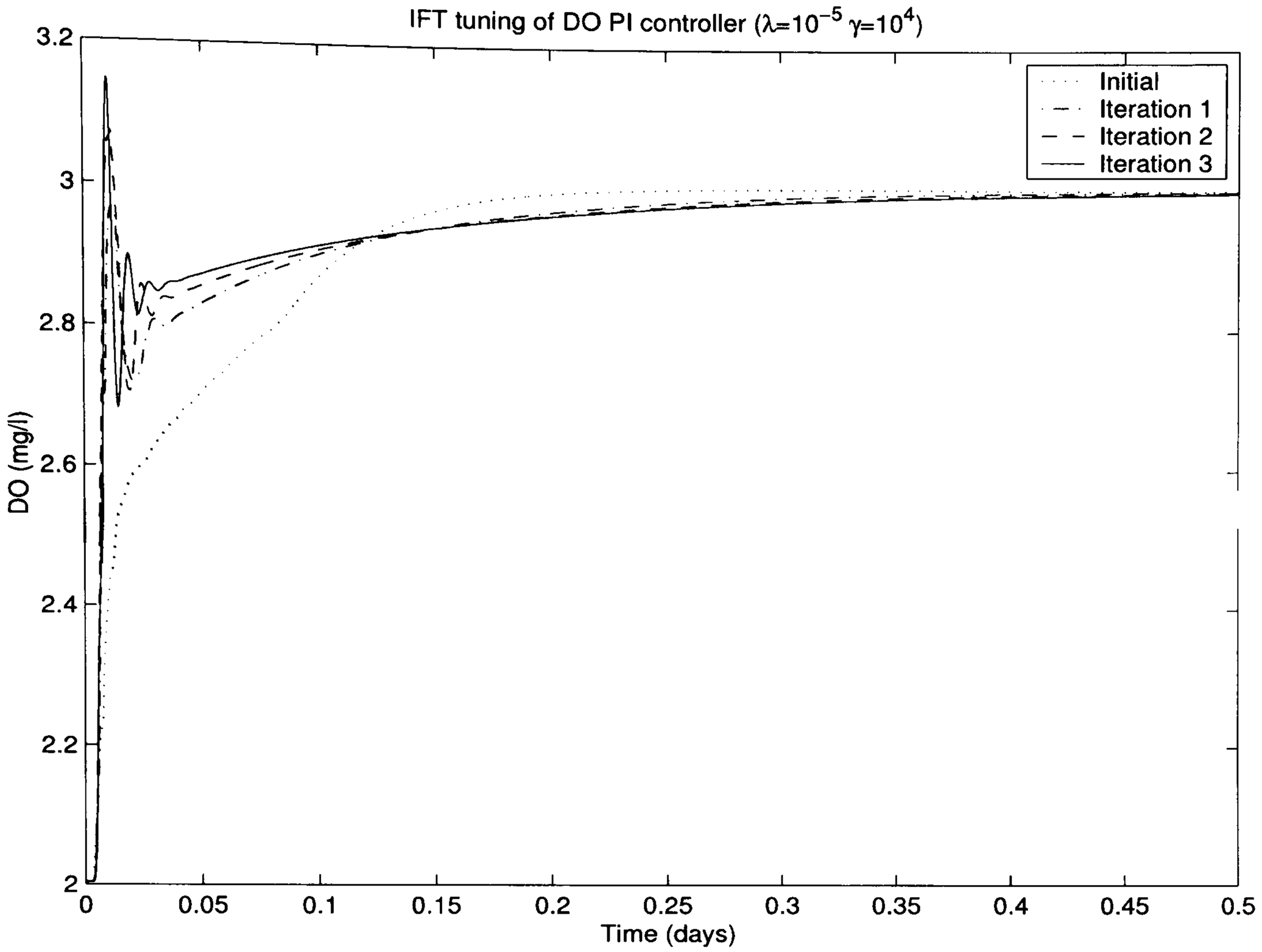

Figure 5.6: Step responses for $\lambda = 10^{-5}$.

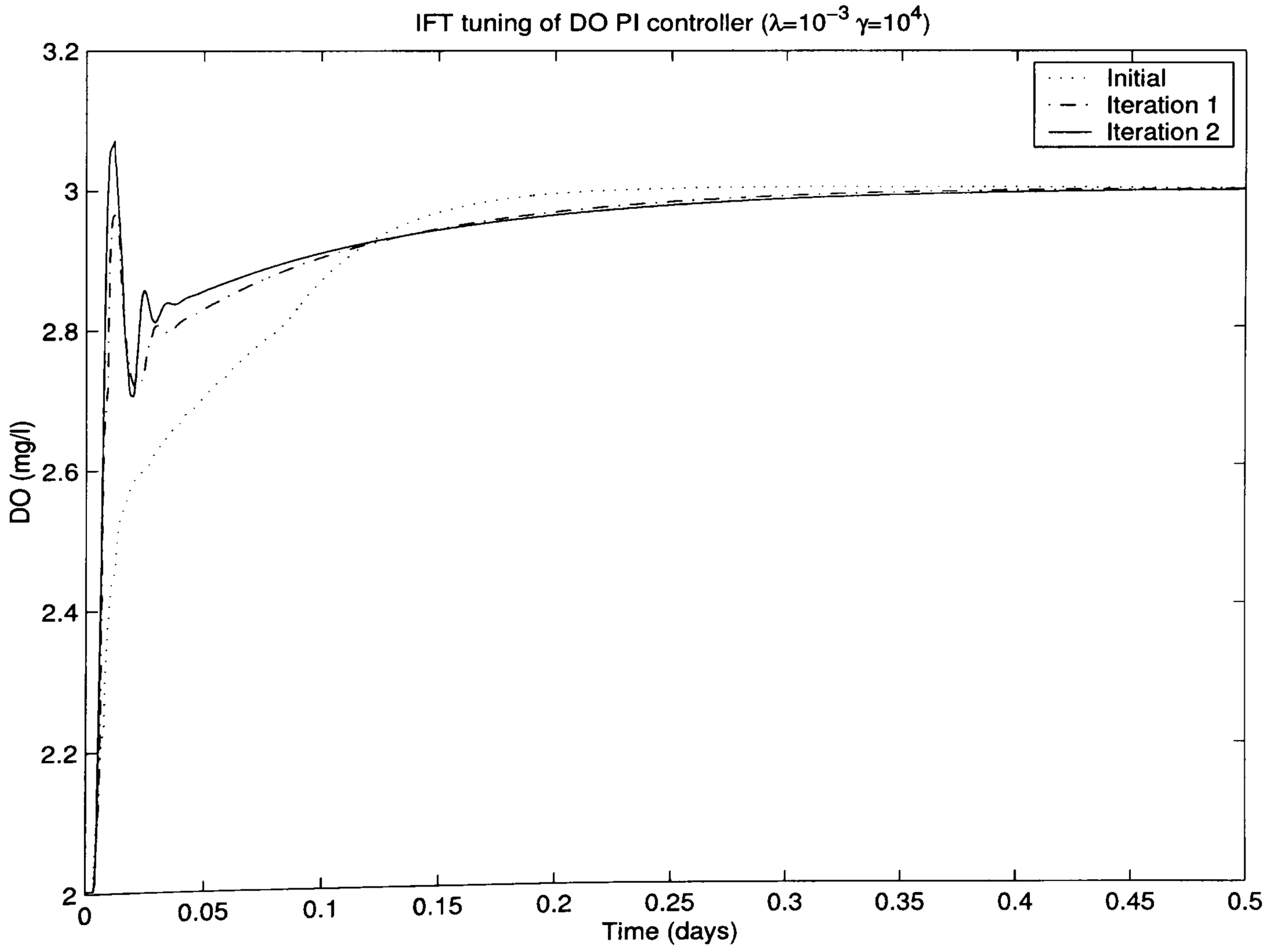

Figure 5.7: Step responses for $\lambda = 10^{-3}$.

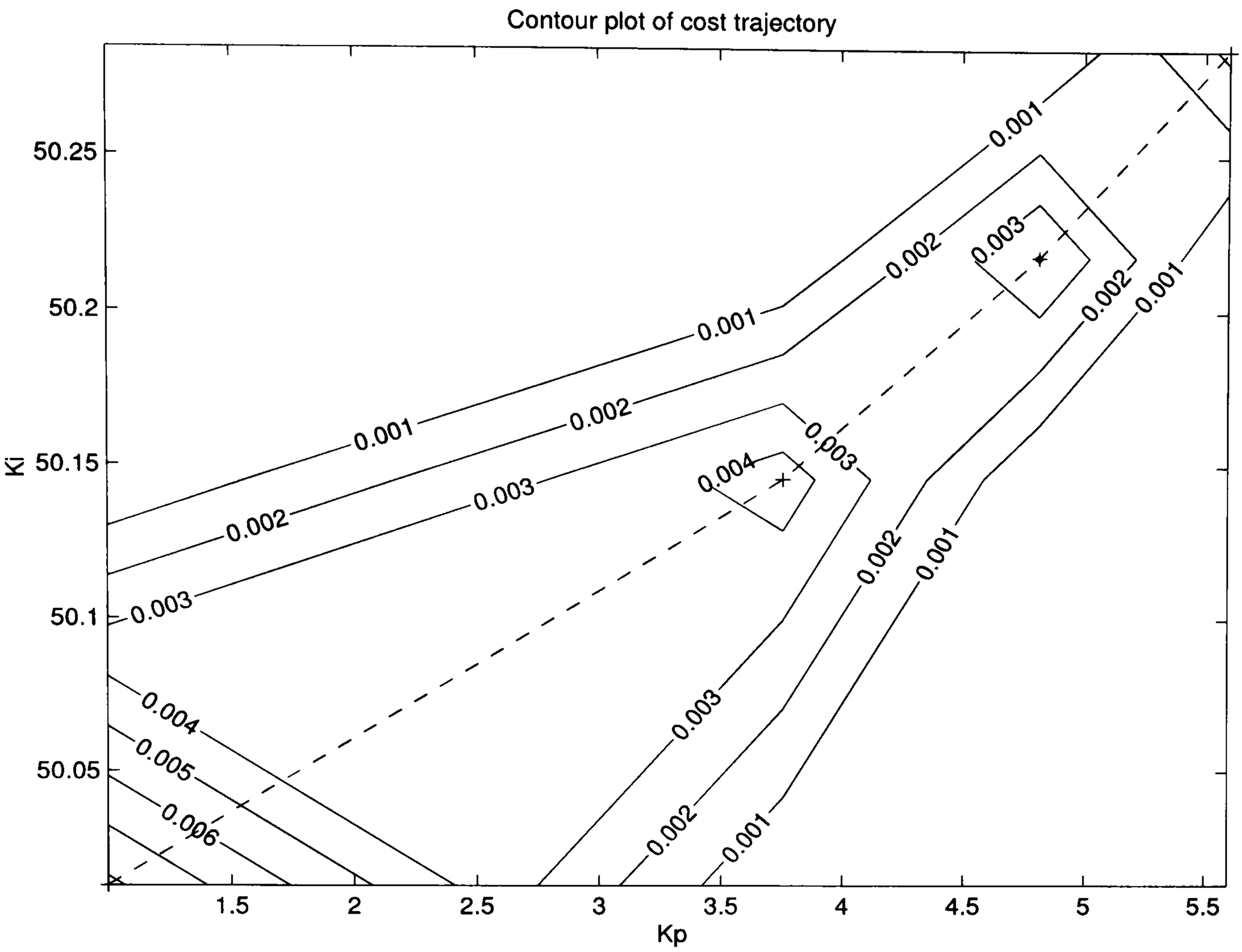

Figure 5.8: Contour level plot for $\lambda = 10^{-5}$.

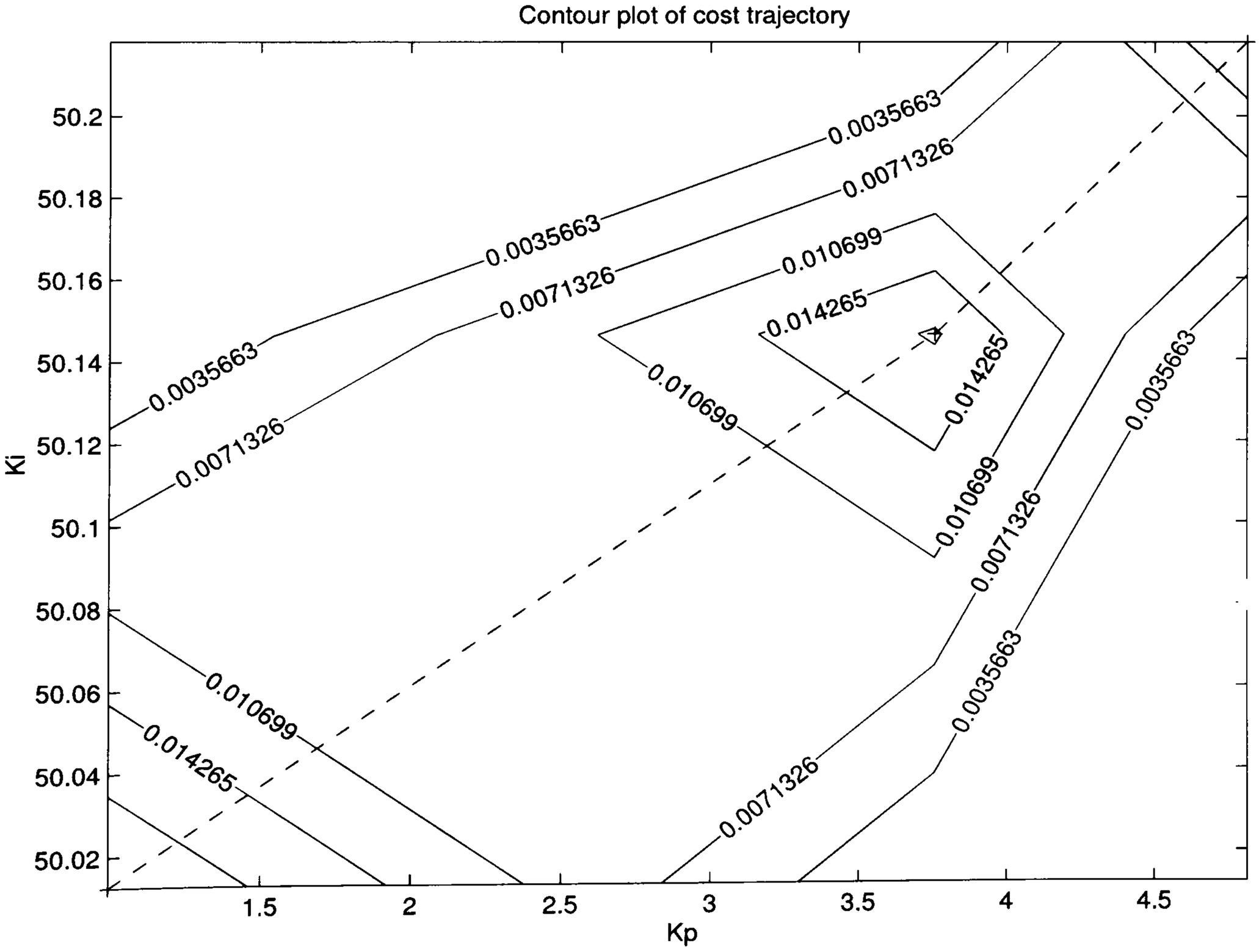

Figure 5.9: Contour level plot for $\lambda = 10^{-3}$.

Table 5.3: IFT algorithm and simulation setup

Parameter	Value	Notes
Time period, T_f	1	Achieve steady state
convergence tol. ε	machine accuracy	
initial controller parameter vector, $\rho(0)$	$\begin{bmatrix} 1.00025 & 50.0125 \end{bmatrix}^T$	
R	H^{-1}	Newton iteration
λ	1 5	
γ_k	Fixed size = 1	

Table 5.4: IFT results

λ	Iterations	K_p	K_i
1	3	10.55	52.38
5	3	5.18	50.80

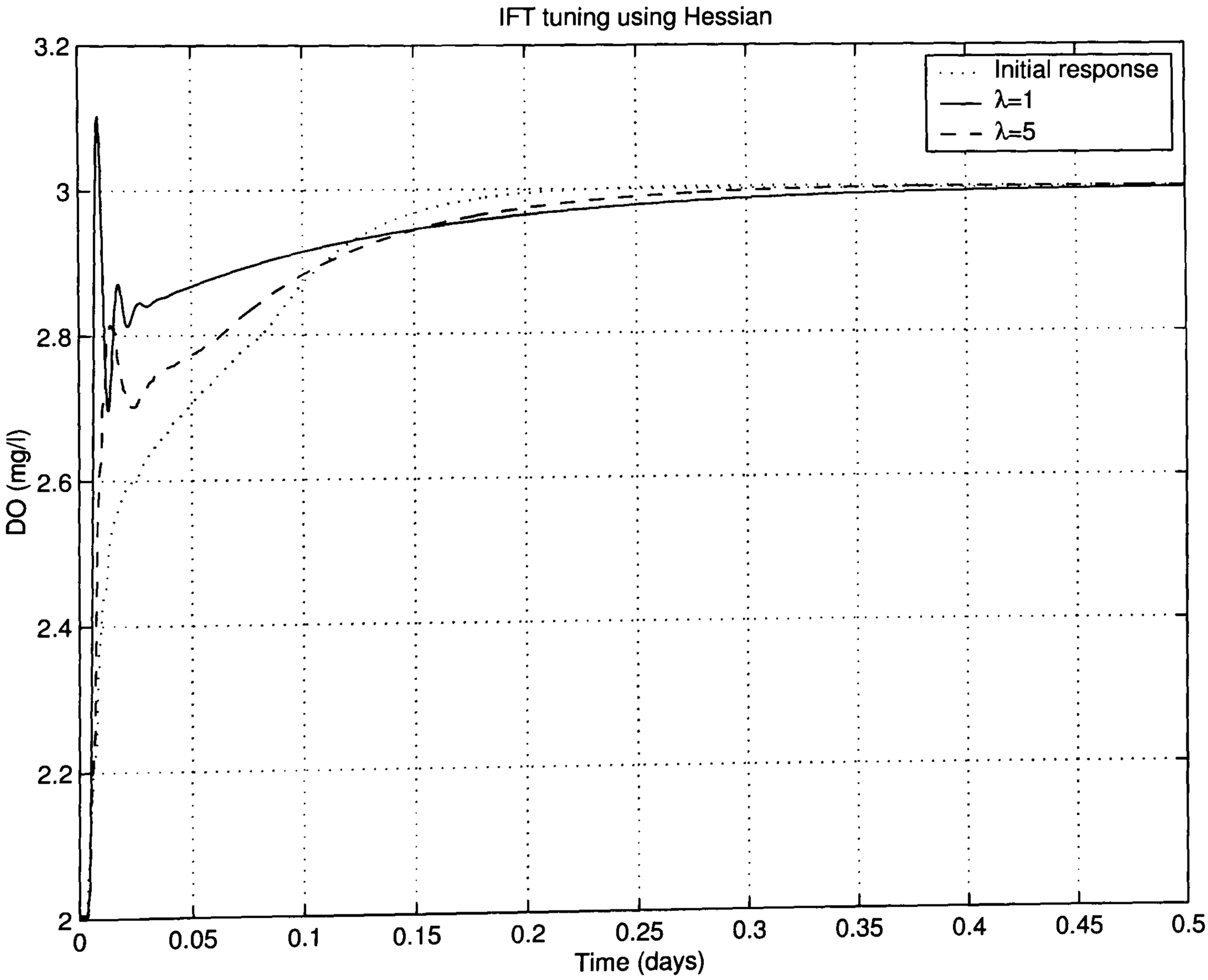

Figure 5.10: Step responses when controller tuned using a Hessian update.

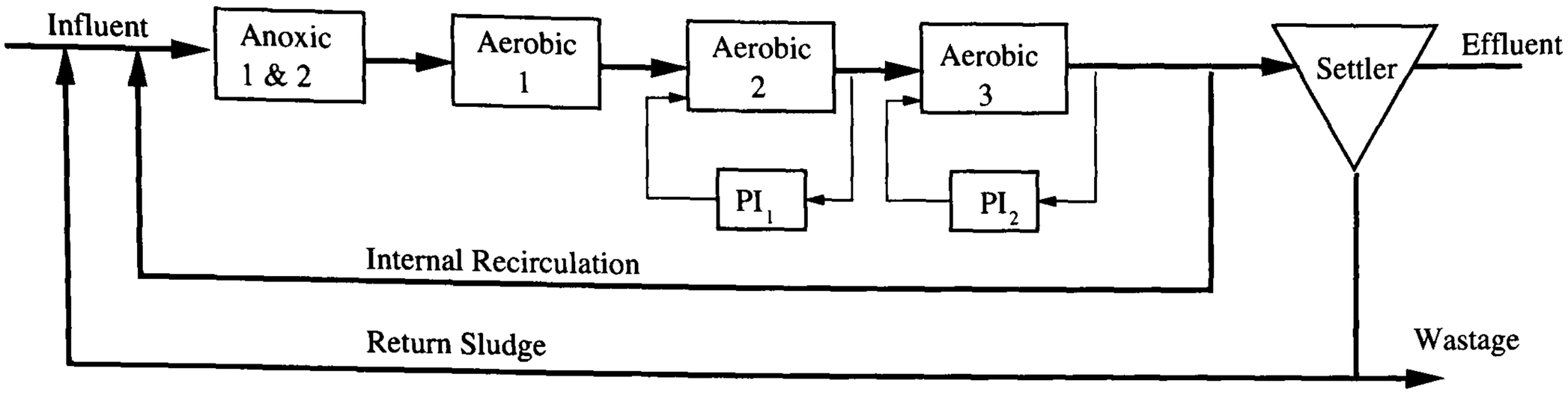

Figure 5.11: Process setup for MIMO IFT.

5.1.3.3 Simulation case study No.3

The problem is now to tune two PI controllers driving the 4^{th} and 5^{th} aeration tanks in the COST WWTP simulation benchmark, as presented in Figure (5.11). Thus, the multivariable controller structure is diagonal, and is presented in equation (5.56). The parameter vector is defined in equation (5.57).

$$C(s,\rho) = \begin{bmatrix} k_{p_1} + \frac{k_{i_1}}{s} & 0 \\ 0 & k_{p_2} + \frac{k_{i_2}}{s} \end{bmatrix} \tag{5.56}$$

$$\rho = \begin{bmatrix} k_{p_1} \\ k_{i_1} \\ k_{p_2} \\ k_{i_2} \end{bmatrix} \tag{5.57}$$

Table (5.5) summarises the problem setup and Table (5.6) the obtained results. Finally, Figures (5.12) and (5.13) present the simulation results.

From the figures, several conclusions can be drawn. First, for example, the effect of the change in λ is to make the controller more aggressive. In Figure (5.12), which has the smallest λ value, the response of the tuned PI controller is much more aggressive in the sense of producing an overshoot for the DO in the 4^{th} reactor and making the response in the 5^{th} reactor faster, compared to the responses in Figure (5.13).

Notice as well, the disparity of the response in the 5^{th} reactor compared with the one of the 4^{th} reactor. Apparently, the optimisation would benefit from different weights

for both reactors. However, with the current formulation of the optimisation problem, it is not possible to produce this effect.

If the responses are compared with the obtained in cases 1 and 2, it appears that the MIMO optimisation problem provides better results in sense of being less aggressive. This effect could be attributed to the couplings between the two tanks, by the knowledge of the system during the optimisation. Therefore, the multivariable case performs better than the SISO case.

Table 5.5: IFT algorithm and simulation setup

Parameter	Value	Notes
Time period, T_f	1	Achieve steady state
convergence tol. ε	machine accuracy	
initial controller parameter vector, $\rho(0)$	$\begin{bmatrix} 1.00025 & 50.0125 \end{bmatrix}^T$	
R	I	steepest descent
λ	10^{-4} 10^{-5}	
γ_k	Fixed size	

Table 5.6: IFT results

λ	Iterations	K_{p_1}	K_{i_1}	K_{p_2}	K_{i_2}
10^{-4}	5	3.26	50.13	0.70	49.70
10^{-5}	5	5.18	50.24	0.99	50.00

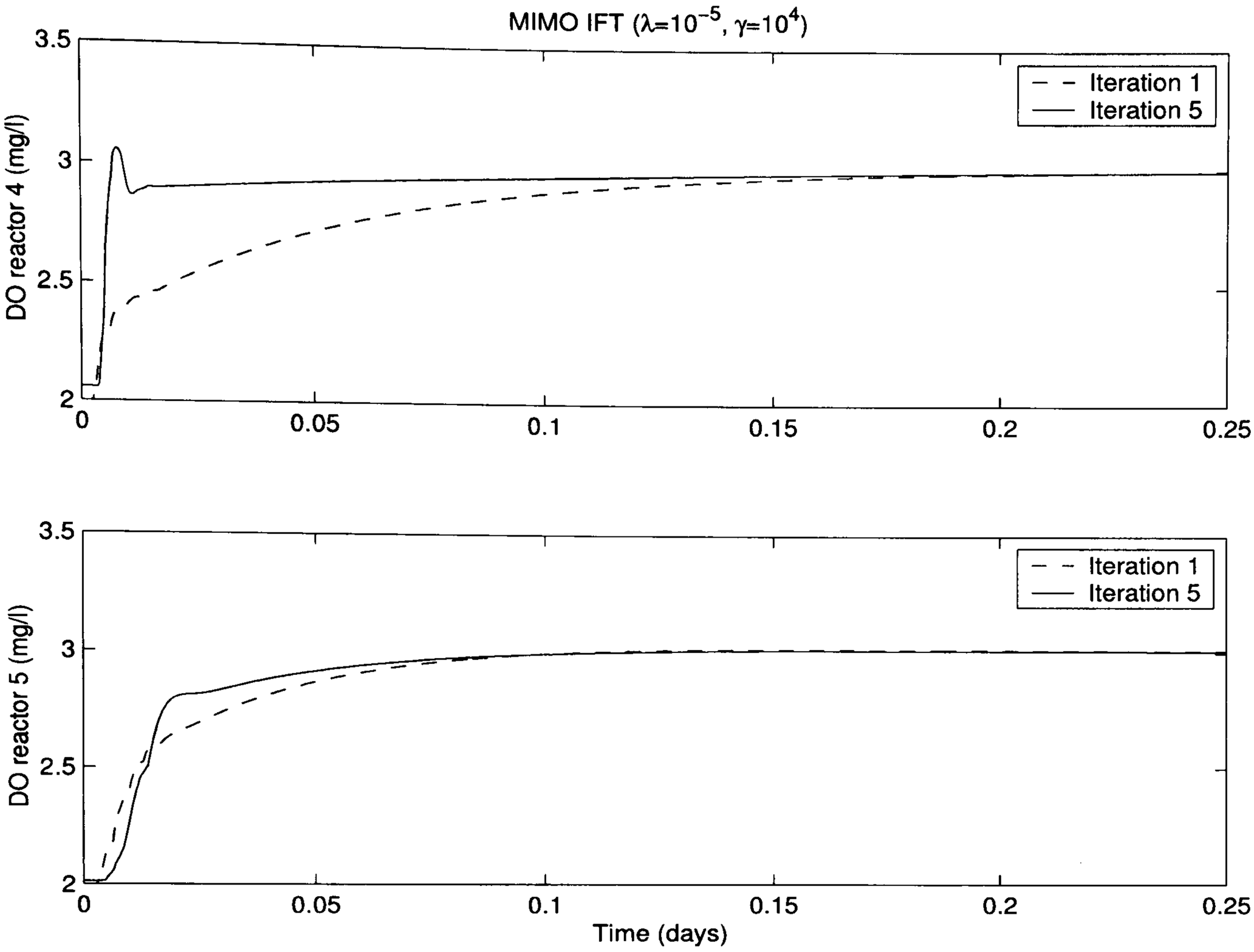

Figure 5.12: Process response $\lambda = 10^{-5}$.

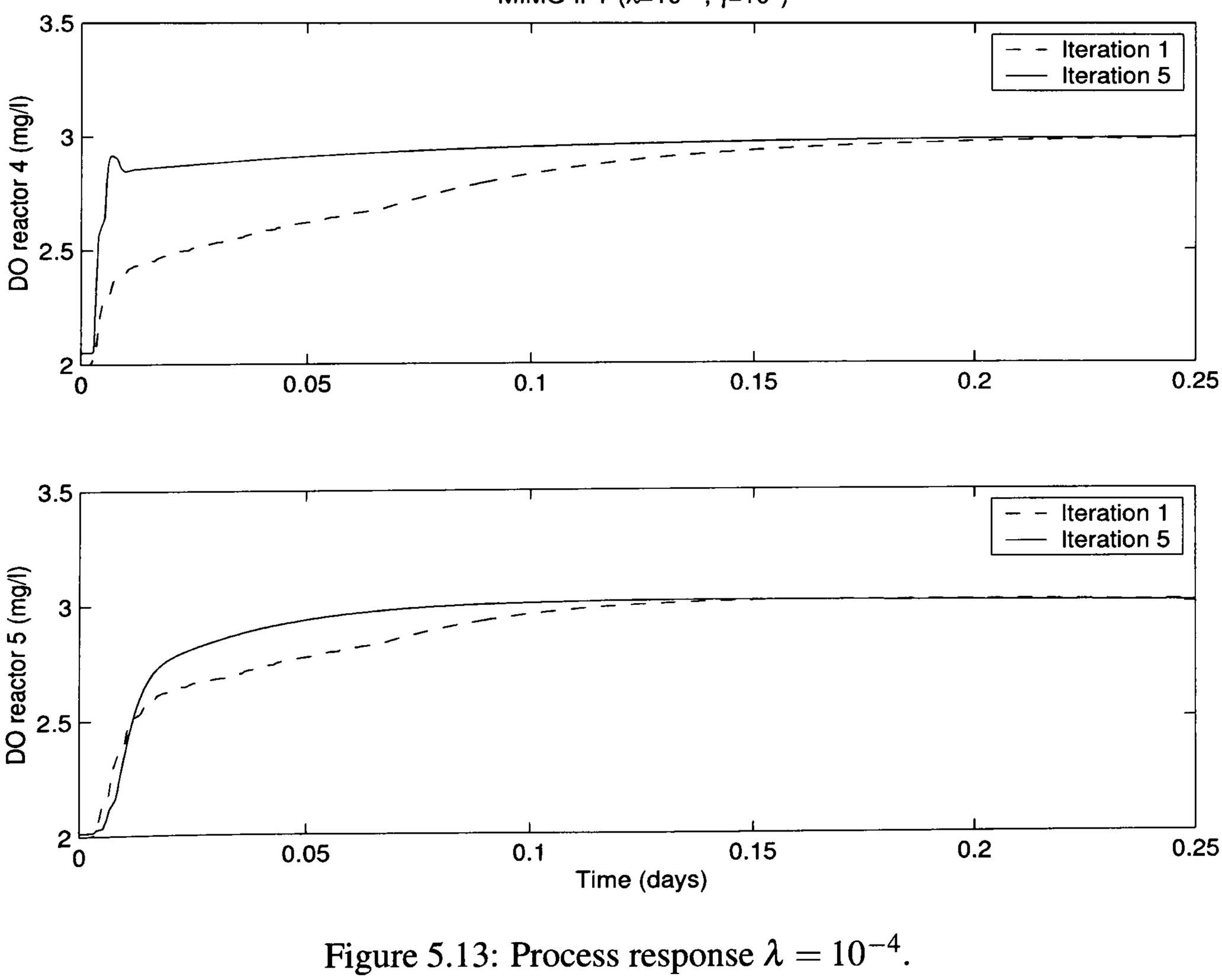

Figure 5.13: Process response $\lambda = 10^{-4}$.

5.2 LQG tuning and process control loop monitoring

The modern treatment of urban wastewater takes places in large-process plant conditions. Control loops are few, but critical, to the success of unit operations like the activated sludge process in the secondary treatment stage. The control problems found in this process usually have noisy measurements, unpredictable disturbances and a corrosive environment leading to sensor degradation, among others. Therefore it is convenient to encourage the development and use of reliable equipment and to support control loops with tuning and monitoring software algorithms able to indicate when controller settings and measurement devices should be examined for possible re-adjustment.

This section introduces a method which allows the calculation (tuning) of PID type controllers for SISO systems. The method employs optimal LQG polynomial theory, nevertheless giving simple explicit formulas. The polynomial approach leads to optimal LQG controllers in transfer function form and therefore can be related to PID structures.

The method has been developed assuming that the process can be described by a simple first order stochastic process model. This simple process model is introduced for two reasons; many key low level processes in wastewater can be described by a first order process description; and a first order model introduces significant simplification into the formulas of the full LQG procedure described in this chapter. In particular, the control law retrieved can be found as PID and explicit formulas can be given for cost function values. The fact that the control law is of PID type is particularly useful since most wastewater control loops are in fact PID-type controllers.

Additionally, to achieve an element of reliability, it might be appropriate to use control loop monitors in conjunction with the controller devices. One possible architecture is to have a loop monitor embedded in the higher levels of the SCADA system running the wastewater treatment works, as in Figure (5.14). However, wastewater treatment

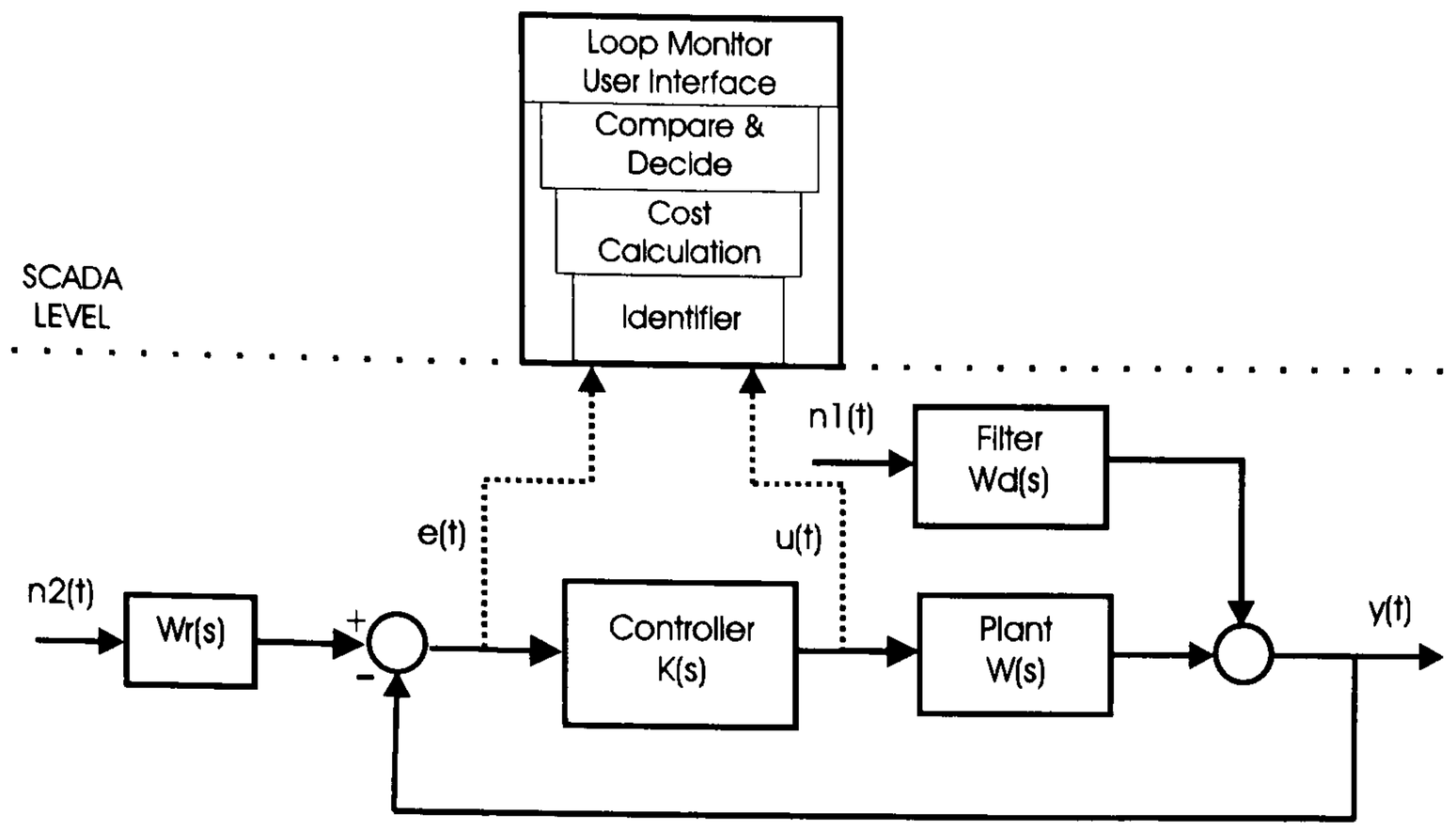

Figure 5.14: The SCADA system control loop monitoring.

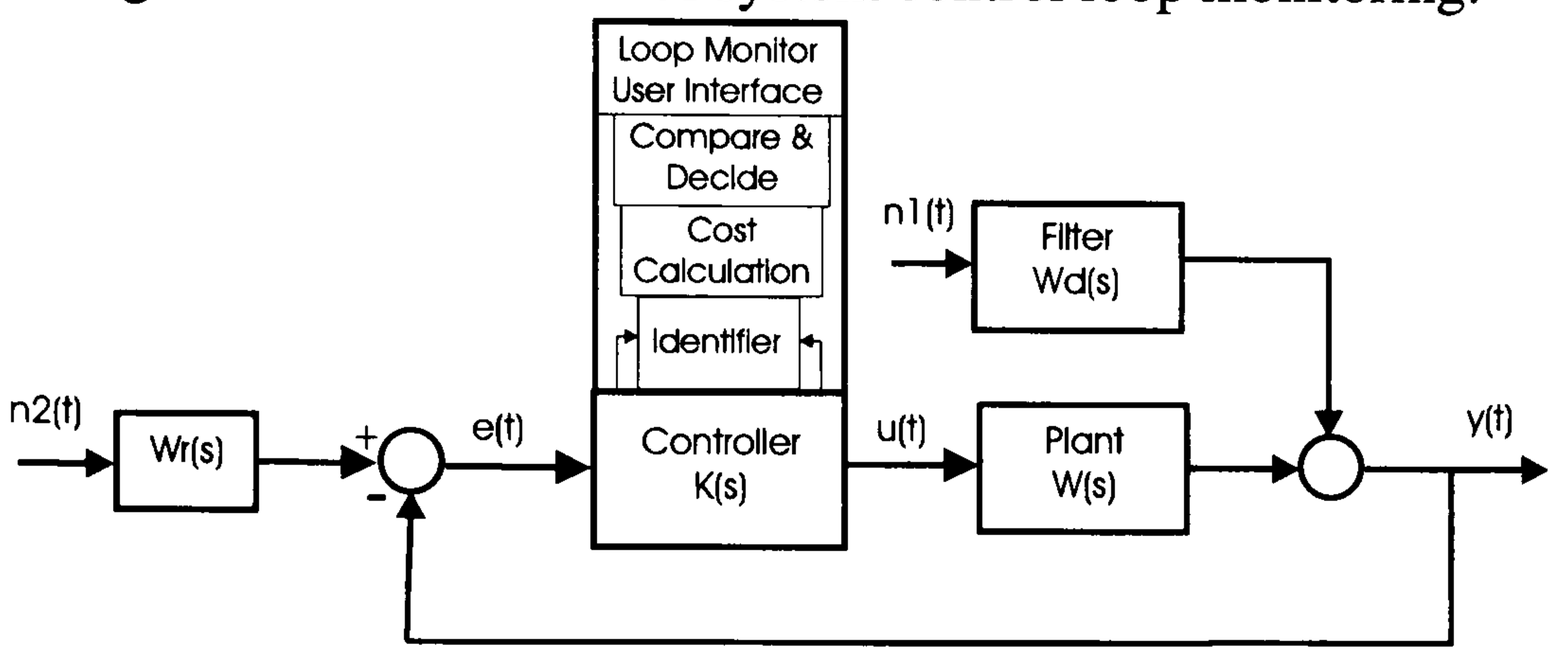

Figure 5.15: The loop-controller-based monitoring.

plant for smaller urban areas and isolated townships tend to be self-contained with PLC based control systems. In this case, an alternative architecture is to have the loop monitor embedded within the local controller hardware level, as in Figure (5.15). It is also possible that new internet and telemetry technology makes the remote monitoring of loops in small treatment plants a possibility, and in this case an extended version of the SCADA architecture of Figure (5.14) could be envisaged. Motivated by the requirement for simple embedded software algorithms to implement a loop monitor that can be used at any level in the process control hierarchy, this thesis reports some simple formulas based on scalar polynomial optimal control theory.

The main content of this section is in subsections 5.2.2 and 5.2.3. The theoretical contribution on LQG optimal control is given in subsection 5.2.3. These theoretical results are captured as some prototype loop monitor algorithms. Implementation is

discussed in subsection 5.2.4 where simulation results are presented.

5.2.1 Benchmarking literature review

One approach to loop controller assessment uses a benchmark index through which actual loop performance is compared with a possible optimised performance benchmark value. Harris and colleagues (Harris, 1989) initiated this conceptual approach using the performance achievable through minimum variance control as the benchmark value. The minimum variance approach has been developed extensively over the last decade or so (Huang and Shah, 1999; Grimble, 2002). The great strength of the minimum variance algorithm is its ability to use online data directly. But, for some time there has been a considerable effort to move away from the minimum variance criterion and use the full Linear Quadratic Gaussian (LQG) optimal cost value as a benchmark index. This study contributes to these research directions.

5.2.2 Process model description

Using the first order system state-space model, the following analysis yields a system configuration for regulation in the presence of a constant (zero) reference input.

$$\dot{x} = -\frac{1}{\tau}x + bu + b_1 n_1 \tag{5.58}$$

$$y = cx \tag{5.59}$$

In transfer function form

$$X(s) = \frac{b}{s+\frac{1}{\tau}}U(s) + \frac{b_1}{s+\frac{1}{\tau}}N_1(s) \tag{5.60}$$

Therefore, by using the output equation, the system model is given as,

$$Y(s) = W(s)U(s) + W_d(s)N_1(s) \tag{5.61}$$

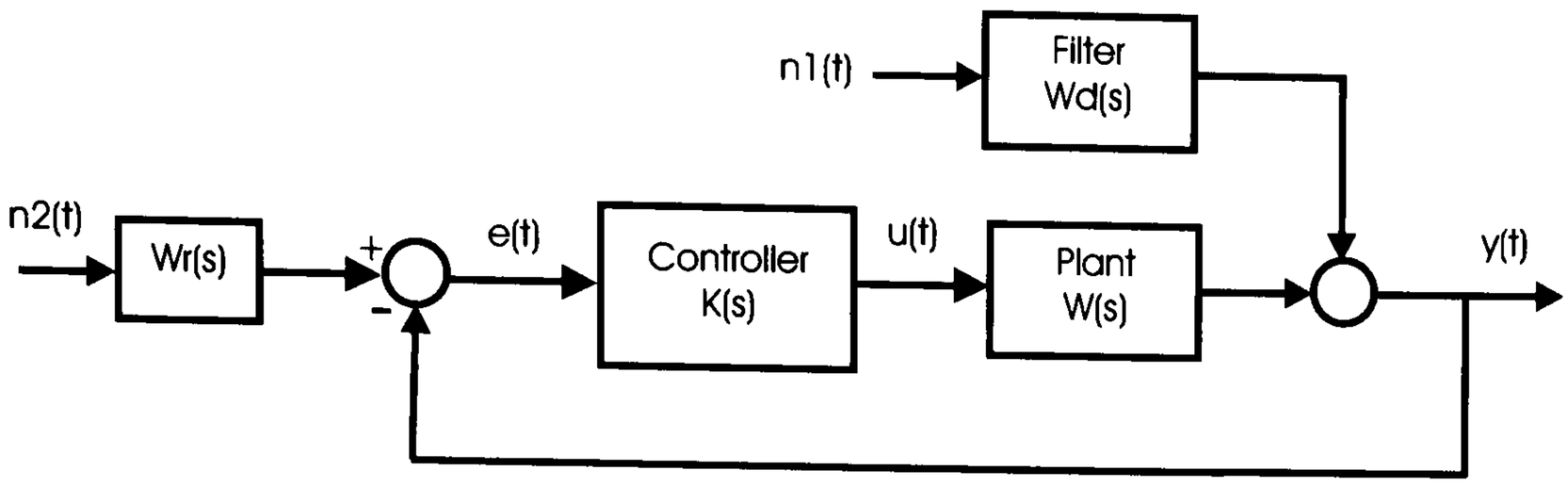

Figure 5.16: LQG system description

where the system transfer functions are given by,

$$W(s) = \frac{K}{\tau s + 1} \tag{5.62}$$

$$W_d(s) = \frac{K_1}{\tau s + 1} \tag{5.63}$$

and the process noise input is $n_1 \in N(0,1)$. The polynomial system structure used in the derivations is shown in Figure (5.16).

5.2.3 LQG optimal control analysis

One finding in wastewater control is that some key low level loops can be modelled by simple common process models, as has been explored in Chapter 3. A second consideration is that most wastewater control loops are PID. The third issue concerns the idea that a simple test is needed to be able to compare actual control loop performance with an optimal benchmark value. Some theoretical results which can be considered as possible solutions for these problems can be derived from the use of a Linear Quadratic Gaussian cost formulation. In this section it is shown that if a scalar polynomial systems formulation for the LQG optimal control problem is assumed for the first order system description, then explicit solutions for the optimal controllers and cost function values can be found. Furthermore the optimal controllers designed all belong to the PID class of controllers. Thus a benchmark value based on optimal PID controllers can be compared against actual industrial PID cost function performance.

This section presents three theorems and a corollary which allow the calculation of the optimal gain for a P controller, PID controller, and a PI controller, as well as the optimal cost. The theorems to be given are derived using the usual steps in the LQG optimal control polynomial system methods (Grimble and Johnson, 1988). These standard steps are:

1. Problem and cost function weight definition

   Cost function:

$$J = \frac{1}{2\pi j}\int_D \{Q_c\Phi_{ee} + R_c\Phi_{uu}\}\,ds \tag{5.64}$$

   Weights:

$$Q_c = \frac{Q_{cn}}{A_q^* A_q} \tag{5.65}$$

$$R_c = \frac{R_{cn}}{A_r^* A_r} \tag{5.66}$$

2. Common denominator model

$$\begin{bmatrix} W(s) & W_d(s) & W_r(s) \end{bmatrix} = A^{-1}\begin{bmatrix} B & C_d & E \end{bmatrix} \tag{5.67}$$

3. Filter spectral factor

$$D_f D_f^* = C_d C_d^* + EE^* \tag{5.68}$$

4. Control spectral factor

$$D_c^* D_c = B^* A_r^* Q_{cn} A_r B + A^* A_q^* R_{cn} A_q A \tag{5.69}$$

5. Diophantine equation solution

$$D_c^* G_o + F_o A A_q = B^* A_r^* Q_{cn} D_f \tag{5.70}$$

$$D_c^* H_o - F_o B A_r = A^* A_q^* R_{cn} D_f \tag{5.71}$$

6. Optimal controller formula and closed-loop relationships

$$K = \left(H_o A_q\right)^{-1} G_o A_r \tag{5.72}$$

$$\rho_{cl}(s) = D_c(s) D_f(s) \tag{5.73}$$

7. Optimal cost function values

$$J_o = J_1 + J_2 \tag{5.74}$$

where

$$J_1 = \frac{1}{2\pi j} \int_D \left\{ \frac{F_o^* F_o}{D_c^* D_c} \right\} ds \tag{5.75}$$

$$J_2 = \frac{1}{2\pi j} \int_D \left\{ \frac{D_f D_f^*}{D_c^* D_c} Q_{cn} R_{cn} \right\} ds \tag{5.76}$$

8. Non-optimal cost function values

$$J = J_{ee} + J_{uu} \tag{5.77}$$

where

$$J_{ee} = \frac{1}{2\pi j} \int_D Q_c T_e T_e^* ds \tag{5.78}$$

$$J_{uu} = \frac{1}{2\pi j} \int_D R_c T_u T_u^* ds \tag{5.79}$$

and

$$T_e = K_d(AK_d + BK_n)^{-1}D_f \tag{5.80}$$

$$T_u = K_n(AK_d + BK_n)^{-1}D_f \tag{5.81}$$

Common denominator process description

A regulation, not a reference tracking, formulation is assumed. Thus benchmarking will occur when the system is in steady conditions and not changing over to a new reference level, hence set, $W_r(s) = 0$. The common denominator system form for the simple system assumptions used in this study is,

$$\begin{bmatrix} W(s) & W_d(s) & W_r(s) \end{bmatrix} = (\tau s + 1)^{-1}\begin{bmatrix} K & K_1 & 0 \end{bmatrix} \tag{5.82}$$

$$= A^{-1}\begin{bmatrix} B & C_d & E \end{bmatrix}$$

Thus, giving the following equivalences

$$A = (\tau s + 1) \tag{5.83}$$

$$B = K \tag{5.84}$$

$$C_d = K_1 \tag{5.85}$$

$$E = 0 \tag{5.86}$$

Theorem 5.2.1. Optimal P controller

Given the following LQG cost function $J = \frac{1}{2\pi j}\int_D \{Q_c\Phi_{ee} + R_c\Phi_{uu}\}ds$, where $Q_c = \frac{Q_{cn}}{A_q^* A_q}$ and $R_c = \frac{R_{cn}}{A_r^* A_r}$; with $Q_{cn} = 1$, $R_{cn} = \rho^2$ and $A_q = A_r = 1$, the optimal control gain can be calculated as,

$$K_{opt} = \frac{\alpha - 1}{K} \tag{5.87}$$

where,

$$\alpha^2 = 1 + \frac{K^2}{\rho^2} \tag{5.88}$$

and the optimal cost function value, J_o, is given by,

$$J_o = \frac{\rho^2 K_1^2 (\alpha - 1)^2}{2\alpha K^2} + \frac{K_1^2}{2\alpha} \tag{5.89}$$

Proof. Given the system defined in common denominator form, and setting $Q_c = 1$ and $R_c = 1$, therefore giving $Q_{cn} = R_{cn} = A_q = A_r = 1$. Replacing the respective values in the filter polynomial spectral factor of equation (5.68), gives:

$$D_f = K_1 \tag{5.90}$$

Replacing the equivalent values into the control spectral factor of equation (5.69) gives:

$$D_c^* D_c = \rho^2 \left(-\tau^2 s^2 + 1 + \frac{K^2}{\rho^2} \right) \tag{5.91}$$

Therefore giving,

$$D_c = \rho (\tau s + \alpha) \tag{5.92}$$

where,

$$\alpha = \sqrt{1 + \frac{K^2}{\rho^2}} \tag{5.93}$$

Evaluating the Diophantine equations (5.70) and (5.71) gives equations (5.94) and

(5.95).

$$\rho(-\tau s+\alpha)g_o+f_o(-\tau s+1) = KK_1 \tag{5.94}$$

$$\rho(-\tau s+\alpha)h_o-f_oK = (-\tau s+1)\rho^2K_1 \tag{5.95}$$

The three unknown factors in these equations (g_o, f_o and h_o) are found by equating each term of each side of both equations, thus giving the overdetermined system of equations (5.96) through (5.99).

$$s^1: \quad -\rho\tau g_o+f_o\tau=0 \tag{5.96}$$

$$s^0: \quad \rho\alpha g_o+f_o=KK_1 \tag{5.97}$$

$$s^1: \quad -\rho\tau h_o=-\tau\rho^2K_1 \tag{5.98}$$

$$s^0: \quad \rho\alpha h_o-f_oK=\rho^2K_1 \tag{5.99}$$

Using equations (5.96), (5.98) and (5.99) yields the following solution to the system:

$$h_o = \rho K_1 \tag{5.100}$$

$$f_o = \frac{K_1(\alpha-1)}{K}\rho^2 \tag{5.101}$$

$$g_o = \frac{K_1(\alpha-1)}{K}\rho \tag{5.102}$$

The remaining equation (5.97) is evaluated in the solution giving the same result on both sides of the equation, thus proving uniqueness of the solution. The optimal gain is then calculated by evaluating equation (5.72) in the solution of the implied Diophantine system, thus proving the first part of the theorem.

The second part of the theorem consists of obtaining the optimal cost function value. To do so, replace the appropriate values into equations (5.75) and (5.76) and evaluate as follows for J_1:

$$J_1 = \frac{\rho^2 K_1^2(\alpha-1)^2}{2\pi K^2 j}\int_D \frac{ds}{(-\tau s+1)(\tau s+1)} \tag{5.103}$$

expanding the integral into partial fractions and using the residue theorem gives,

$$J_1 = \frac{\rho^2 K_1^2(\alpha-1)^2}{2\alpha K^2} \tag{5.104}$$

Similarly for J_2,

$$J_2 = \frac{K_1^2}{2\alpha} \tag{5.105}$$

Thus, making $J_o = J_1 + J_2$ proves the second part of this theorem. □

Theorem 5.2.2. Optimal PID controller

Given the following LQG cost function $J = \frac{1}{2\pi j}\int_D \{Q_c\Phi_{ee} + R_c\Phi_{uu}\}\,ds$, where $Q_c = \frac{Q_{cn}}{A_q^* A_q}$ and $R_c = \frac{R_{cn}}{A_r^* A_r}$; with $Q_{cn} = 1$, $R_{cn} = \rho^2$, $A_q = s$, and $A_r = (\tau_r s + 1)$; the optimal control gains can be calculated as

$$K_p = \frac{(g_o + g_1\tau_r)}{h_o} \tag{5.106}$$

$$K_i = \frac{g_1}{h_o} \tag{5.107}$$

$$K_d = \frac{g_o\tau_r}{h_o} \tag{5.108}$$

where

$$g_o = \frac{2\tau K_1\rho}{\rho(1+b)+K\tau_r} \tag{5.109}$$

$$g_1 = K_1 \tag{5.110}$$

$$h_o = K_1\rho\left[\frac{\rho(1+b)-K\tau_r}{\rho(1+b)+K\tau_r}\right] \tag{5.111}$$

$$b = \sqrt{1+\frac{2\tau K}{\rho}+\left(\frac{K\tau_r}{\rho}\right)^2} \tag{5.112}$$

Proof. Given the system defined in common denominator form, and setting $Q_c = 1$ and $R_c = 1$, therefore giving $Q_{cn} = R_{cn} = A_q = A_r = 1$. Replacing the respective values in the filter polynomial spectral factor of equation (5.68), gives:

$$D_f = K_1 \tag{5.113}$$

Replacing the equivalent values into the control spectral factor of equation (5.69) gives:

$$D_c^* D_c = K^2 (-\tau_r s + 1)(\tau_r s + 1) + \rho^2 (-s^2)(-\tau s + 1)(-\tau s + 1) \tag{5.114}$$

Reorganising equation (5.114) and grouping terms gives:

$$D_c^* D_c = \rho^2 (\tau s^2 - bs + c)(\tau s^2 + bs + c) \tag{5.115}$$

where

$$b = \sqrt{1 + \frac{2\tau K}{\rho} + \left(\frac{K\tau_r}{\rho}\right)^2} \tag{5.116}$$

$$c = \frac{K}{\rho} \tag{5.117}$$

Evaluating the Diophantine equations (5.70) and (5.71) gives equations (5.118) and (5.119).

$$\rho (\tau s^2 - bs + c) g(s) + f(s)(\tau s + 1) s = K(-\tau_r s + 1) K_1 \tag{5.118}$$

$$\rho (\tau s^2 - bs + c) h(s) - f(s) K (\tau_r s + 1) = (-\tau s + 1)(-s) \rho^2 K_1 \tag{5.119}$$

By selecting the structure $f(s) = f_o s + f_1$, $g(s) = g_o s + g1$ and $h(s) = h_o$, equations (5.118) and (5.119) yield,

$$\rho\left(\tau s^2 - bs + c\right)(g_o s + g_1) + (f_o s + f_1)\left(\tau s^2 + s\right) = KK_1(-\tau_r s + 1) \quad (5.120)$$

$$\rho\left(\tau s^2 - bs + c\right)h_o - K(f_o s + f_1)(\tau_r s + 1) = \rho^2 K_1\left(\tau s^2 - s\right) \quad (5.121)$$

Equating powers of s in equations (5.120) and (5.121) give respectively:

$$s^3: \quad \rho\tau g_o + f_o\tau = 0 \quad (5.122)$$

$$s^2: \quad \rho\tau g_1 - \rho b g_o + f_o + f_1\tau = 0 \quad (5.123)$$

$$s^1: \quad -\rho b g_1 + \rho c g_o + f_1 = -KK_1\tau_r \quad (5.124)$$

$$s^0: \quad \rho c g_1 = KK_1 \quad (5.125)$$

$$s^2: \quad \rho\tau h_o - f_o K\tau_r = \rho^2 K_1\tau \quad (5.126)$$

$$s^1: \quad -\rho b h_o - f_o K - f_1 K\tau_r = -\rho^2 K_1 \quad (5.127)$$

$$s^0: \quad \rho c h_o - K f_1 = 0 \quad (5.128)$$

The resulting system yields 7 equations and 5 unknowns (f_o, f_1, g_o, g_1, and h_o). The solution of this overdetermined system is unique, and gives the following solution:

$$h_o = K_1\rho\left[\frac{\rho(1+b) - K\tau_r}{\rho(1+b) + K\tau_r}\right] \quad (5.129)$$

$$f_o = -\rho g_o \quad (5.130)$$

$$f_1 = h_o \quad (5.131)$$

$$g_o = \frac{2\tau K_1\rho}{\rho(1+b) + K\tau_r} \quad (5.132)$$

$$g_1 = K_1 \quad (5.133)$$

Therefore, the optimal controller gains are:

$$K_{opt}(s) = \frac{G_o A_r}{H_o A_q} \tag{5.134}$$

$$= \frac{(g_o s + g_1)(\tau_r s + 1)}{h_o s} \tag{5.135}$$

Finally giving:

$$K_{opt}(s) = \left(\frac{g_o + g_1 \tau_r}{h_o}\right) + \left(\frac{g_1}{h_o}\right)\frac{1}{s} + \left(\frac{g_o \tau_r}{h_o}\right) s \tag{5.136}$$

Thus, $K_p = \frac{g_o + g_1 \tau_r}{h_o}$, $K_i = \frac{g_1}{h_o}$, and $K_d = \frac{g_o \tau_r}{h_o}$. □

For the case of PI controller, the controller gains can be determined by using the following corollary from the previous theorem.

Corollary 5.2.1. Optimal PI controller

If the parameter τ_r is set to zero, then the results of Theorem 5.2.2 are those for optimal proportional and integral control.

The corollary shows that with $\tau_r = 0$, the optimal PI occurs, and hence increasing the value of τ_r slowly introduces the D term in the controller.

Theorem 5.2.3. Optimal PID control cost

For the optimal controller of theorem 5.2.2, the optimal cost value is given by,

$$J_o = J_1 + J_2 \tag{5.137}$$

with,

$$J_1 = \frac{A_1}{\rho^2} \tag{5.138}$$

$$J_2 = K_1^2 A_2 \tag{5.139}$$

and,

$$A_1 = \frac{cf_o^2 + f_1^2\tau}{2bc\tau} \tag{5.140}$$

$$A_2 = \frac{\tau}{2bc} \tag{5.141}$$

Proof. The closed-loop poles of the PID optimal system are given by,

$$p_{cl}(s) = D_c(s)D_f(s) \tag{5.142}$$

$$= K_1\rho\left(\tau s^2 + bs + c\right) \tag{5.143}$$

where b and c have been defined in the previous theorem in equations (5.117) and (5.116) respectively. Complex poles occur if:

$$\frac{2K\tau}{\rho} - \frac{K^2\tau_r^2}{\rho^2} - 1 > 0 \tag{5.144}$$

To determine the cost function values, it will be necessary to perform a partial function expansion of the kernel prior to the use of the residue theorem to evaluate any of the two integrals in J_1 and J_2. Examining first J_1,

$$J_1 = \frac{1}{2\pi j}\int_D \frac{f^* f}{D_c^* D_c} ds \tag{5.145}$$

Therefore performing a partial fraction expansion over the kernel,

$$\frac{f^* f}{D_c^* D_c} = \frac{(-f_o s + f_1)(f_o s + f_1)}{\rho^2(\tau s^2 - bs + c)(\tau s^2 + bs + c)} \tag{5.146}$$

$$= \left(\frac{1}{\rho^2}\right)\left[\frac{-As + B}{\tau s^2 - bs + c} + \frac{As + B}{\tau s^2 + bs + c}\right] \tag{5.147}$$

Hence,

$$(-f_o s + f_1)(f_o s + f_1) = (-As + B)(\tau s^2 + bs + c) + (As + B)(\tau s^2 - bs + c) \quad (5.148)$$

Equating coefficients of powers of s:

$$s^2: \quad -f_o^2 = 2(-Ab + B\tau) \quad (5.149)$$

$$s^0: \quad f_1^2 = 2Bc \quad (5.150)$$

Hence,

$$B = \frac{f_1^2}{2c} \quad (5.151)$$

$$A = \frac{1}{2b}\left(f_o^2 + \frac{f_1^2 \tau}{c}\right) \quad (5.152)$$

Thus using the notation $A_1 = -A$ and $B_1 = B$ yields,

$$J_1 = \frac{1}{2\pi j}\left(\frac{1}{\rho^2}\right)\int_D \frac{A_1 s + B_1}{(\tau s^2 - bs + c)} ds \quad (5.153)$$

Examining the second term (J_2),

$$J_2 = \frac{1}{2\pi j}\int_D \left\{\frac{D_f D_f^*}{D_c^* D_c} Q_{cn} R_{cn}\right\} ds \quad (5.154)$$

$$= \frac{K_1^2}{2\pi j}\int_D \frac{ds}{(\tau s^2 - bs + c)(\tau s^2 + bs + c)} \quad (5.155)$$

Doing partial fractions over the kernel,

$$\frac{1}{(\tau s^2 - bs + c)(\tau s^2 + bs + c)} = \frac{-As + B}{(\tau s^2 - bs + c)} + \frac{As + B}{(\tau s^2 + bs + c)} \quad (5.156)$$

Equating coefficients of powers of s:

$$s^2: \qquad 0 = 2(-Ab + B\tau) \tag{5.157}$$

$$s^0: \qquad 1 = 2Bc \tag{5.158}$$

Therefore, $B = \frac{1}{2c}$ and $A = \frac{\tau}{2bc}$. Finally, use the notation $A_2 = -A$, and $B_2 = B$, which yields,

$$J_2 = \frac{1}{2\pi j}(K_1^2)\int_D \frac{A_2 s + B_2}{(\tau s^2 - bs + c)} ds \tag{5.159}$$

This concludes the proof of the theorem. □

Loop monitoring

To conclude with this section, the loop monitoring algorithm is presented, with the purpose of providing a summarised guideline for a a real-time implementation.

A benchmark index is created using the optimal cost value. The cost will satisfy the relation, $0 < J_{opt} \leq J_{act}$ and this can be used in the form,

$$0 < BI_{LQG} \leq 1 \tag{5.160}$$

with benchmark index $BI_{LQG} = J_{opt}/J_{act}$. Hence, the installed PID controller can be considered to be operating with near optimal PID control performance if BI_{LQG} is close to unity. A small value of BI_{LQG} would indicate re-tuning of the installed controller advisable. The actual plant cost to be calculated on-line would be,

$$J_{act} = \lim_{T_f \to \infty} E\left\{\int_0^{T_f} \left[(Qe(t))^2 + (Ru(t))^2\right] dt\right\} \tag{5.161}$$

The generic loop monitor algorithm for this type of LQG loop assessment can be given in terms of the following steps.

Algorithm 5.2.1. Loop monitoring algorithm

1. Process estimation

   On-line data used to estimate $\hat{K}$, $\hat{K}_1$, and $\hat{\tau}$.

2. Optimal LQG cost calculation for the particular P, PI, PID controller.

3. Process PID cost (equation (5.161)) is calculated with on-line data.

4. Assessment and decision by computing the benchmark index (BI_{LQG}) and comparing to unity to quantify performance achieved.

5.2.4 Case studies

Three case studies are presented in order to exemplify the use of this tuning method and monitoring technique. The first two examples compare the optimal LQG tuning with the sustained oscillation version of the Ziegler-Nichols methods, and then compares their responses. The third case, is an example of the loop monitoring algorithm.

All three examples employ a linear continuous model, which can be representative of a wastewater treatment process. In particular, as discussed in Chapter 3 and in (Mahathanakiet *et al.*, 2002), dissolved oxygen can easily approximated to a first or second order linear model. Therefore this case studies assume that dissolved-oxygen can be modelled by a first order system for the controller design, and the plant process is of third order.

5.2.4.1 Process model

The test system is chosen to illustrate some features of interest. The design plant is,

$$G_m = \frac{4}{3s+1}$$

The noise process is,

$$G_d = \frac{0.5}{3s+1}$$

The simulated plant includes some model mismatch,

$$G_p = \frac{4}{(3s+1)(0.2s+1)^2}$$

The LQG optimal control uses $W(s) = G_m(s)$, $W_d(s) = G_d(s)$ and $W_r(s) = 0$. Thus, the design studies uses the models, $G_m(s)$ and $G_d(s)$, but the simulation trials use the mismatched model, $G_p(s)$.

5.2.4.2 Ziegler-Nichols design

A sustained oscillation experiment was performed and PID controls designed using the Quarter Amplitude decay rules. The ultimate gain and period were identified as, $K_u = 8.5$ and $P_u = 1.2$. The calculated PID-type controller gains and time constants are presented in Table (5.7).

Table 5.7: Ziegler-Nichols design

Controller	K_p	τ_i	τ_d
P	4.250	-	-
PI	3.825	0.9996	-
PID	5.100	0.600	0.15

5.2.4.3 LQG PID control

The model data was used with the theory of Theorems 5.2.1 and 5.2.2 to find the optimal gains. Very little design data is required: the model parameters $K = 4$, $K_1 = 0.5$ and $\tau = 3$ are needed and Table (5.8) shows the interpretations of the design parameters. Table (5.9) presents the calculated parameters for a set of values of ρ and τ_r.

Finally Figures (5.17), (5.18) and (5.19) show the step response simulation results of the designed controllers.

Table 5.8: LQG PID design rules

Controller	Notes
P	The relative speed of response, and the amount of DC offset is given by $\rho > 0$. Smaller ρ values give faster responses and smaller DC offsets but increases the overshoot.
PI	The relative speed of response is given by $\rho > 0$. Smaller ρ values give faster responses. Set $\tau_r = 0$.
PID	The relative speed of response is given by $\rho > 0$. Smaller ρ values give faster responses. Parameter $\tau_r > 0$ increases the amount of derivative control introduced.

Table 5.9: LQG PID control design

	ρ	τ_r	K_p	τ_i	τ_d
P	0.4	N/A	2.2625	-	-
	0.7	N/A	1.2003	-	-
PI	0.4	0	1.7026	0.6810	-
	0.7	0	1.2350	0.8645	-
PID	0.7	0.2	1.8475	0.9324	0.1571
	0.7	0.6	4.1187	1.1316	0.2819

5.2.4.4 Loop monitoring

The Loop Monitor concept depends on being able to compute benchmark cost values. Recall the online cost function to be computed as equation (5.161). Then the theoretical value of J_{opt} is used to find the Benchmark index BI_{LQG}. Steady system regulation conditions are necessary for a valid comparison to be made across different control designs. Table (5.10) shows some selected cost and benchmark calculations obtained from 10 independent trials with PI control and different noise sequences. The ZN PI control is only 56% as effective as the LQG PI optimal control and re-tuning is strongly advisable. Figure (5.20) shows the cost computation traces for the two controllers for one of the ten trials made. The clear separation shows significant difference in ZN and optimal PI control performance.

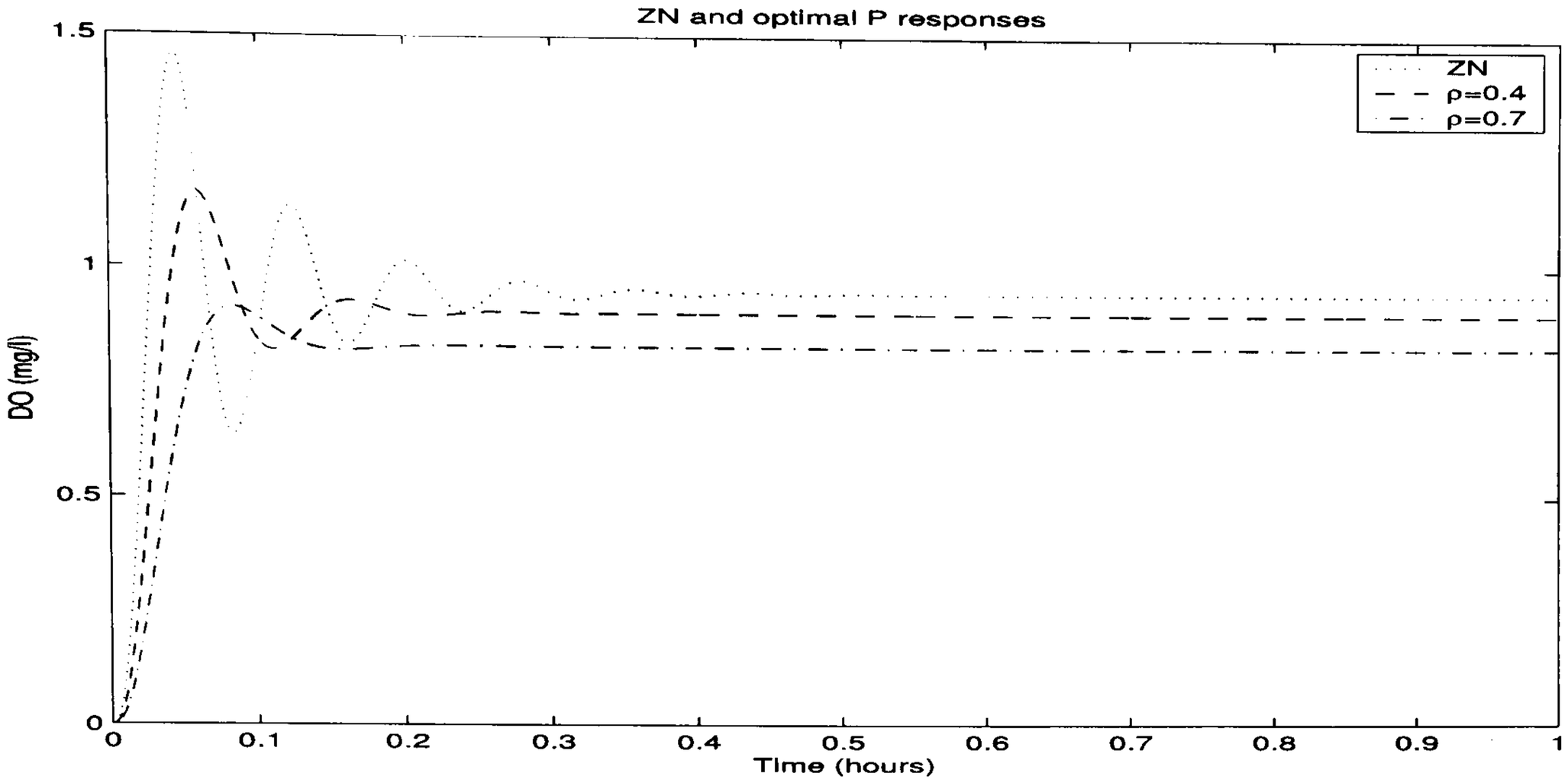

Figure 5.17: ZN and optimal P responses

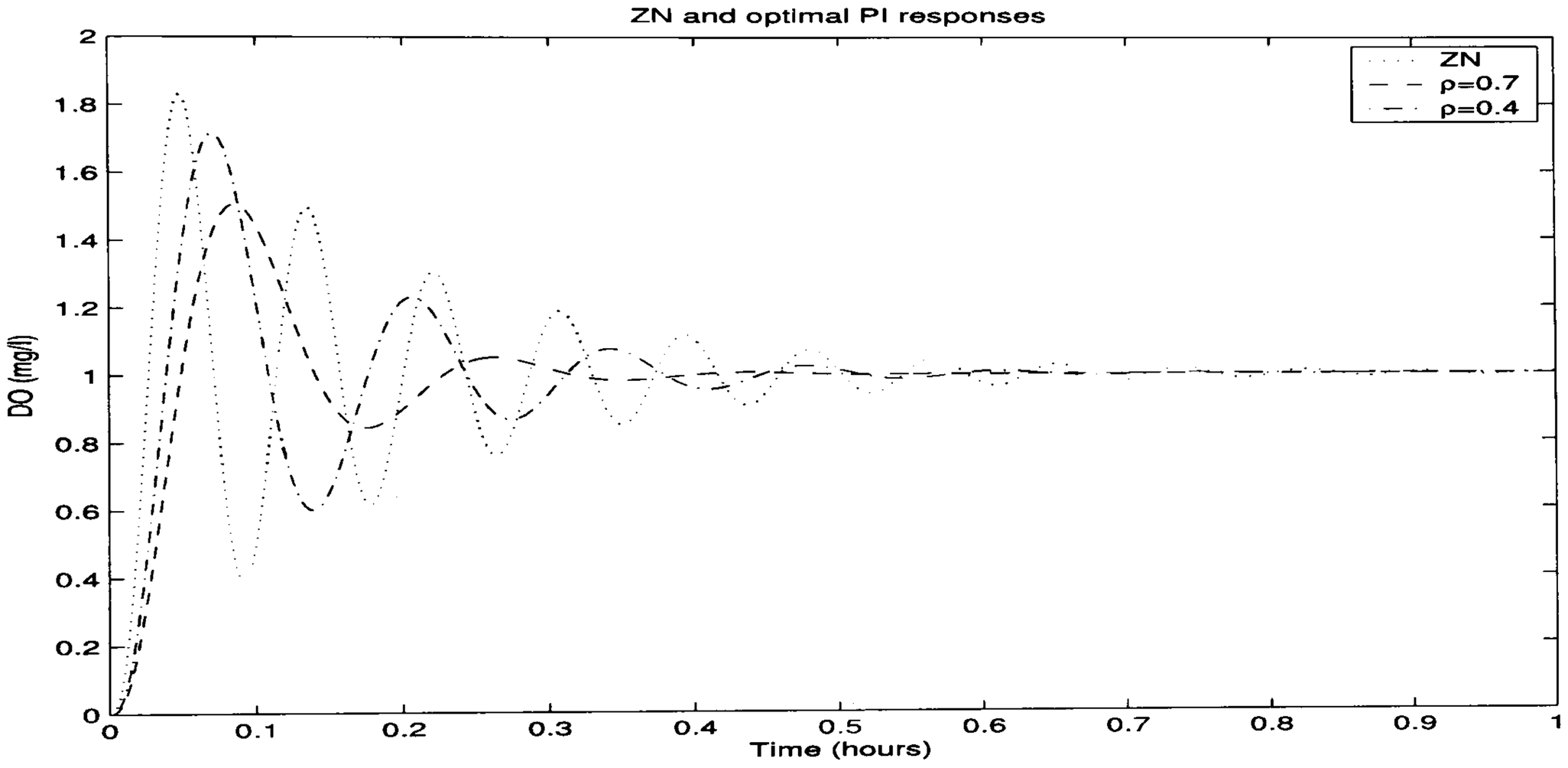

Figure 5.18: ZN and optimal PI responses

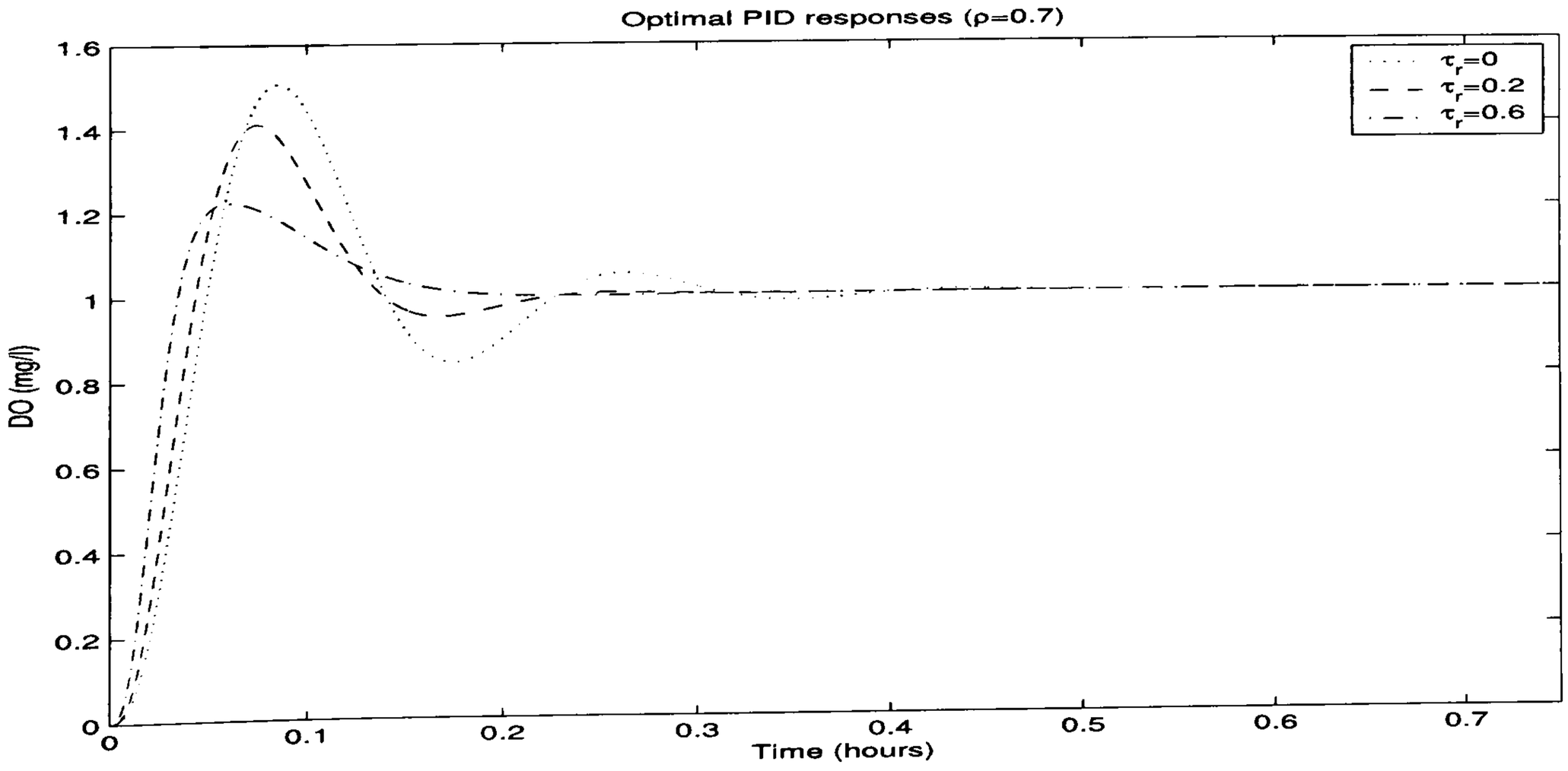

Figure 5.19: Optimal PID designs

Table 5.10: Benchmark computations for PI control

	Design	model	Mismatched	model
	J_{opt}	J_{ZN}	J_{opt}	J_{ZN}
Estimated cost	0.0126	0.0222	0.0220	0.0951
Theoretical cost	0.0126	-	-	-
BI_{LQG}	1	0.5676	-	-
Estimated values	-	-	1	0.2313

Degree of mismatch=J_{opt}(design model)/J_{opt}(actual plant)=0.5727

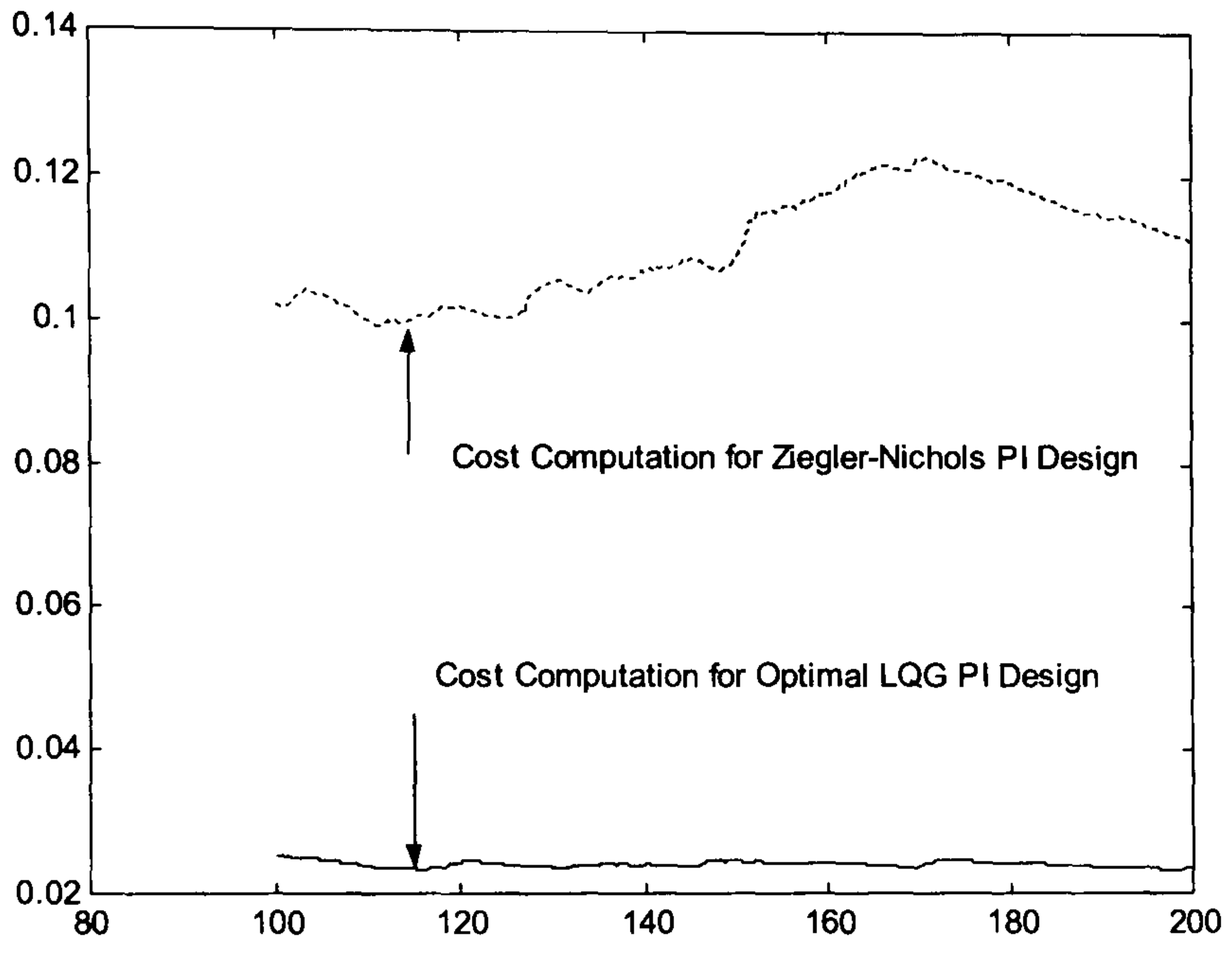

Figure 5.20: Process and optimal cost

5.3 Summary

Some loops in process control and in wastewater treatment plant can be modelled by simple first order stochastic system processes. However in certain conditions it will not always be possible to identify directly the system. Also, the controllers on such loops are almost always of PID form. This chapter has presented the development of two methods to tune PID type controllers: IFT and LQG, and process loop monitoring algorithm. This section summarises the results and contributions presented throughout the chapter.

IFT tuning

The first section of this chapter is the development of a continuous-time deterministic formulation of Iterative Feedback Tuning (IFT) for SISO and MIMO systems, which was originally introduced by Hjalmarsson *et al.* (1994); Hjalmarsson (1999) for discrete stochastic systems and for SISO deterministic continuous systems by Mahathanakiet *et al.* (2002). Therefore the main contribution of this section is the development of the MIMO continuous deterministic formulation.

IFT does not require explicit models, but a set of experiments to be performed over the plant. A cost function is minimised by implicitly calculating the gradient using signals recorded from the experiments.

For the SISO case, two case studies are presented: (a) steepest descent optimisation (b) Newton type update. The Newton type update requires the calculation of the Hessian matrix. The calculation of the Hessian matrix is performed by an additional experiment. Simulation results of both approach demonstrate that the use of Hessian produces an optimisation procedure which is more sensitive than the standard formulation. The calculation of the Hessian is however considerably more complex than for the standard formulation.

Simulation results of the MIMO formulation show the versatility of the method. However, the method showed some limitations in the degrees of freedom to improve the optimisation routine for each individual loop. Also, there is a considerable amount of experiments required. The number of experiments depends on the number of parameters to tune.

LQG tuning and process loop monitoring

The second method presented in this chapter employs simple models which lead to explicit solutions of a LQG polynomial system optimal control problem. An original

theoretical contribution of this chapter was to show that the optimal LQG controller belongs to the PID family. Resorting to a benchmark framework then enables optimal LQG-PID controllers and corresponding benchmark values to be compared with actual PID controller performance.

The method developed provides very simple equations for the design of a PID type controller. The simulation results corraborate the theoretical results. Further, the method development also gives explicit equations for the calculation of the cost function, and therefore a simple to use benchmark is easy to derive.

Simulation results present the comparative effects of changing the tuning parameters and also a comparison with the sustained oscillation Ziegler-Nichols method.

Chapter 6

Data-driven design of restricted structure controllers

Restricted-structure controllers are those whose structure has been fixed independently of the plant order. Even further, in some cases the controller parameters are fixed to a certain range of allowed values. In general, these controllers are of a lower order than the plant they control. Typical examples of such controllers, commonly employed in industry, are phase lead, phase lag, phase lead-lag and PID type controllers.

Design methods for these controllers use a wide range of control theory and computational approaches. Therefore, it is quite difficult to provide a classification of all the design techniques developed. This chapter will consider the broad classifications of model-based and model-free methods. Model based methods can be further categorised into parametric model methods and non-parametric model methods. The parametric model methods can use either transfer function models, state-space models or, as in this section, data-based identified subspace models. Non-parametric methods usually use one or two frequency response data points on which to base the controller design. On the other hand, model-free methods manage to use plant responses directly without the need for an intervening model description.

The general problem presented in this chapter is the tuning of conventional deterministic controllers, for example of PID type, such that their performance is as close as possible to that obtainable from a full order LQG controller. The solution procedure involves optimised restricted structure multivariable controllers and models from the subspace identification paradigm. The restriction of the multivariable structure of the process controller allows: (a) the controller input-output structure, (b) the number of controller parameters within the individual controller elements and (c) the numerical range of the controller parameters all to be defined. The subspace identification part of the method involves the use of closed loop plant data and subspace identification techniques to develop a linear process model in matrix form.

The chapter is organised in the following way: first, the subspace framework employed to develop the method is briefly introduced in section 6.1. Section 6.2 discusses the characterisation of a univariate control system. Section 6.3, extends the univariate results to the multivariable case thus providing a more general perspective. Later, in section 6.4 the method and its conditions are developed. Several simulation case studies are presented in section 6.5. The chapter ends with a brief summary of the main results achieved.

6.1 A subspace framework

This section presents a subspace framework necessary for the development of the tuning method in the following sections. This framework is the intermediate system representation between the input-output equations and the full state-space matrix representation as presented in Chapter 3.

Consider a plant described by the set of equations (6.1-6.2), operating in closed-loop driven by a controller described by equations (6.3-6.4), as in Figure (6.1).

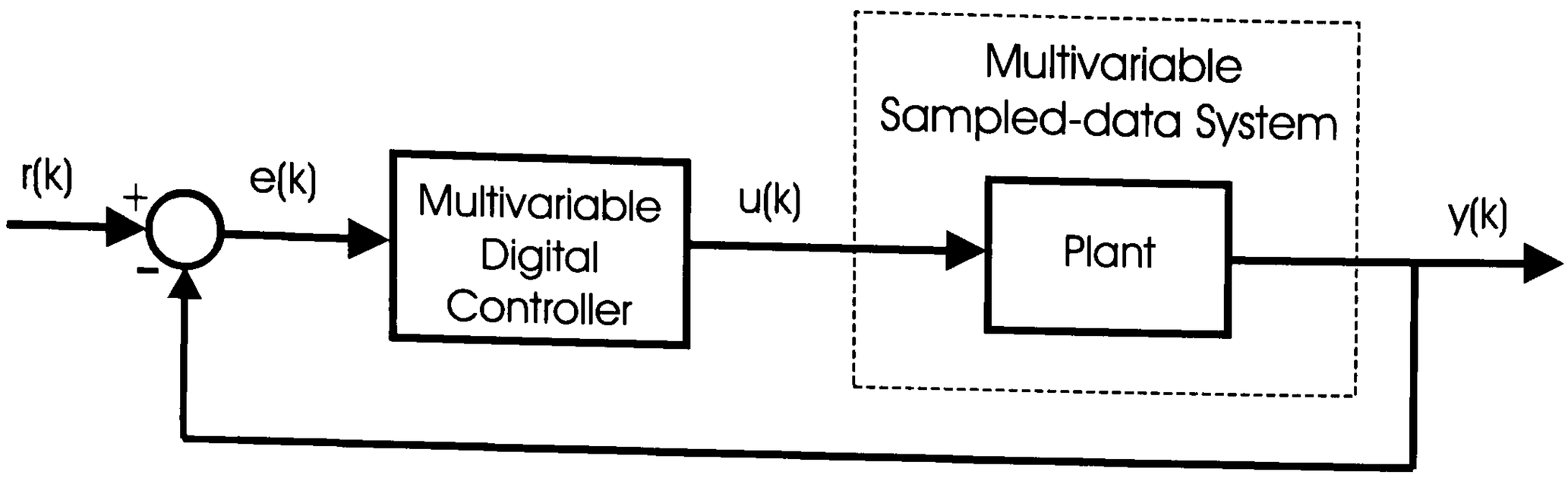

Figure 6.1: Closed-Loop System

$$x(k+1) = Ax(k)+Bu(k)+Kv(k) \tag{6.1}$$

$$y(k) = Cx(k)+Du(k)+v(k) \tag{6.2}$$

$$x_c(k+1) = A_c x_c(k)+B_c[r(k)-y(k)] \tag{6.3}$$

$$u(k) = C_c x_c(k)+D_c[r(k)-y(k)] \tag{6.4}$$

where $u(k) \in \mathbb{R}^m$, $y(k) \in \mathbb{R}^m$, $x(k) \in \mathbb{R}^n$ and $x_c(k) \in \mathbb{R}^l$ are the process inputs, outputs, states and controller states respectively. K is the the Kalman filter gain and $v(k)$ is an unknown innovation sequence with covariance equal to:

$$E\left[v(k)v^T(k)\right] = S \tag{6.5}$$

The problem can then be formulated as how to identify the plant parameters using closed-loop data and knowledge from the controller. There are several ways of solving this problem as for example the method presented by Verhaegen (1993) using MOESP. The method presented in (Favoreel *et al.*, 1998; van Overschee and De Moor, 1996*a*) has been adopted in this study, and is described below.

Consider a sufficiently large amount of data $\{u(k)\}$ and $\{y(k)\}$ and knowledge of the controller parameters so the past and future block Hankel matrices for $u(k)$ and $y(k)$ can be constructed by considering a backward and future horizon of dimension N, as in equations (6.6) and (6.7) for the case of $u(k)$, where j is the number of columns.

$$U_p = \frac{1}{\sqrt{j}} \begin{bmatrix} u_0 & u_1 & \cdots & u_{j-1} \\ u_1 & u_2 & \cdots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{N-1} & u_N & \cdots & u_{N+j-2} \end{bmatrix} \tag{6.6}$$

$$U_f = \frac{1}{\sqrt{j}} \begin{bmatrix} u_N & u_{N+1} & \cdots & u_{N+j-1} \\ u_{N+1} & u_{N+2} & \cdots & u_{N+j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2N-1} & u_{2N} & \cdots & u_{2N+j-2} \end{bmatrix} \tag{6.7}$$

The factor $\frac{1}{\sqrt{j}}$ has been added for statistical purposes. In general, subspace identification assumes that there are long time series of data ($j \to \infty$), and the data is ergodic. Due to this, the expectation operator E (average over a finite number of experiments) can be replaced with a different operator E_j applied to the sum of variables. This operator can therefore be defined as in equation (6.8). Thus, the factor $\frac{1}{\sqrt{j}}$ has the function of preprocessing the data matrices.

$$E_j = \lim_{j \to \infty} \frac{1}{j} [\bullet] \tag{6.8}$$

The matrix input-output equation for the plant, (equation (6.9)) and for the controller (equation (6.10)) can be obtained by recursive substitution of equations (6.1-6.2) and (6.3-6.4) respectively.

$$Y_f = \Gamma_N X_f + H_N U_f + H_N^s E_f \tag{6.9}$$

$$U_f = \Gamma_N^c X_f^c + H_N^c \left(R_f - Y_f \right) \tag{6.10}$$

The matrices H_N, Γ_N, H_N^c and Γ_N^c are the lower block triangular Toeplitz matrices and

the extended observability matrices for the process and the controller respectively, and are defined as in equations (6.11) and (6.12) for the process in (6.1) and (6.2).

$$H_N = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B & CA^{N-3}B & \cdots & D \end{bmatrix} \tag{6.11}$$

$$\Gamma_N = \begin{bmatrix} C & CA & CA^2 & \cdots & CA^{N-1} \end{bmatrix}^T \tag{6.12}$$

Substituting equation (6.10) in (6.9), gives the expression for the system operating in closed-loop as presented in equation (6.13).

$$Y_f = T_N \Gamma_N X_f + T_N H_N N_f + H_N^s E_f \tag{6.13}$$

where $T_N^{-1} = I + H_N H_N^c$ and $N_f = \Gamma_N^c X_f^c + H_N^c R_f$. Additionally from equation (6.10) it is evident that equation (6.14) holds, and N_f is uncorrelated to E_f since $X_f^c E_f^T = 0$ and $R_f E_f^T = 0$.

$$N_f = U_f + H_N^c Y_f \tag{6.14}$$

Therefore, the output prediction $\hat{Y}_f$ can be estimated when $j \to \infty$, as:

$$\hat{Y}_f = L_w^c W_p + L_u^c N_f \tag{6.15}$$

where

$$W_p = \begin{bmatrix} Y_p \\ U_p \end{bmatrix}$$

The term $L_w^c W_p$, is a bank of Kalman filters, as proven in (van Overschee and De

Moor, 1996*b*, p. 69-72), therefore $\hat{Y}_f$ is considered to be the best possible prediction (estimate) of Y_f. The matrices L_w^c and L_u^c are calculated by minimising the Frobenius norm as in (6.16).

$$\min_{L_w^c, L_u^c} \left\| Y_f - \begin{bmatrix} L_w^c & L_u^c \end{bmatrix} \begin{bmatrix} W_p \\ N_f \end{bmatrix} \right\|_F^2 \tag{6.16}$$

The numerical implementation to find L_w^c and L_u^c is a *RQ decomposition* as defined in equation (6.17). Using this decomposition it is possible to prove that L_w^c and L_u^c can be calculated as in equation (6.18), where † denotes the Moore-Penrose pseudo inverse (van Overschee and De Moor, 1996*b*; Ruscio, 2000).

$$\begin{bmatrix} W_p \\ U_f \\ Y_f \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{21} & R_{22} & 0 \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{bmatrix} \tag{6.17}$$

$$\begin{bmatrix} L_w^c & L_u^c \end{bmatrix} = \begin{bmatrix} R_{31} & R_{32} \end{bmatrix} \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix}^{\dagger} \tag{6.18}$$

The closed-loop matrices L_w^c and L_u^c are related to the open-loop matrices L_w and L_u by equations (6.19) and (6.20), which are derived by comparing the closed-loop model with its equivalent in open-loop.

$$L_u^c = T_N L_u \tag{6.19}$$

$$L_w^c = T_N L_w \tag{6.20}$$

Finally, the open-loop matrices L_w and L_u can be found by using equations (6.21) and (6.22), which are obtained by substituting the expression for T_N into (6.19-6.20) and rearranging. L_w must be approximated to a rank-n deficient matrix, where 'n' is estimated by a *Singular Value Decomposition*, as in (6.23).

$$L_u = L_u^c (I - H_N^c L_u^c)^{-1} \tag{6.21}$$

$$L_w = (I + L_u H_N^c) L_w^c \tag{6.22}$$

$$L_w = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_n & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{6.23}$$

The plant prediction model is then given as a function of the future input vector $\hat{u}_f$ and the past input-output vector w_p as in equation (6.24). Note that the identification method is valid even if the signals are generated by purely deterministic systems. This, however, can lead to rank deficient Hankel matrices, which can produce numerical problems depending on the decomposition algorithm employed. This phenomenon is produced when a straightforward implementation of the Schur algorithm is used to compute the R factor in a fast RQ decomposition using the *Hankel structure*. As stated in(van Overschee and De Moor, 1996*b*, p. 163), this is not often the case in practice, however systems with many outputs or with heavily coloured input signals can generate Hankel matrices that are nearly rank deficient.

$$\hat{y}_f = L_w w_p + L_u \hat{u}_f \tag{6.24}$$

where

$$\hat{u}_f = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix}^T$$

$$w_p = \left[\begin{array}{ccc|ccc} y_{-N+1} & \cdots & y_0 & u_{-N+1} & \cdots & u_0 \end{array}\right]^T$$

6.1.1 Incremental subspace representation

Equation (6.24) gives the best prediction of the output $\hat{y}_f$ given the future inputs $\hat{u}_f$ and past output-inputs w_p. However, it is sometimes more useful to have a model defined

in terms of the changes in the control signal rather than the signal itself. To do such a modification, several approaches have been suggested in the literature as in Ruscio (1997*a*) and Kadali *et al.* (2003). Both approaches yield the same formulas; however, they consider different signal frameworks. In Ruscio (1997*a*), a deterministic approach is considered, while in Kadali *et al.* (2003) a stochastic framework is employed. The two approaches are presented in this section.

6.1.1.1 Deterministic case

Consider the process described by equations (6.1) and (6.2), where $v(k)$ has been set to zero to assume a deterministic system. Define then a new state variable $z(k)$ such that $z(k) = x(k) - x(k-1)$. Then, the system equations can be transformed into equations (6.25) and (6.26), and therefore the process matrix input-output equation is as presented in (6.27). Note, that the extended observability matrix and the block lower triangular Toeplitz matrix have not changed since the system matrices are the same.

$$z(k+1) = Az(k) + B\Delta u(k) \tag{6.25}$$

$$\Delta y(k) = Cz(k) + D\Delta u(k) \tag{6.26}$$

$$\Delta Y_f = \Gamma_N Z_f + H_N \Delta U_f \tag{6.27}$$

Using equations (6.25) and (6.26) recursively it is possible to calculate the output prediction $\hat{y}$ at instant k+N, as presented in equation (6.28).

$$\begin{aligned} \hat{y}(k+N) = {} & y(k) + \left(CA^{N-1} + \cdots + C\right) z(k+1) + \\ & \left(CA^{N-2}B + \cdots + CB + D\right) \Delta u(k+1) + \\ & \cdots + \left(CB + D\right) \Delta u(k+N-1) + D\Delta u(k+N) \end{aligned} \tag{6.28}$$

Therefore the output prediction is:

$$\hat{y}_f = y_t + \Gamma_N^{\Delta} z(k+1) + H_N^{\Delta} \Delta \hat{u}_f \tag{6.29}$$

where

$$\Gamma_N^{\Delta} = \begin{bmatrix} C \\ CA + C \\ \vdots \\ CA^{N-1} + \cdots + C \end{bmatrix} \tag{6.30}$$

$$H_N^{\Delta} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB + D & D & \cdots & 0 \\ CAB + CB + D & CB + D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B + \cdots + D & CA^{N-3}B + \cdots + D & \cdots & D \end{bmatrix} \tag{6.31}$$

By comparing Γ_N^{Δ} and H_N^{Δ} with Γ_N and H_N, it is simple to verify the relations in (6.32) and (6.33), where l_{w_i} is the i^{th} row block of dimension *'m'* of L_w. Therefore, the incremental form of the system is presented in equation (6.34).

$$L_u^{\Delta} = L_u \begin{bmatrix} I_{mN} & 0 & \cdots & 0 \\ I_{mN} & I_{mN} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ I_{mN} & I_{mN} & \cdots & I_{mN} \end{bmatrix} \tag{6.32}$$

$$L_w^{\Delta} = \begin{bmatrix} l_{w_1} \\ l_{w_2} + l_{w_1} \\ \vdots \\ \sum_{i=1}^{N} l_{w_i} \end{bmatrix} \tag{6.33}$$

$$\hat{y}_f = y_t + L_w^{\Delta} \Delta w_p + L_u^{\Delta} \Delta \hat{u}_f \tag{6.34}$$

6.1.1.2 Stochastic case

Assuming the process defined by equations (6.1) and (6.2), using an integrating white noise model as in equations (6.35) and (6.36), and making the same change of variable as in the previous case, the incremental stochastic process model is then described by equations (6.37) and (6.38). Consequently, the input-output matrix equation is presented in (6.39).

$$v(k+1) = v(k) + a(k) \tag{6.35}$$

$$v(k) = \frac{a(k)}{\Delta} \tag{6.36}$$

$$z(k+1) = Az(k) + B\Delta u(k) + Ka(k) \tag{6.37}$$

$$\Delta y(k) = Cz(k) + D\Delta u(k) + a(k) \tag{6.38}$$

$$\Delta Y_f = \Gamma_N Z_f + H_N \Delta U_f + H_N^s A_f \tag{6.39}$$

The output prediction at sampling instant $k+N$ can then be calculated as in (6.40).

$$\begin{aligned} \hat{y}(k+N) = {} & y(k) + \left(CA^{N-1} + \cdots + C\right) z(k+1) + \\ & \left(CA^{N-2}B + \cdots + CB + D\right) \Delta u(k+1) + \cdots \\ & + (CB + D)\, \Delta u(k+N-1) + D\Delta u(k+N) \\ & + (a(k) + \cdots + a(k+N)) \end{aligned} \tag{6.40}$$

This can be written in the same way as in equation (6.34); therefore the best output prediction can be calculated in the same manner as in the deterministic case.

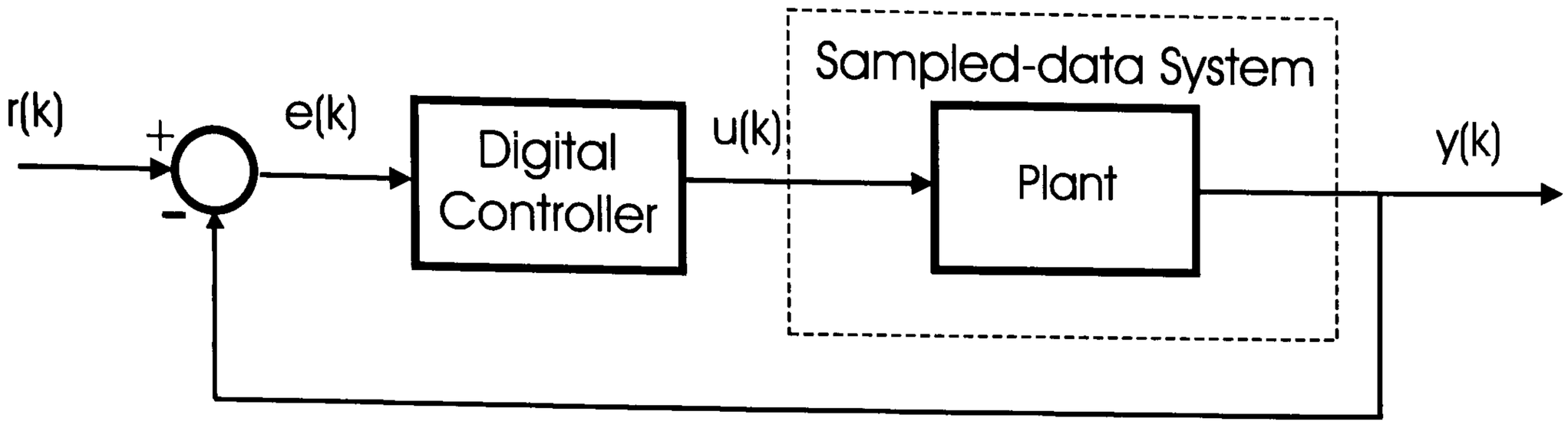

Figure 6.2: Closed-loop SISO system

6.2 Univariate restricted structure controller characterisation

This section presents an introduction to the problem of the parameterisation and definition of the structure for a single input - single output (SISO) plant. First the parameterisation of a typical PID type discrete controller is presented. Later, the controller structure for a SISO type system is introduced. A more complete and general description of both of these topics is given for the multivariable controller characterisation.

6.2.1 Controller parameterisation

Consider the SISO closed-loop system in Figure (6.2), with plant model described by equation (6.34). Assume a sufficiently extensive database of the signals r, y and u. A discrete PID controller can be defined by equation (6.41), and the incremental control action by equation (6.42), where k_p, k_i and k_d are the proportional, integral and derivative gains respectively.

$$u(k) = k_p \cdot e(k) + k_i \sum_{n=1}^{k} e(n) + k_d \left[e(k) - e(k-1) \right] \tag{6.41}$$

$$\Delta \hat{u}(k) = u(k) - u(k-1) \tag{6.42}$$

From equations (6.41) and (6.42), it is easy to find the time expression for the control action increment as in equation (6.43).

$$\Delta\hat{u}(k) = \rho^{(1)}e(k)+\rho^{(2)}e(k-1)+\rho^{(3)}e(k-2) \tag{6.43}$$

where: $\rho^{(1)} = k_p + k_i + k_d$, $\rho^{(2)} = -k_p - 2k_d$ and $\rho^{(3)} = k_d$.

Equation (6.43) can be parameterised as in equation (6.44).

$$\Delta\hat{u}(k) = \begin{bmatrix} e(k) & e(k-1) & e(k-2) \end{bmatrix} \begin{bmatrix} \rho^{(1)} \\ \rho^{(2)} \\ \rho^{(3)} \end{bmatrix} \tag{6.44}$$

Which in compact notation can be written as:

$$\Delta\hat{u}_0 = \begin{bmatrix} e_0 & e_{-1} & e_{-2} \end{bmatrix} \cdot \rho \tag{6.45}$$

To comply with a digital PID structure the controller parameter vector ρ must comply with the following linear constraints:

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \rho \leq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.46}$$

or equivalently:

$$\varphi \cdot \rho \leq 0 \tag{6.47}$$

The future increment control action can then be written as in equation (6.48) and in compact form in equations (6.49):

$$\Delta\hat{u}_f = \begin{bmatrix} e_1 & | & e_0 & | & e_{-1} \\ e_2 & | & e_1 & | & e_0 \\ \vdots & | & \vdots & | & \vdots \\ e_N & | & e_{N-1} & | & e_{N-2} \end{bmatrix} \rho \tag{6.48}$$

$$\Delta\hat{u}_f = \varepsilon(\rho) \cdot \rho \tag{6.49}$$

where

$$\varepsilon(\rho) = \begin{bmatrix} \xi_1 & | & \xi_2 & | & \xi_3 \end{bmatrix} \tag{6.50}$$

6.2.2 Univariate controller structure

In equation (6.49) it is still necessary to calculate the matrix of closed-loop errors $\varepsilon(\rho)$. This matrix can be directly derived from the error definition. Since r_f has been set to zero, it is evident that the matrix ε is equal to:

$$\begin{aligned} \varepsilon &= \begin{bmatrix} \xi_1 & | & \xi_2 & | & \xi_3 \end{bmatrix} \\ &= -\begin{bmatrix} T_{f_1} \cdot \hat{y}_f & | & T_{f_2} \cdot \hat{y}_f & | & T_{f_3} \cdot \hat{y}_f \end{bmatrix} \\ &- \begin{bmatrix} T_{p_1} \cdot y_p & | & T_{p_2} \cdot y_p & | & T_{p_3} \cdot y_p \end{bmatrix} \end{aligned} \tag{6.51}$$

where:

$$T_{f_1} = I_N \tag{6.52}$$

$$T_{f_2} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \tag{6.53}$$

$$T_{f_3} = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \\ 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \end{bmatrix} \tag{6.54}$$

$$T_{p_1} = 0_N \tag{6.55}$$

$$T_{p_2} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{6.56}$$

$$T_{p_3} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{6.57}$$

As will be discussed in a later section, the matrices T_f and T_p can be directly calculated from the controller structure. At present, these matrices do not show the controller structure by simple inspection. The discussion in this section will be limited to say that a SISO controller of this type will have a structure as presented in equation (6.58). Later in the chapter, a complete discussion regarding the controller structure and its importance in the controller design is presented.

$$\Lambda = \langle \{1\} \rangle \tag{6.58}$$

By using equations (6.34) , (6.49), (6.51), and defining $\Phi = r_f - y_t - L_w^{\Delta} \Delta w_p$, the following set of equations are obtained:

$$\Omega(\rho)\begin{bmatrix}\xi_1\\ \xi_2\\ \xi_3\end{bmatrix} = \omega \tag{6.59}$$

where

$$\Omega = \begin{bmatrix} I_N+\rho^{(1)}L_u^{\Delta} & \rho^{(2)}L_u^{\Delta} & \rho^{(3)}L_u^{\Delta} \\ \rho^{(1)}T_{f_2}L_u^{\Delta} & I_N+\rho^{(2)}T_{f_2}L_u^{\Delta} & \rho^{(3)}T_{f_2}L_u^{\Delta} \\ \rho^{(1)}T_{f_3}L_u^{\Delta} & \rho^{(2)}T_{f_3}L_u^{\Delta} & I_N+\rho^{(3)}T_{f_3}L_u^{\Delta} \end{bmatrix} \tag{6.60}$$

$$\omega = \begin{bmatrix} \Phi \\ T_{f_2}\Phi - T_{p_2}y_p \\ T_{f_3}\Phi - T_{p_3}y_p \end{bmatrix} \tag{6.61}$$

It is interesting to see that the left hand side of (6.59) are signals to be predicted over the forward horizon (i.e. future), while the right hand side are signals previously recorded in the past.

6.3 Multivariable restricted-structure controller characterisation

The following section is divided into two parts. The first part considers the characterisation of a restricted-structure controller in terms of a finite number of parameters and an incremental control action. The second part concerns the definition of the multivariable controller structure by establishing the interactions between inputs and outputs.

6.3.1 Controller parameterisation

The method assumes that the controllers can be parameterised and expressed in an incremental form. A general form for this type of parameterisation is presented in equation (6.62), where ρ is a vector containing the controller parameters.

$$\Delta\hat{u}(k) = \Xi(k)\rho \tag{6.62}$$

This situation is very common and many industrial controllers can be written in this way. Since the plant has m inputs and outputs, let $\Delta\hat{u}_{i,j}(k+1)$ represent the incremental control action i, due to the error sequence of output j, where i and j vary between 1 and m. Then, the effective incremental control action for input i (i.e. $\Delta\hat{u}_i(k+1)$) can be calculated as the sum of each partial contribution, as in (6.63).

$$\Delta\hat{u}_i(k+1) = \sum_{j=1}^{m} \Delta\hat{u}_{i,j}(k+1) \tag{6.63}$$

By assuming the parameterisation in (6.62), the incremental control actions for all the plant inputs at sampling instant $k+1$ can be written as in (6.64), which has the general form presented in (6.62).

$$\Delta\hat{u}(k+1) = \mathscr{E}(k+1)\rho \tag{6.64}$$

where:

$$\rho = \left[\begin{array}{ccccccc|c} \rho_{1,1}^{(1)} & \rho_{1,1}^{(2)} & \rho_{1,1}^{(3)} & \cdots & \rho_{1,m}^{(1)} & \rho_{1,m}^{(2)} & \rho_{1,m}^{(3)} & \cdots \\ \multicolumn{7}{c|}{} & \rho_{m,1}^{(1)} & \rho_{m,1}^{(2)} & \rho_{m,1}^{(3)} & \cdots & \rho_{m,m}^{(1)} & \rho_{m,m}^{(2)} & \rho_{m,m}^{(3)} \end{array} \right]^T$$

$$\mathscr{E}(k+1) = \begin{bmatrix} \varepsilon(k+1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varepsilon(k+1) \end{bmatrix}$$

and

$$\varepsilon(k+1) = \begin{bmatrix} e_1(k+1) & e_1(k) & e_1(k-1) & \cdots & e_m(k+1) & e_m(k) & e_m(k-1) \end{bmatrix}$$

The indices in the notation $\rho_{i,j}^{(k)}$ indicate that the coefficient is the k^{th} parameter of the controller acting over input i using the error sequence from output j. Notice that in general $\rho \in \mathbb{R}^{(p \cdot m^2)}$, $\mathscr{E}(k+1) \in \mathbb{R}^{m \times (p \cdot m^2)}$ and $\varepsilon(k+1) \in \mathbb{R}^{1 \times (p \cdot m)}$ where p is the number of parameters of each individual controller. For the specific case of a PID controller, $p = 3$.

6.3.2 Multivariable controller structure

The controller characterisation leads to the definition of the structure of the multivariable controller. The definition of the multivariable controller structure allows the specification of the output errors which are used to compute each input. The following definition in conjuction with the parameterisation completely defines the controller.

Definition 6.3.1. Let a controller structure be defined by a matrix $\Lambda \in \mathtt{M}_m$ as in (6.65), which links the output error of the plant to its inputs through a restricted-structure multivariable controller. In matrix Λ, any $a_{i,j} \in \Lambda$, $a_{i,j} = 1$ implies an interconnection between the input i and the output j; $a_{i,j} = 0$ implies no interconnection.

$$\Lambda = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{bmatrix} \quad a_{i,j} \in \{0,1\} \tag{6.65}$$

The matrix Λ can then be decomposed into the sum of f matrices, where f is the number

of non-zero elements in Λ as in equations (6.66) and (6.67).

$$\Lambda = \begin{bmatrix} a_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{m,m} \end{bmatrix} \tag{6.66}$$

$$\Lambda = \Lambda_{1,1} + \cdots + \Lambda_{m,m} \tag{6.67}$$

Therefore, it is also possible to characterise the controller structure by the set enumerated in (6.68). This definition allows the description of any multivariable controller structure.

$$\Lambda = \langle \{\Lambda_{1,1}, \Lambda_{1,2}, \cdots, \Lambda_{m,m}\} \rangle \tag{6.68}$$

6.4 Parameter calculation

The approach developed in this thesis calculates a set of controllers by minimising a finite horizon LQG cost index. The algorithm uses the plant model described by equation (6.34) and the controller characterised by (6.64) and a subset $\mathscr{B} \subseteq \Lambda$. The method ensures that if the plant is linear, the controller is well characterised (no uncontrollable modes), and assuming that the optimisation converges, the solution of the problem is a controller which stabilises the plant when operating in closed-loop. The following section focuses in developing the method and the conditions to comply with the prior statement.

6.4.1 Cost index

Consider the closed-loop system in Figure (6.1), with plant model described by equation (6.34). Assume sufficient knowledge of the signals r, y and u. The problem is

to find a set of controller parameters such that the cost function of equation (6.69) is minimised.

$$J = \left(r_f - \hat{y}_f\right)^T \mathscr{Q} \left(r_f - \hat{y}_f\right) + \Delta \hat{u}_f^T \mathscr{R} \Delta \hat{u}_f \tag{6.69}$$

To simplify the numerical problem, it is assumed that the system is regulated around an operating point and all the signals are normalised. By using equation (6.64), the future incremental control output can be written as in (6.70).

$$\Delta \hat{u}_f = \varphi \rho \tag{6.70}$$

where

$$\varphi = \begin{bmatrix} \mathscr{E}(k+1) \\ \vdots \\ \mathscr{E}(k+N) \end{bmatrix}$$

The dimensions of the matrices and vectors are $\Delta \hat{u}_f \in \mathbb{R}^{(mN)}$ and $\varphi \in \mathbb{R}^{(mN)\times(pm^2)}$. By replacing equation (6.70) in (6.34) and in (6.69) and defining $\Phi = r_f - y_t - L_w^{\Delta} \Delta w_p$ it is possible to obtain an equivalent expression for the cost function given by (6.71).

$$\begin{aligned} J = {} & \rho^T \left(\varphi^T L_u^{\Delta^T} \mathscr{Q} L_u^{\Delta} \varphi + \varphi^T \mathscr{R} \varphi \right) \rho \\ & -2\rho^T \left(\varphi^T L_u^{\Delta^T} \mathscr{Q} \Phi \right) \\ & +\Phi^T \mathscr{Q} \Phi \end{aligned} \tag{6.71}$$

Equation (6.71) has the quadratic form $x^T Ax + x^T b + c$, which can be minimised either using efficient numerical methods or by directly computing the derivative of J with respect to ρ. However, the problem is not straightforward. There is still the problem of computing matrix φ, and also the provision of conditions so that the computed

multivariable controller gives a closed-loop stable system.

6.4.2 Formulation as a least-squares problem

The matrices φ and L_u^{Δ} are usually large and ill-conditioned, therefore the choice of a robust numerical algorithm to calculate the minimum of J in equation (6.71) is of vital importance. An efficient solution to this problem is to use a least-squares approach. Since $\mathscr{Q}$ and $\mathscr{R}$ are positive-definite or at least positive-semidefinite, it is possible to find matrices $S_{\mathscr{R}}$ and $S_{\mathscr{Q}}$ such that:

$$\begin{aligned} \mathscr{Q} &= S_{\mathscr{Q}}^T S_{\mathscr{Q}} \\ \mathscr{R} &= S_{\mathscr{R}}^T S_{\mathscr{R}} \end{aligned} \tag{6.72}$$

It is relatively simple to prove that the minimisation of (6.71) is equivalent to the minimisation of the norm of the vector of equation (6.73).

$$\min_{\rho} \left\| \begin{bmatrix} S_{\mathscr{Q}} \left\{ L_u^{\Delta} \varphi \rho - \Phi \right\} \\ S_{\mathscr{R}} \varphi \rho \end{bmatrix} \right\|^2 \tag{6.73}$$

So, ρ is the least-squares solution to:

$$\begin{bmatrix} S_{\mathscr{Q}} L_u^{\Delta} \varphi \\ S_{\mathscr{R}} \varphi \end{bmatrix} \rho = \begin{bmatrix} S_{\mathscr{Q}} \Phi \\ 0 \end{bmatrix} \tag{6.74}$$

6.4.3 Closed-loop condition

In equation (6.74) the problem of how to calculate the Hankel matrix of errors φ such that the system is stable remains to be considered. A solution to this problem is to calculate the future errors based on past data by satisfying the closed-loop equations. To do so the matrix φ can be decomposed into pm^2 column vectors as in (6.75), where

the indices hold the same meaning as in $\rho_{i,j}^{(k)}$.

$$\varphi = \left[\begin{array}{ccccccc|c} \xi_{1,1}^{(1)} & \xi_{1,1}^{(2)} & \xi_{1,1}^{(3)} & \cdots & \xi_{1,m}^{(p-2)} & \xi_{1,m}^{(p-1)} & \xi_{1,m}^{(p)} & \cdots \\ \hline \xi_{m,1}^{(1)} & \xi_{m,1}^{(2)} & \xi_{m,1}^{(3)} & \cdots & \xi_{m,m}^{(p-2)} & \xi_{m,m}^{(p-1)} & \xi_{m,m}^{(p)} & \end{array} \right] \tag{6.75}$$

Each error vector $\xi_{i,j}^{(k)}$ can be calculated by solving equation (6.76) using (6.75) for $i = 1 \cdots m$, $j = 1 \cdots m$ and $k = 1 \cdots p$.

$$\xi_{i,j}^{(k)} = -T_{f_{i,j}}^{(k)} \left\{ L_u^{\Delta} \varphi \rho - \Phi \right\} - T_{p_{i,j}}^{(k)} y_p \tag{6.76}$$

The resulting system of equations is:

$$\Omega(\rho)\, \upsilon = \omega \tag{6.77}$$

where

$$\Omega(\rho) = \begin{bmatrix} I + \rho_{1,1}^{(1)} T_{f_{1,1}}^{(1)} L_u^{\Delta} & \rho_{1,1}^{(2)} T_{f_{1,1}}^{(1)} L_u^{\Delta} & \cdots & \rho_{m,m}^{(p)} T_{f_{1,1}}^{(1)} L_u^{\Delta} \\ \rho_{1,1}^{(1)} T_{f_{1,1}}^{(2)} L_u^{\Delta} & I + \rho_{1,1}^{(2)} T_{f_{1,1}}^{(2)} L_u^{\Delta} & \cdots & \rho_{m,m}^{(p)} T_{f_{1,1}}^{(2)} L_u^{\Delta} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,1}^{(1)} T_{f_{m,m}}^{(p)} L_u^{\Delta} & \rho_{1,1}^{(2)} T_{f_{m,m}}^{(p)} L_u^{\Delta} & \cdots & I + \rho_{m,m}^{(p)} T_{f_{m,m}}^{(p)} L_u^{\Delta} \end{bmatrix}$$

$$\upsilon = \begin{bmatrix} \xi_{1,1}^{(1)} \\ \xi_{1,1}^{(2)} \\ \vdots \\ \xi_{m,m}^{(p)} \end{bmatrix}$$

$$\omega = \begin{bmatrix} T_{f_{1,1}}^{(1)} \Phi - T_{p_{1,1}}^{(1)} y_p \\ T_{f_{1,1}}^{(2)} \Phi - T_{p_{1,1}}^{(2)} y_p \\ \vdots \\ T_{f_{m,m}}^{(p)} \Phi - T_{p_{m,m}}^{(p)} y_p \end{bmatrix}$$

The problem now relies on determining the matrices $T_{f_{i,j}}^{(k)}$ and $T_{p_{i,j}}^{(k)}$. The construction of these matrices is simple and comes as an immediate result from the controller structure definition. The controller structure is defined as a combination of the matrices of the set (or a subset) (6.68). The following Lemma states how the matrices $T_{f_{i,j}}^{(k)}$ and $T_{p_{i,j}}^{(k)}$ are constructed based on the controller structure.

Lemma 6.4.1. Let Λ be the set that generates all the possible combinations of controller structures of order m. Let $\mathscr{B}$ a multivariable controller structure defined as a subset of Λ ($\mathscr{B} \subseteq \Lambda$), such that $B_{i,j} \in \mathscr{B}$.

1. The matrix $T_{f_{i,j}}^{(k)}$ can be calculated as:

$$T_{f_{i,j}}^{(k)} = \left[\begin{array}{c|c} 0 & 0 \\ - & - \\ \beta & 0 \end{array}\right]$$

where

$$\beta = \left[\begin{array}{ccc} B_{i,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_{i,j} \end{array}\right]_{m(N-k+1)}$$

2. The matrices $T_{p_{i,j}}^{(k)}$ can be calculated as:

$$T_{p_{i,j}}^{(k)} = \left[\begin{array}{c|c} 0 & \gamma \\ - & - \\ 0 & 0 \end{array}\right]$$

where

$$\gamma = \begin{cases} \begin{bmatrix} B_{i,j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_{i,j} \end{bmatrix}_{(k-1)m} & for \quad k = 2 \cdots p \\ 0_m & for \quad k = 1 \end{cases}$$

Proof. This lemma can be verified by constructing the error vectors $\xi_{i,j}^{(k)}$, from the definition,

$$\xi_{i,j}^{(k)} = -T_{f_{i,j}}^{(k)} \hat{y}_f - T_{p_{i,j}}^{(k)} y_p \tag{6.78}$$

□

6.4.4 Stability condition

The problem of stability can be addressed by using a result presented by Giovanini and Marchetti (1999). This paper presents a proof that to assure exponential stability of restricted-structure digital controllers it is sufficient to comply with the following condition:

$$|\Delta \hat{u}(k+N)_i| \leq \sigma \qquad for\, i = 1 \cdots m \tag{6.79}$$

This condition transforms into the constraint (6.80). The constraint in (6.80) has to be solved simultaneously with the closed-loop condition of equation (6.77).

$$\Theta \rho \leq \Psi \tag{6.80}$$

where

$$\Theta = \begin{bmatrix} \mathscr{E}(k+N) \\ -\mathscr{E}(k+N) \end{bmatrix}$$

$$\Psi = \begin{bmatrix} \sigma \\ \vdots \\ \sigma \end{bmatrix}$$

The use of this condition in the optimisation will produce a controller with which the system is closed-loop stable; however, it might be possible that this condition cannot be met and in that case the optimisation will be infeasible. Some suggestions to avoid infeasible optimisations are to enlarge either the horizon N or the domain of attraction σ, or both.

Algorithm 6.4.1. Subspace multivariable restricted structure tuning algorithm

1. Record sufficient data of $y(k)$ and $u(k)$, and normalise the data around the operating point.

2. Estimate the forward-backward horizon (N) based on the settling time required.

3. Construct the data Hankel matrices and the controller Toeplitz matrix H_N^c.

4. Calculate N_f.

5. Calculate L_u^c and L_w^c by performing the RQ-decomposition and then use equation (6.18).

6. Calculate L_u and L_w using equations (6.21) and (6.22).

7. Calculate L_u^Δ and L_w^Δ using equations (6.32) and (6.33).

8. Define the number of parameters (p).

9. Define the controller structure.

10. Define the parameters range and express as an inequality constraint.

11. Calculate the matrices T_f and T_p, using Lemma 6.4.1.

12. Estimate initial values for the weight matrices $\mathscr{Q}$ and $\mathscr{R}$.

13. Estimate the stability condition (σ).

14. Estimate initial values for the controller parameters (ρ).

15. Solve $\begin{bmatrix} S_{\mathscr{Q}} L_u^{\Delta} \varphi \\ S_{\mathscr{R}} \varphi \end{bmatrix} \rho = \begin{bmatrix} S_{\mathscr{Q}} \Phi \\ 0 \end{bmatrix}$ in the least squares sense, subject to $\Theta\rho \leq \Psi$ (stability condition) and $\Omega(\rho)\upsilon = \omega$ (closed-loop condition), and any additional restrictions over the parameters range.

6.5 Simulation case studies

This section examines four controller structures which are used as examples of the method developed in this chapter. The examples comprise the control of dissolved oxygen in a simulation benchmark for an activated sludge wastewater treatment plant introduced in section 2.2. Air is pumped into the reactors through blowers (actuators) which are commanded by a PID type controller.

The case studies examine the design of four possible controller configurations for the treatment plant. The examples present simulation results that have been obtained using MATLAB/SIMULINK R11.

6.5.1 Univariate controller structure

The following simulation exemplifies the use of the tuning algorithm for SISO type plant. The simulation benchmark is used by considering that only the last aerated

Table 6.1: Optimisation Specifications

N	M	R	Q	σ
100	100	8	5	0.01

Table 6.2: Controller Parameters

	$\rho^{(1)}$	$\rho^{(2)}$	$\rho^{(3)}$
Initial parameters	1	-0.9308	0
Optimal parameters	8.2673	-12.2142	4.5825

reactor is controlled. The input to the model is the airflow rate scaled to a base of 10 and the output is the oxygen concentration in the reactor. The airflow is pumped into the reactor through blowers (actuators) which are commanded by a PID controller. Figure (3.1) shows details of the control loop in the fifth aerobic reactor. The control loop also accounts for unmeasured disturbances as changes in the plant load. These disturbances are included in the form of the signal $d(k)$. The identification considers a 1 minute sampling rate with an oxygen sensor with 1 minute time delay. The initial controller, with which the plant was identified in closed-loop, is a PI with parameters shown in Table (6.2) .

For the identification, 1200 points of data have been collected with the system excited by a pseudo random binary signal (PRBS) of zero mean and 0.5 [mg/l] amplitude around a 1 [mg/l] setpoint. The forward and backward horizon have been set to a length of 100. The system is approximated by the first five singular values ($n = 5$) in the singular value decomposition of the identification algorithm. The algorithm was implemented in MATLAB, and the solution takes around 10 to 20 iterations depending on how stringent the stability constraints are, as shown in Figure (6.3). Figure (6.4) shows the response of the obtained controller for a unit step when designed with the specifications in Table (6.1) , while Table (6.2) shows the initial and optimal controller parameters.

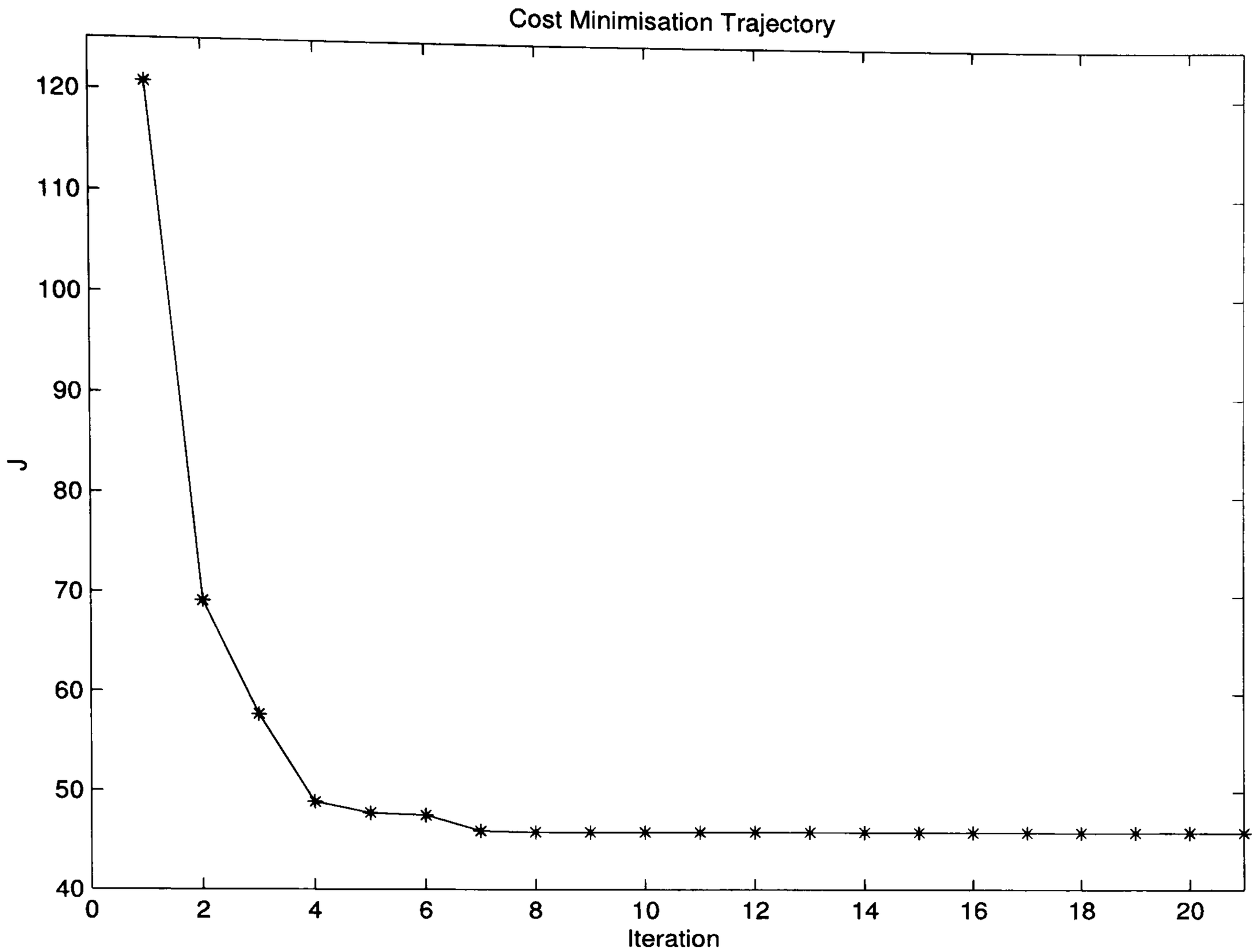

Figure 6.3: Cost minimisation trajectory.

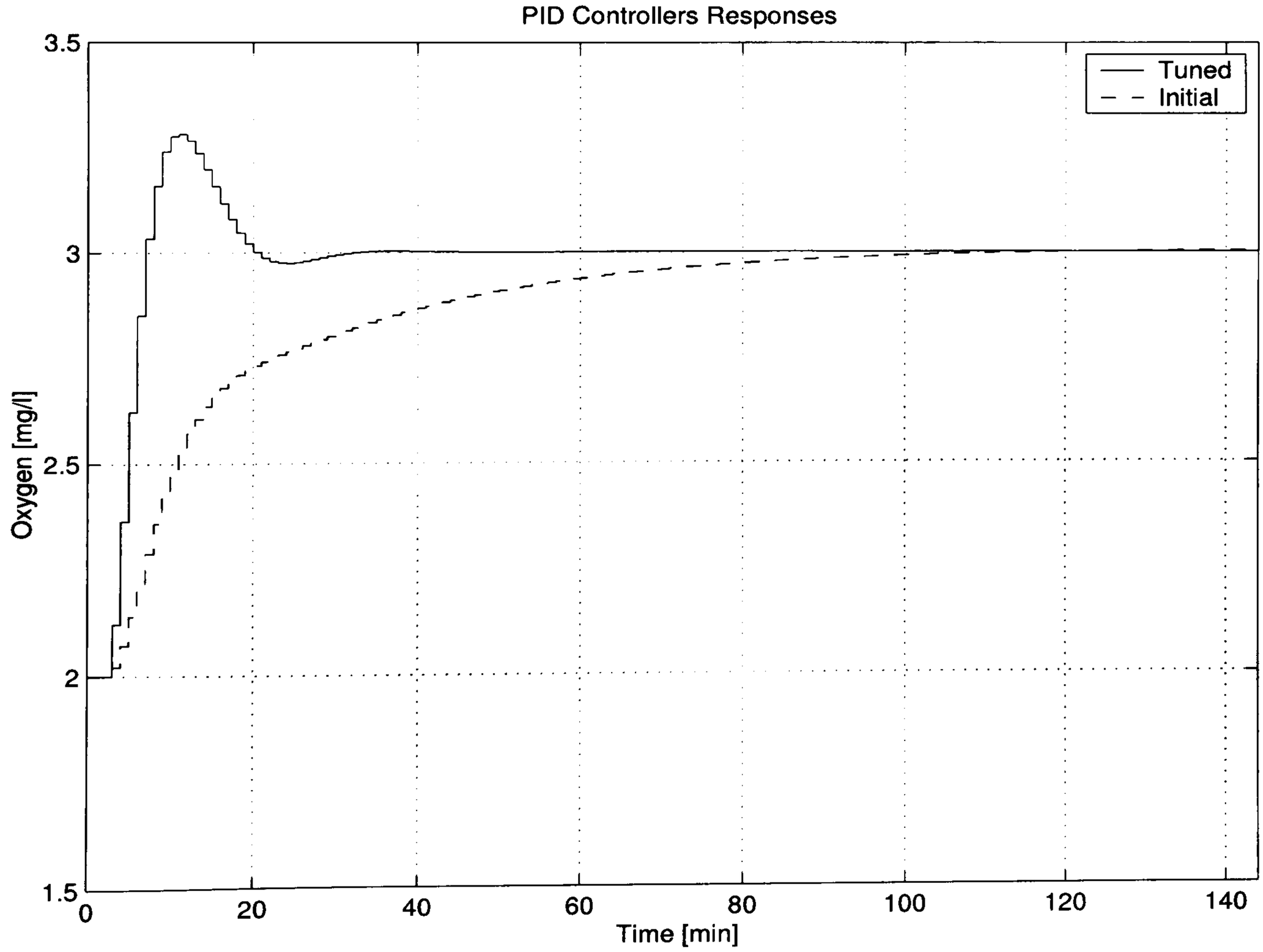

Figure 6.4: Comparison between initial PID response and tuned PID response.

6.5.2 Control of two reactors with a lower triangular controller structure

The following example illustrates the design of a controller for the last two aerated reactors (4 and 5). Therefore, input-output data from reactors 4 and 5 is required, as well as knowledge from their PID controllers. One day of input-output data has been collected at a sampling rate of 1 minute (1440 samples). The system has been excited with a pseudo random binary signal (PRBS) around the setpoint (2 ± 1 mg/l). The initial PID controllers for all three reactors have been considered to be the same, and their parameters are presented in Table 6.2.

The controller structure has been considered to be of a lower-triangular type, as presented in equation (6.81).

This structure defines an interaction between the error signal in the 4th reactor and the control signal in the 5th reactor. Simulation results have been performed for three different cases as presented in Table (6.3), where n is the order approximation in the *SVD*. Figures (6.5) and (6.6) present the simulation results.

$$\mathscr{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{6.81}$$

Table 6.3: Optimisation specifications

	$\mathscr{R}$	$\mathscr{Q}$	N	n	σ
case 1	diag(10 10)	diag(10 10)	60	5	0.01
case 2	diag(10 10)	diag(10 20)	60	5	0.01
case 3	diag(10 20)	diag(10 20)	60	5	0.01

Notice that the response in case 1 is much faster and with an acceptable overshoot compared to case 2 and 3. However, the effect of the coupling controller produces a much more aggressive response in the oxygen concentration in reactor 5.

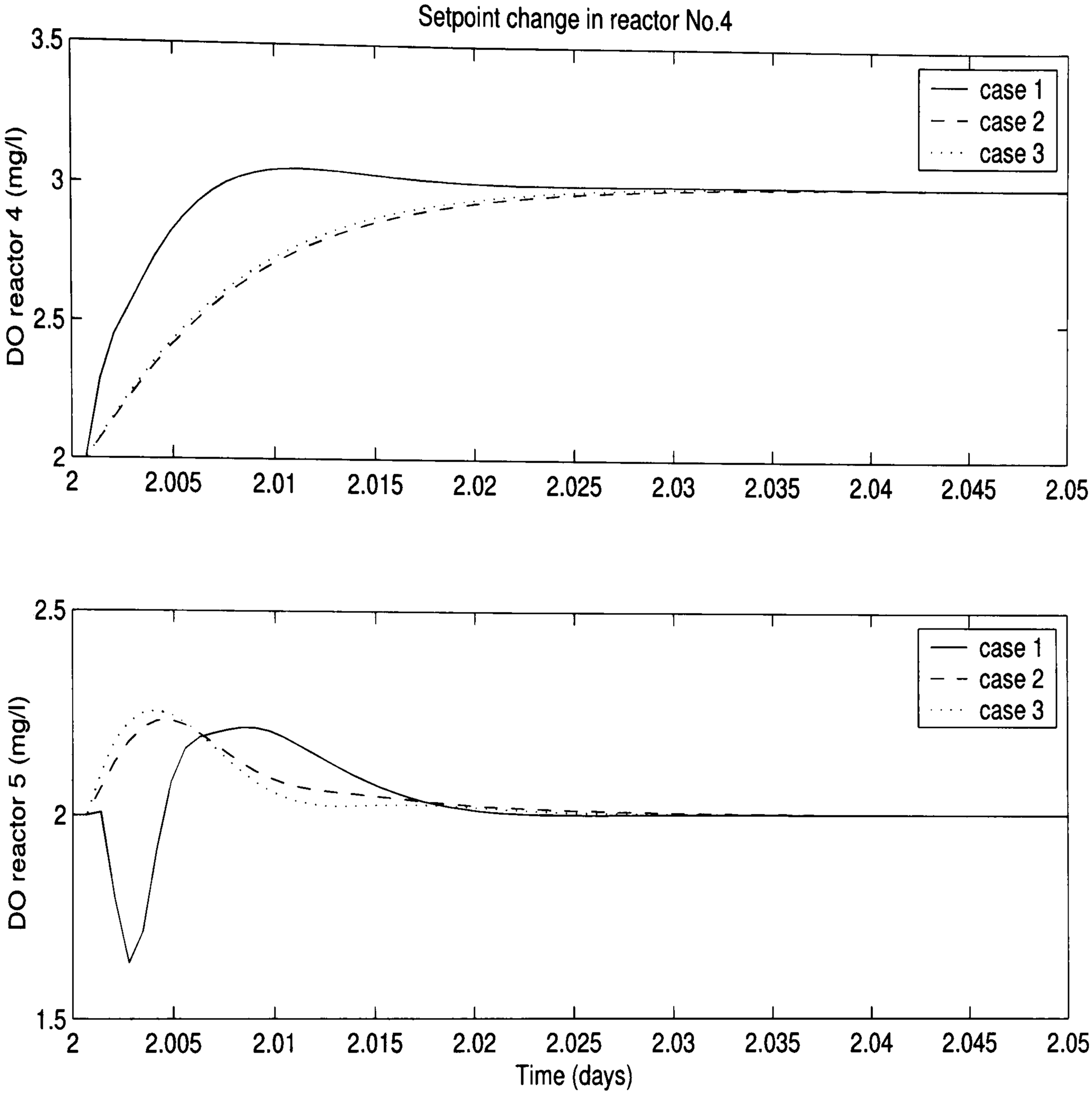

Figure 6.5: Effects of a setpoint change in reactor No.4.

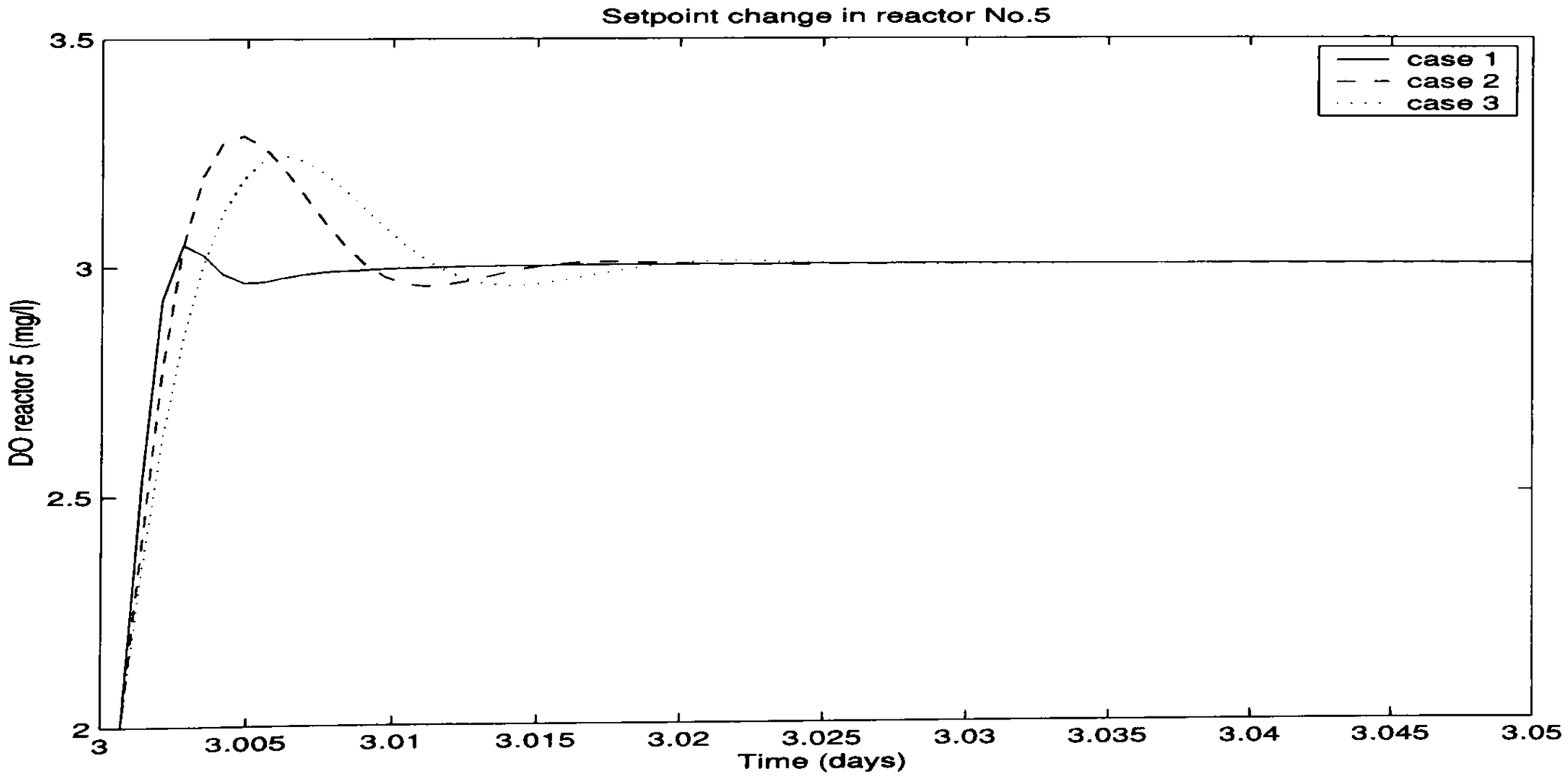

Figure 6.6: Effects of a setpoint change in reactor No.5.

6.5.3 Control of three reactors with a diagonal controller structure

This example considers the tuning of the three PID controllers with a diagonal controller structure. The same amount of data has been collected as in the previous case, and using the same procedure. The controller structure in this case is defined by equation (6.82). In this example however, simulations have been performed for a prediction horizon of 40 and 60. For each horizon, four cases of different weights are considered. Table 6.4 summarises the different cases, where the weight matrices $\mathscr{Q}$ and $\mathscr{R}$ have been chosen such that all their elements are the same. Simulation results are presented in Figures (6.7) to (6.9) for $N = 40$ and in Figures (6.10) to (6.12) for $N = 60$.

$$\mathscr{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.82}$$

Table 6.4: Optimisation specifications

	$\mathscr{R}$	$\mathscr{Q}$	n	σ
case 1	1	1	5	0.01
case 2	1	3	5	0.01
case 3	3	1	5	0.01
case 4	10	10	5	0.01

The most significant observation in this example is that there was no major difference between the simulations with different horizons. This suggests that bigger horizons do not contribute significantly to the optimality of the solution, however they will increase the computation requirements.

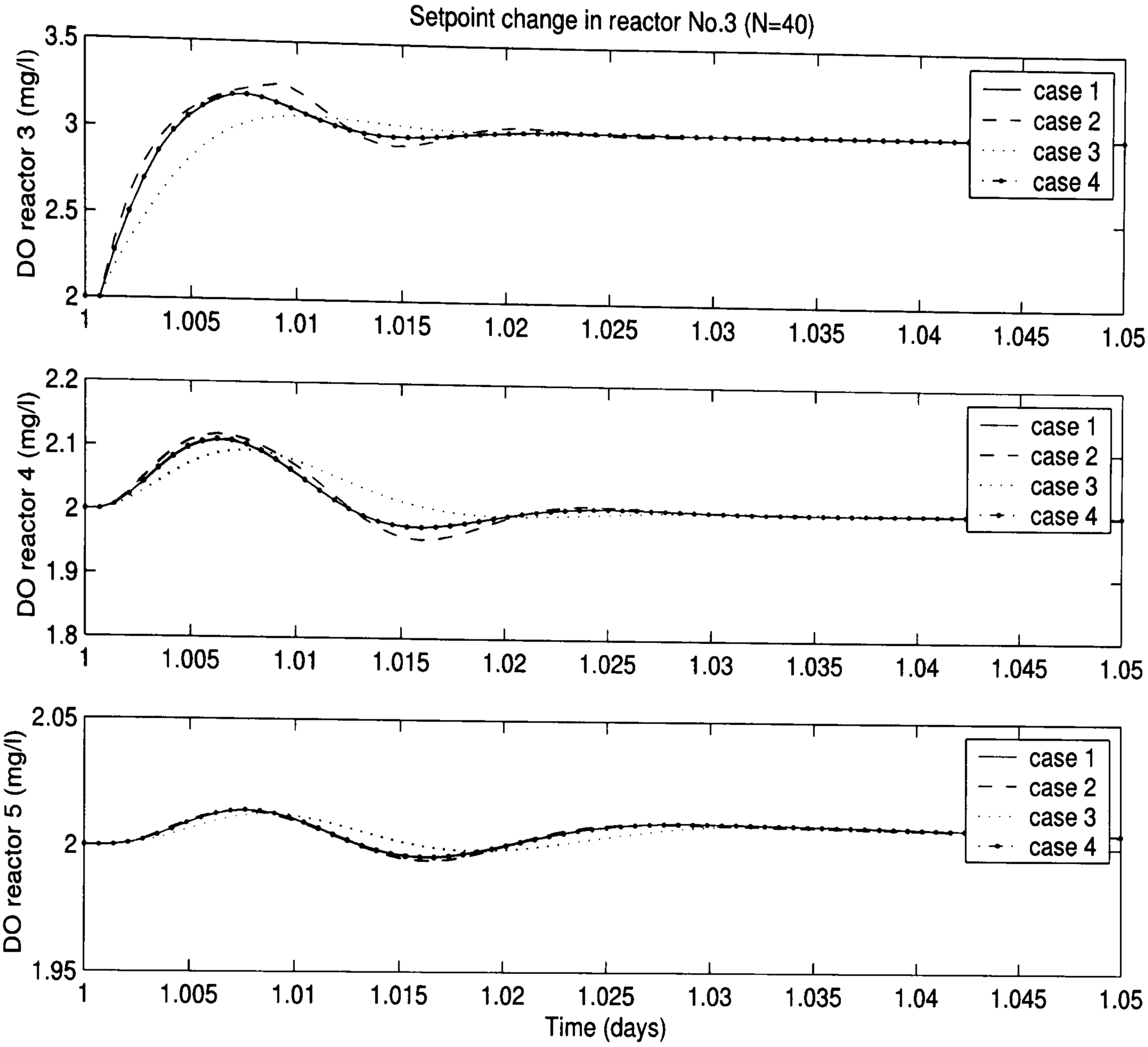

Figure 6.7: Effects of a setpoint change in reactor No.3, with N=40.

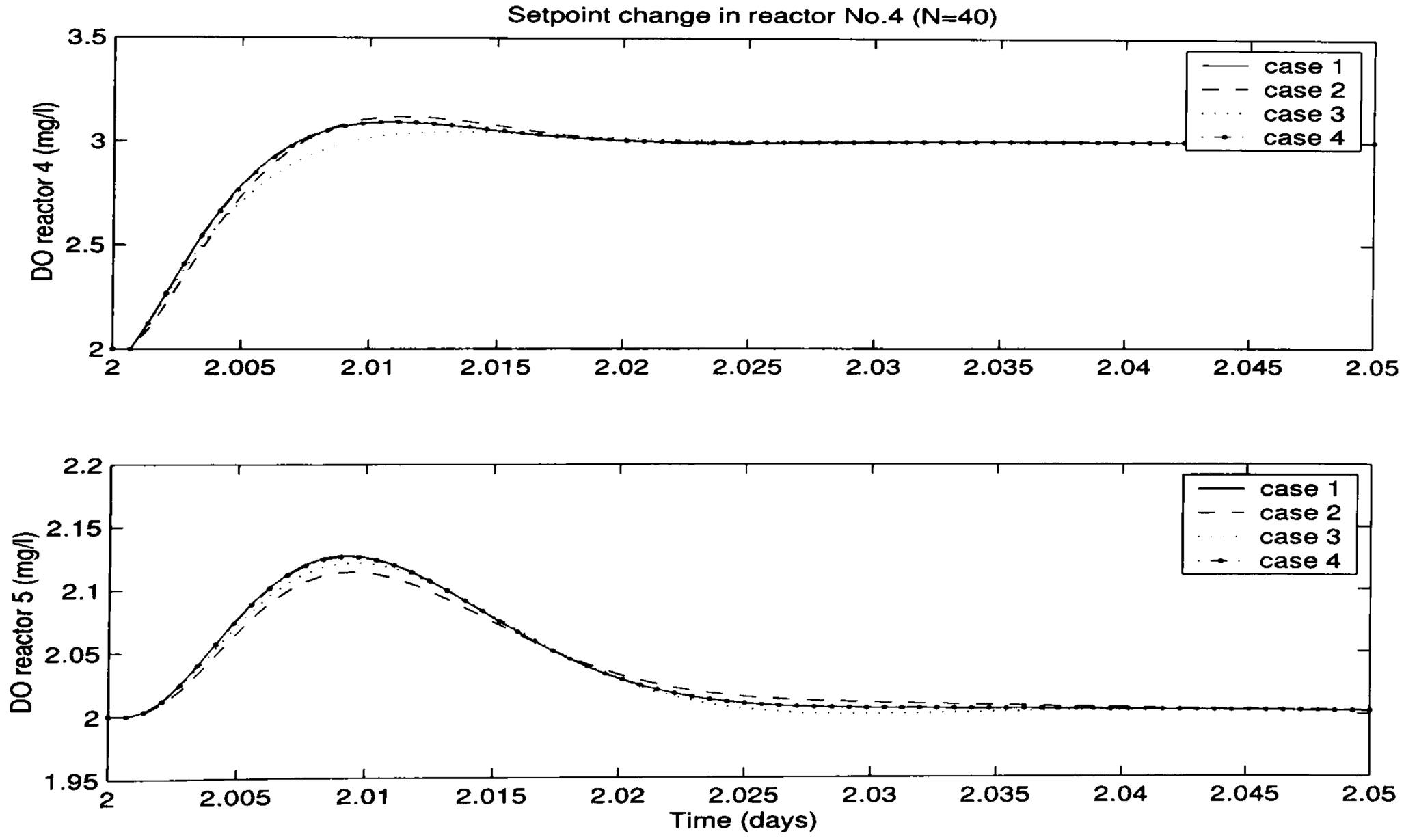

Figure 6.8: Effects of a setpoint change in reactor No.4, with N=40.

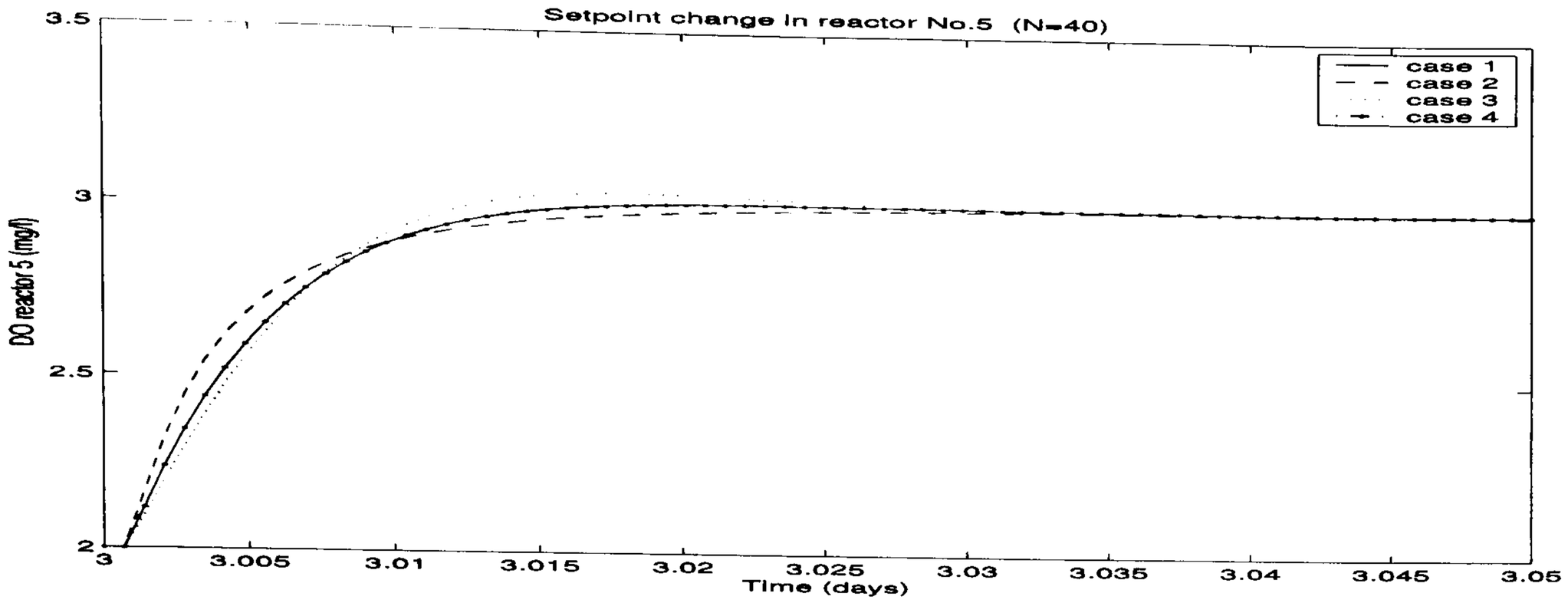

Figure 6.9: Effects of a setpoint change in reactor No.5, with N=40.

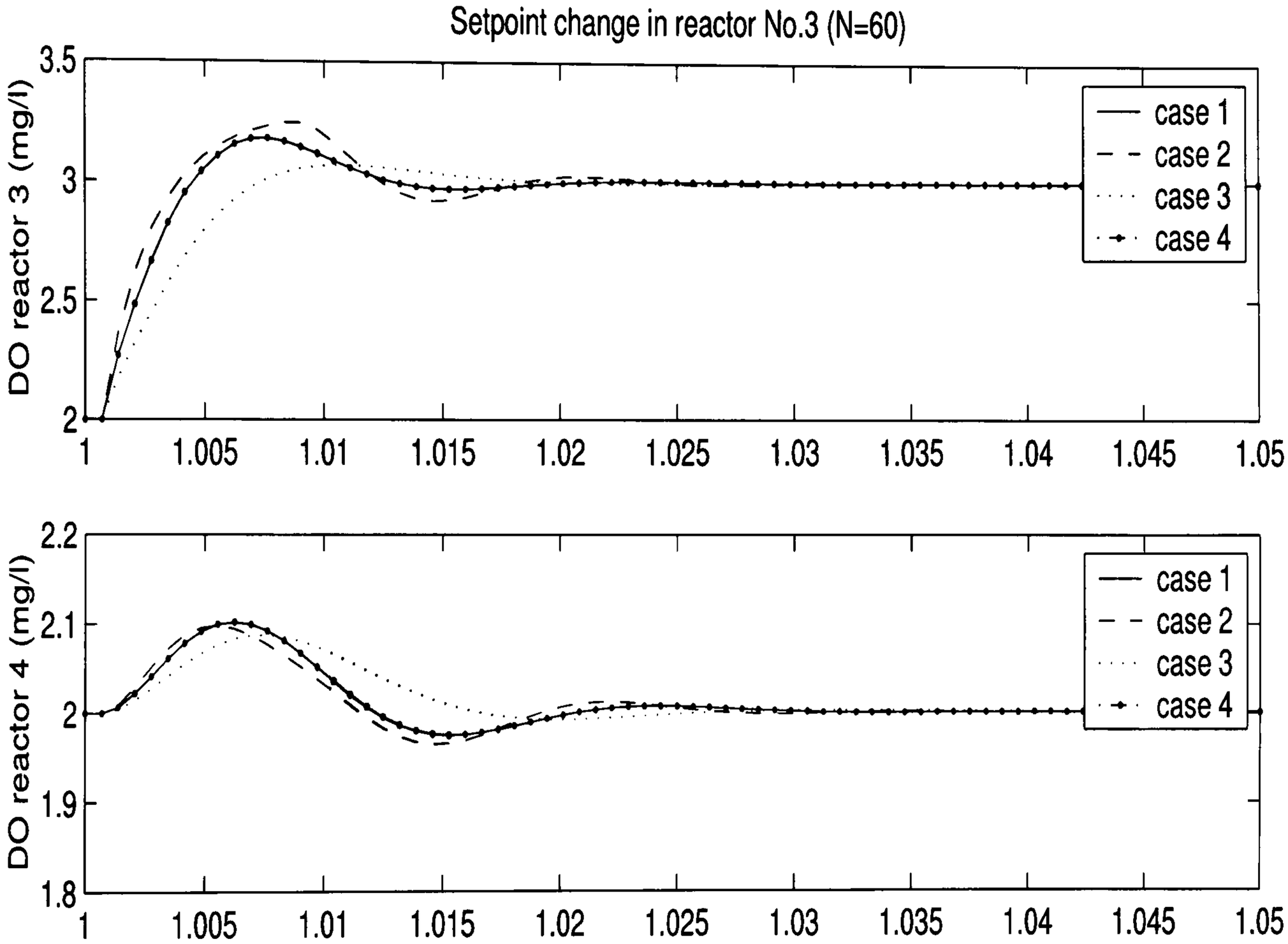

Figure 6.10: Effects of a setpoint change in reactor No.3, with N=60.

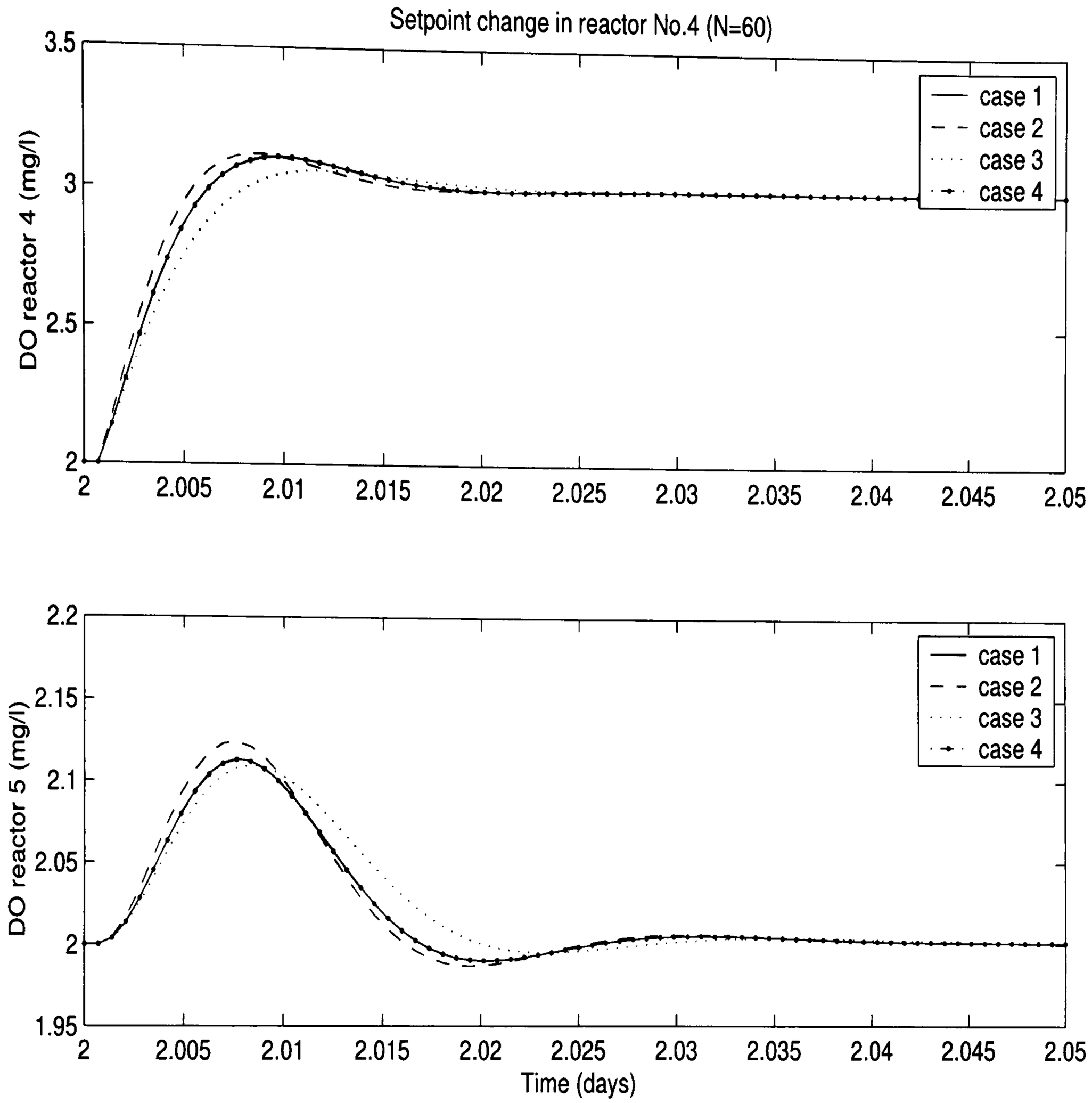

Figure 6.11: Effects of a setpoint change in reactor No.4, with N=60.

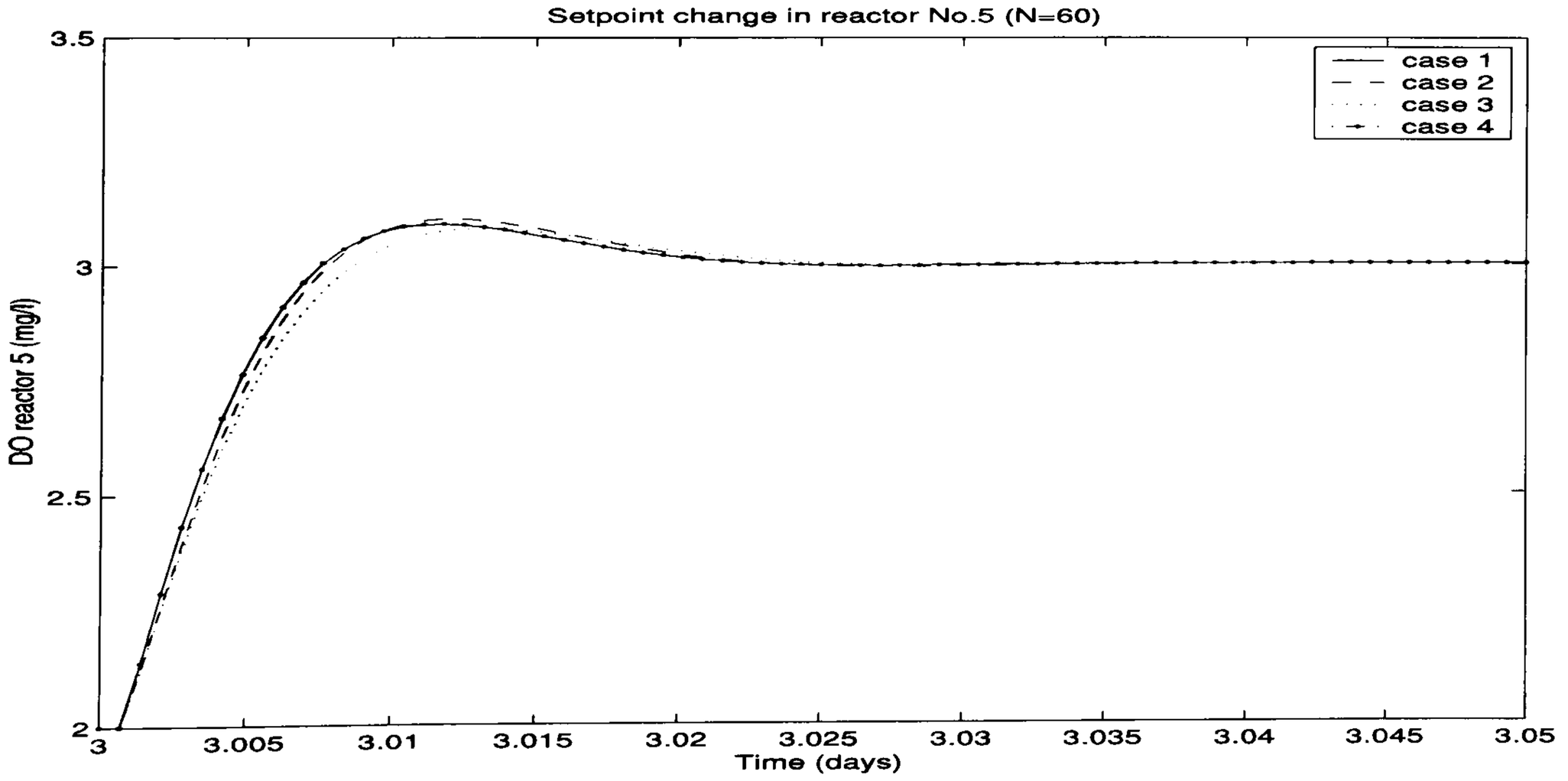

Figure 6.12: Effects of a setpoint change in reactor No.5, with N=60.

6.5.4 Control of three reactors with a lower triangular controller structure

This last example examines the design of a lower triangular controller, similar to the first example, but considering all three reactors. The controller structure is shown in equation (6.83). Two simulations for different sets of weights are presented as shown in Table 6.5. Both simulations assume a horizon of $N = 40$ and are presented in Figures (6.13), (6.14) and (6.15).

$$\mathscr{B} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{6.83}$$

Table 6.5: Optimisation specifications

	$\mathscr{R}$	$\mathscr{Q}$	n	σ
case 1	diag(3 3 3)	diag(3 3 3)	5	0.01
case 2	diag(10 20 30)	diag(10 20 30)	5	0.01

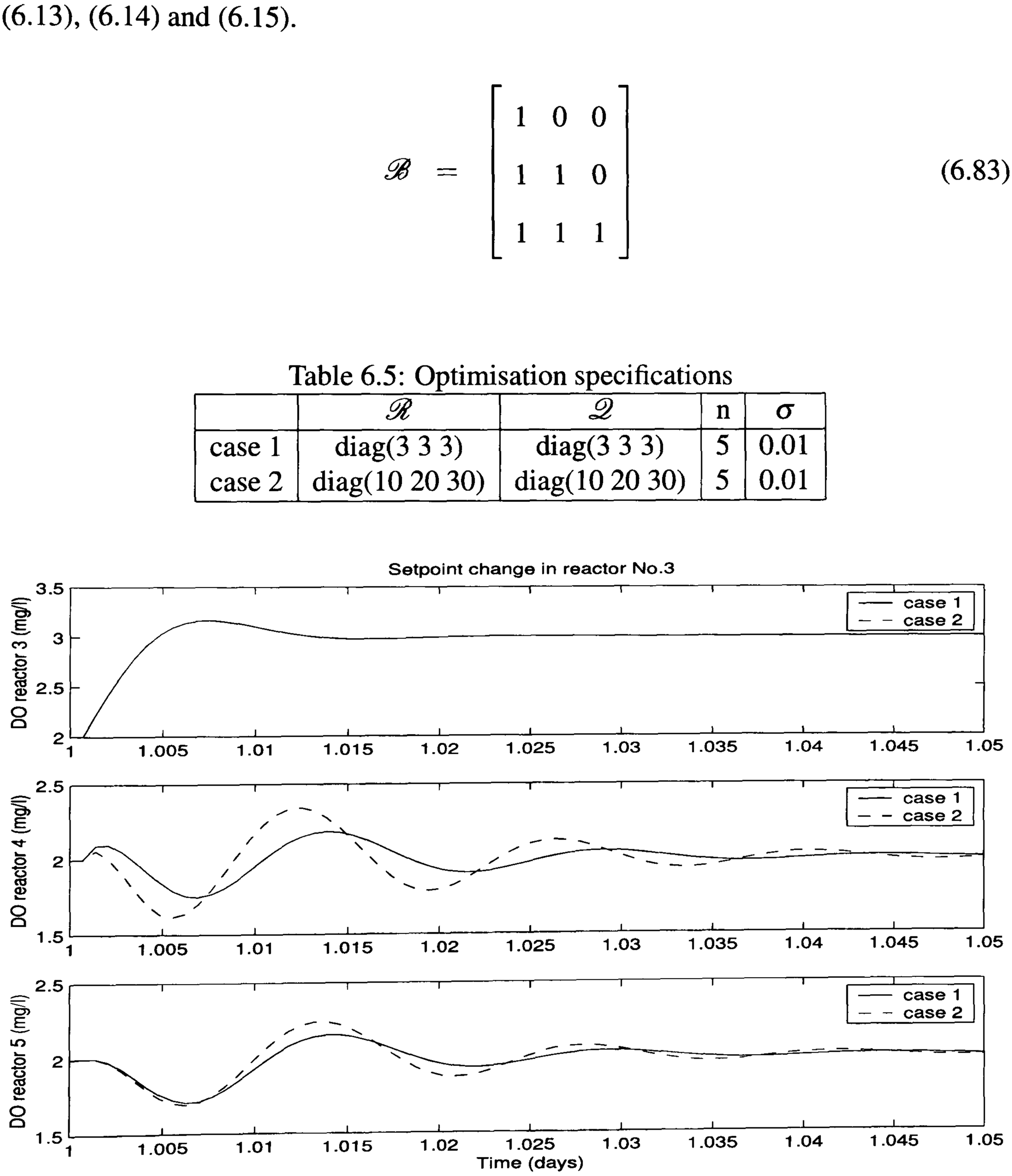

Figure 6.13: Effects of a setpoint change in reactor No.3.

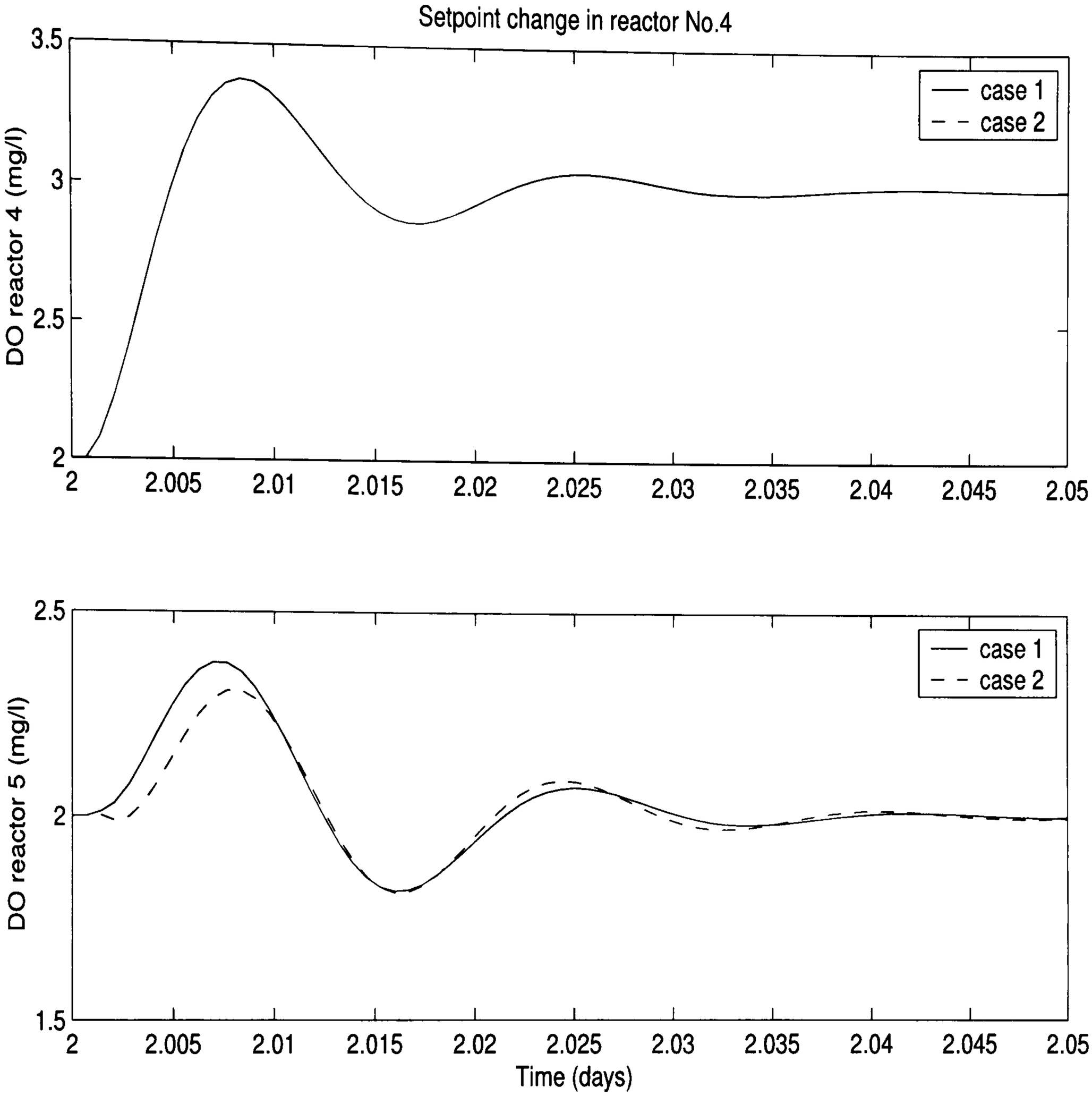

Figure 6.14: Effects of a setpoint change in reactor No.4.

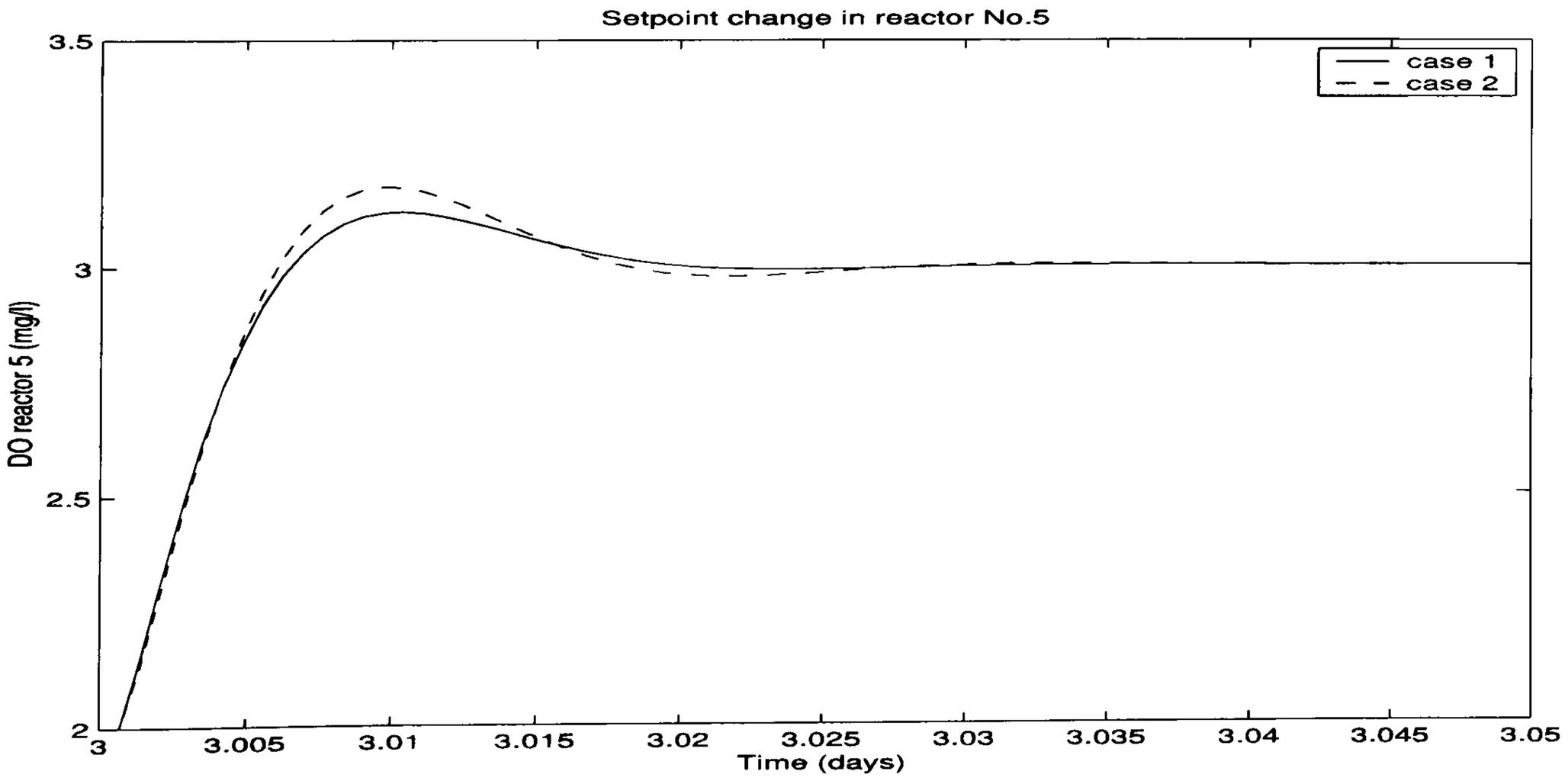

Figure 6.15: Effects of a setpoint change in reactor No.5.

6.6 Summary

This chapter presents a new algorithm which allows the design of a restricted-structure multivariable controllers within a subspace identification framework. The algorithm can be divided in two steps: identification of the plant model using closed-loop data followed by the controller parameter calculation.

The controller parameters are calculated as the result of minimising a LQG criterion over a finite forward horizon. The method ensures that the resulting control system response will be asymptotically stable over the horizon.

The chapter begins presenting the framework in which the method is developed. The chapter then presents the characterisation of the controller structure for SISO and MIMO systems. The characterisation for MIMO systems is developed as a generalisation of the SISO case.

Section 6.4 presents the main formulation of the method. In this section the algorithm to calculate the controller parameters and the optimisation is presented. Finally, several simulation case studies are presented at the end of the chapter. Results demonstrate the versatility of the method towards the design of different possible configurations.

Even though there is a high mathematical complexity in the implementation, the use of the algorithm is easy and allows a number of degrees of freedom for the individual calculation of the parameters.

The main advantage of the method, is that it allows the direct calculation the parameters from data, therefore giving a direct data-driven approach to the problem solution.

Chapter 7

Real-time control in Swinstie WWTP

In Chapter 3 simulation data from the COST benchmark and real-data from Helsingor WWTP where employed to identify models of dissolved-oxygen and nutrient removal. Later, Chapter 4 studied the design of MPC controllers for the same process variables, by using a hierarchical control structure. This control structure allowed to improve performance of the system without modifying the low-level control loops (programmed in the PLC). This performance increase however is limited by the performance of the low-level control loops; which can in general benefit from a re-tuning, if the parameters are accessible. Therefore, Chapters 5 and 6 contributed with the development of three tuning algorithms, two for multivariable systems, and one for SISO systems with the purpose of performance assessment.

All these developments however would be incomplete if left in a simulation level. Therefore, this chapter presents the development of a software platform and the implementation of identification, control and monitoring algorithms for real-time execution in Swinstie WWTP.

The development of software for industrial use has to comply with several characteristics in order to be reliable, and robust. In addition, such a software should be implemented in such a way that costs are minimised. Due to the numerical complexity of some algorithms, the use of a computationaly efficient machines should be pursued.

Almost every industrial facility employ programable logic controllers (PLCs) to control their operations and machines. PLCs are industrial computers which allow the implementation of basic control functions and with limited numerical precision. This controllers appeared as the *microcontroller era* predecessors of relay-contactor industrial control, and they usually work in a sequential-deterministic manner. Nowadays, PLCs have increased considerably their performance by using faster and robust processors and incorporating more options and modules (e.g. PID control modules, fuzzy control modules). However, the programming of any new function or algorithm is difficult, thus expensive and extreme caution should be exercised if performing it on-line.

In addition, most industries are operated by a network of PLCs commanding especific areas or processes of a plant, due to the large amount of control loops and input-outputs. Due to this reason, most industrial plants employ a supervisory control and data acquisiton system (SCADA), which collects some of the most important information into a database in a computer (or cluster of computers) and allows very limited control modification.

In the case of the wastewater industry, the introduction of SCADA systems has facilitated the control and monitoring of processes with several hundreds and even thousands of control loops. However, it is only in the last decade or so, with effluent quality standards becoming increasingly stringent, that large-scale urban wastewater treatment plants have been equipped with similar supervisory control systems. Some of the principal characteristics of these systems are the human-machine interface (HMI), which allows the operator to have a plant-wide perspective of the operation of the plant and large databases to collect important plant information such as alarms, events and process variables. However, due to security and time constraints, control actions in SCADA systems are often limited to very simple functions such as pump and valve scheduling and modifying some parameter settings as setpoints, with the more sophisticated control algorithms being in the field hardware such as PLCs or temperature controllers (Katebi *et al.*, 1999).

Time constraints appear because of the necessity of maintaining real-time operation. In most cases, communication between the high-level control and low-level hardware is slow and performed by polling due to the high channel density so direct real-time control is difficult and unreliable from this point of view. Security issues appear when the upper-level control operates over unreliable operating systems. In some cases, where security is life critical, upper-level control systems operate over industrial hardware and software, which improves reliability and reduces long term costs.

In the case of the wastewater treatment industry, many of the processes and variables are very slow and high-level security is not imperative. However, since in activated sludge wastewater treatment plants, biological processes and chemical reactions are involved, the system performance and operation is susceptible to external disturbances. Disturbances can come in the form of increased influent flow, and nutrient or chemical loading due to weather conditions or industry discharges. Erroneous handling of the plant can produce total inhibition or even death of the plant biological components and therefore halt operation for weeks or even months until sludge inventories are recovered. Therefore in these types of processes, it is very important to have plant wide information in order to assess the operation and take appropriate control actions to avoid plant mismanagement.

This chapter presents the development and implementation of advanced process control and process monitoring in real-time. Much of the chapter is concentrated in the development of the software and its interfacing with the industrial plant. The software has been developed employing two commercial software packages: Matlab with Simulink (Mathworks) and LabVIEW with the Datalogging supervisory control module plus some additional toolkits (National Instruments). Matlab has been employed to write the code of many of the algorithms and controllers. This code has been compiled into dynamic link libraries (DLLs) and included in the LabVIEW schematics. The interfaces with the PLC and the user, have been programmed in LabVIEW.

The chapter is organised in the following manner: The software-hardware architec-

ture is first presented. This section introduces the reader into some of the basics of the software and the hardware integration and the communication protocols and interfaces. Section 7.2 presents a detailed description of the available modules, their programming, and their interfacing within the software. Sections 7.3 and 7.4 present results obtained by executing the algorithms in real-time in the plant, for identification and control respectively. Finally, the chapter ends with a summary of the main achievements and conclusions from the chapter.

7.1 Software platform architecture

The wastewater process industry has been the focus of intensive research and development in the last 20 years. The state-of-the-art in sensor technology to acquire online measurements of nutrients as for example ammonia, nitrate, phosphate, and physical variables such as dissolved oxygen or suspended solids, has reached a state in which advanced data quality management and control strategies can be used to improve process performance. It is claimed in (Yuan *et al.*, 2001) that innovative process design and optimised process control is the solution to many of the problems involving nutrient removal. However, the use of sophisticated control algorithms and data quality management is still not widely used. For example, as concluded by Yuan *et al.* (2001), the use of model based on-line control is still in its infancy, and it is just in the past 15 years that on-line monitoring and data acquisition has helped to generate and verify mathematical process models which can be used for control. Data monitoring and validation is a significant area of research. Its use has helped to improve plant performance by improving plant design and operation. Many new approaches such as Aeration Tank Settling (ATS) (Nielsen *et al.*, 2000) or step-feed to increase hydraulic capacity to cope with large weather disturbances have arisen from detailed data monitoring and analysis.

Research in integrated wastewater treatment systems (sewer and treatment plant) are

also now of great interest (Nielsen *et al.*, 1996), and data monitoring plays a very important role, since by analysing plant data it is possible, for example, to identify influent flow models (Nielsen, 2002, SMAC project report). Plant monitoring can also be used for knowledge extraction, which allows the assessment of plant operation by identifying deviations from the normal operation conditions (Nielsen, 2002, SMAC project report).

In this context, a supervisory control system for a wastewater treatment plant would not only allow the collection of data and its analysis but also allow the use of the acquired knowledge to improve plant performance. It would be expected that under this scheme it will be possible to implement more sophisticated control algorithms and not only be limited to conventional SCADA control functions. The platform introduced in this chapter provides an alternative solution to the lack of versatility for computational requirements in existing SCADA systems. The platform has tools that can be interfaced with an existing SCADA or used independently, since for security reasons a plant can usually operate without the SCADA. This solution can only be implemented when the plant time constants of the specific loops are larger than the polling time of the existing SCADA as is the case of the activated sludge wastewater treatment process. The platform is conceived to work in parallel with the existing SCADA. This might however not be always possible, depending on the functionalities of the existing supervisory system.

The platform architecture comprises four units, the plant wide process monitoring unit, the identification unit, the process control unit and the database; these are shown in Figure (7.1). Each of these units is supported by a set of mathematical tools to assist in the implementation of control algorithms, data analysis and plant operation.

A direct link with the PLC uses an OLE for Process Control Server (OPC). OPC servers are commercially available and are usually supplied by the PLC manufacturer.

The following section gives a more detailed explanation of the software development

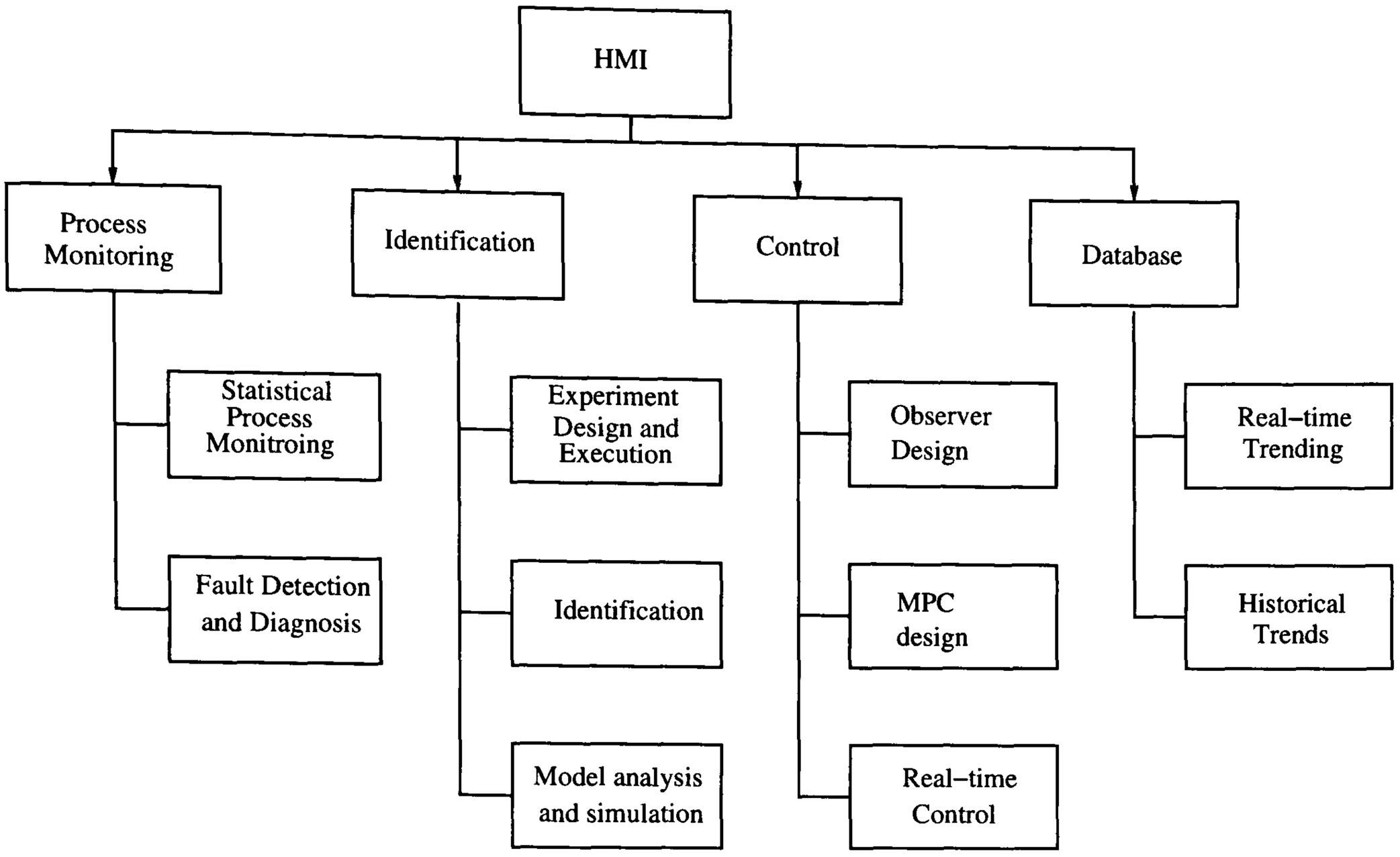

Figure 7.1: Control platform architecture

and interfaces as well as the tools that have been incorporated and are to be tested in the plant.

7.2 Software platform implementation

Due to the large number of commercially available SCADA systems in the market, it is nowadays very difficult to implement a system that is compatible with purchased proprietary software. The main problem within the architecture of the system is to interface in an adequate manner with the existing software and hardware.

Some SCADAs are designed in a closed way that only provides communication protocols with software from the same manufacturer. Even if the manufacturer provides some type of well-known industry standard communication protocol, the allowed functions are usually very limited.

The software developed employs OPC technology to interface with the PLC. An OPC server allows direct access to the PLC registries and all events, alarms, and variables

Table 7.1: Lab-VIEW and MATLAB interfacing technologies

Algorithm	Technology	Unit
RPCA	MATLAB script node (Activex)	Process monitoring
PRBS	Simulation Interface Toolkit (DLL)	Identification
Subspace Id.	MATLAB Compiler (DLL)	Identification
Observer design	MATLAB script node (Activex)	Control
MPC design	MATLAB compiler (DLL)	Control

are then directly accessible.

The software is programmed over LabVIEW and makes use of the Datalogging and Supervisory Control (DSC) module to interface with the OPC server. Some of the algorithms have been programmed over MATLAB and compiled into C shared libraries (DLLs) using the MATLAB compiler and some additional C code to interface with common C standards and not particular to MATLAB. All of the routines have been tested in simulation exercises to ensure their correct execution. Appendix A contains a *HOW-TO* guide for the compilation of MATLAB functions into DLLs to be used in LabVIEW.

Other modules have been implemented in SIMULINK and integrated into LabVIEW employing the Simulation Interface toolkit. Table (7.1), gives a summary of the technologies employed to implement different algorithms from MATLAB into LabVIEW.

The following sections give a detailed description of the system interfacing and a discussion on the implementation of each unit.

7.2.1 System interfacing and database population

The interfacing of the system with the PLC or PLC network is perfomed through a communication protocol known as OPC. OPC stands for OLE for Process Control, and OLE for Object Linking and Embedding. OLE is a Microsoft component which allows the automation of certain processes and applications like communication servers. OPC was designed for industrial application purposes.

The software platform interfaces with PLC by registering with an OPC server which handles the communication flow. Each type of PLC requires a particular OPC server, which is usually provided by the manufaturer. In the case of Swinstie WWTP, the PLC is TI 565 which communicates using a Tiway protocol. Therefore the selected OPC server should support this protocol, and the communication interface, which in the case of Swinstie is a RS-232. If a network were present, then the interface would be through a ethernet link in most cases.

With the OPC server correctly configured and running, the next step is to populate the database. PLCs contain memory registers from where they read or save information which is to be used by the internal program. In general, each input-output of the PLC is also mapped into a memory location (or register). The access to these memory locations is inmmediate using the OPC server, however, it is important to map it into the database, so the information read or written to the register is mirrored in the client. LabVIEW and the DSC module employ a tag engine concept. A tag is a variable associated to a register, however it is already located in the client and not in the PLC, therefore providing faster access. The tag engine assures that the register is accessed at the required sampling times or when the program requires. This helps the system to have a deterministic access to the variables of the process, and therefore perform operations in real-time.

7.2.2 Process monitoring unit

This unit monitors the process state in real-time using on-line measurements and historical information stored in the database. By monitoring the state of the plant in real-time it is possible to detect faults and even provide a diagnosis. Process monitoring for fault detection and diagnosis is a complex area of study with continuous developments, see for example Wade (2004).

The unit retrieves data from the database and on-line measurements, and processes it

with the purpose of extracting knowledge of the process behaviour and plant operation. Process knowledge extraction accumulates information from expert knowledge of the plant (plant operator), on-line data and historical data, to find possible ways of globally optimising the plant by a better scheduling of actuators or by enhancing set-points by identifying current or future operating states. It also provides means by which abnormal plant conditions can be predicted and alarms raised. Therefore this unit comprises software algorithms to perform the following tasks:

1. recursive statistical process monitoring

2. Fault detection and diagnosis

Recursive statistical process monitoring employs a recursive formulation of principal component analysis (PCA). PCA is a multivariate statistical tool, which allows the comprension of data and examination of its statistical properties, employing geometrical tools. The recursive principal component method employed in the software is the algorithm proposed by Li *et al.* (2000) for the recursive update of the correlation matrix R_k at each sampling time. The algorithm is summarised in Appendix B.

Using the recursive update of the correlation matrix it is possible to perform the following:

1. Statistics monitoring, by calculating the Q and Hotteling's T^2 statistics of the compressed data set (Chiang *et al.*, 2001).

2. Fault detection, by comparing the statistics with limits within a 95 and 99 percentile confidence interval.

3. Diagnosis, by observing the loads and scores from the data set decomposition.

The recursive update of the correlation matrix has been implemented in Matlab and is passed into LabVIEW by using a script node. Figure (7.2), shows a block diagram of the execution of the algorithm.

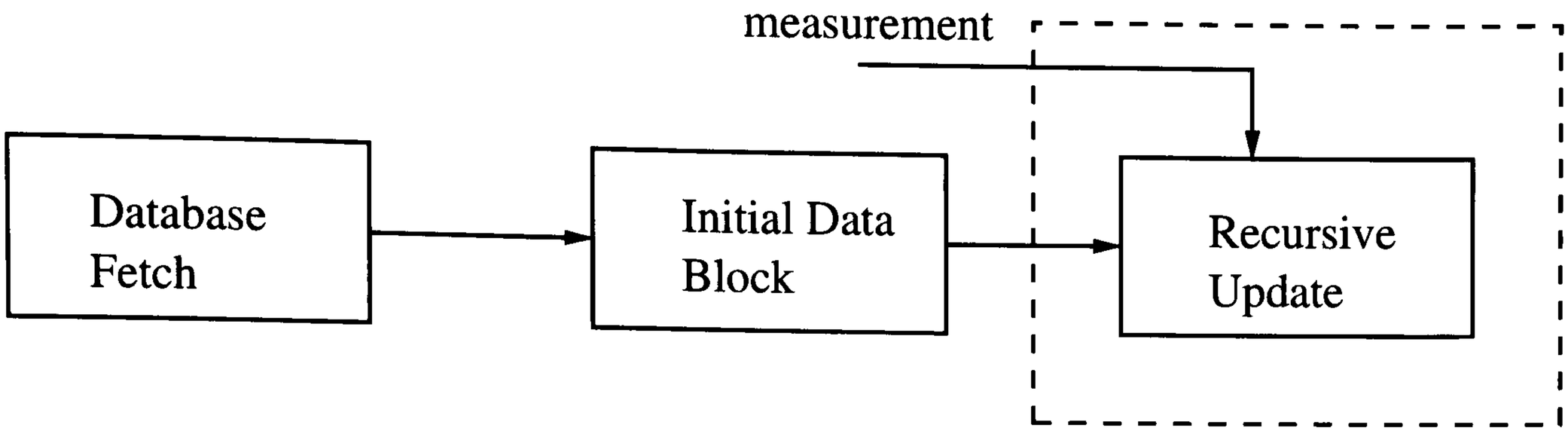

Figure 7.2: RPCA execution.

The first step is to feed data into an initial data block set, by looking into the database and selecting appropriate data. This data must representative of the process operation and must not contain important failures in the process operation. Once the variables and the data have been retrieved and fed into the initial data-block set, the recursive update of the correlation matrix begins.

Once the correlation matrix R_k and a normalised data block matrix X_k are calculated, they are used in the fault diagnosis module to identify the system state and detect where and what type of fault is occuring.

Fault detection is performed by calculating Hotteling's T^2 statistic and the Q statistic using the normalised data matrix X_k and the correlation matrix R_k (see (Chiang *et al.*, 2001) for the definition of the statistics). The calculation of these statistics is perfomed by decomposing the correlation matrix. Two decomposition methods have been implemented: (a) singular value decomposition (SVD) (b) Lanczos tridiagonalisation.

The choice of the decomposition matrix depends on the amount of data and type of process. Lanczos tridiagonalisation allows the the SVD decomposition in large sparse data matrices. The reason for implementing both methods is to investigate their suitability for the activated sludge process. The Lanczos algorithm has been implemented using a freeware toolbox for Matlab called LANPRO ver. 1.0 available at (http://soi.stanford.edu/~rmunk/PROPACK/index.html).

The detection of faults is performed by comparing the value of the statistics with a confidence interval of 95 and 99 percentile. The decomposition of the correlation

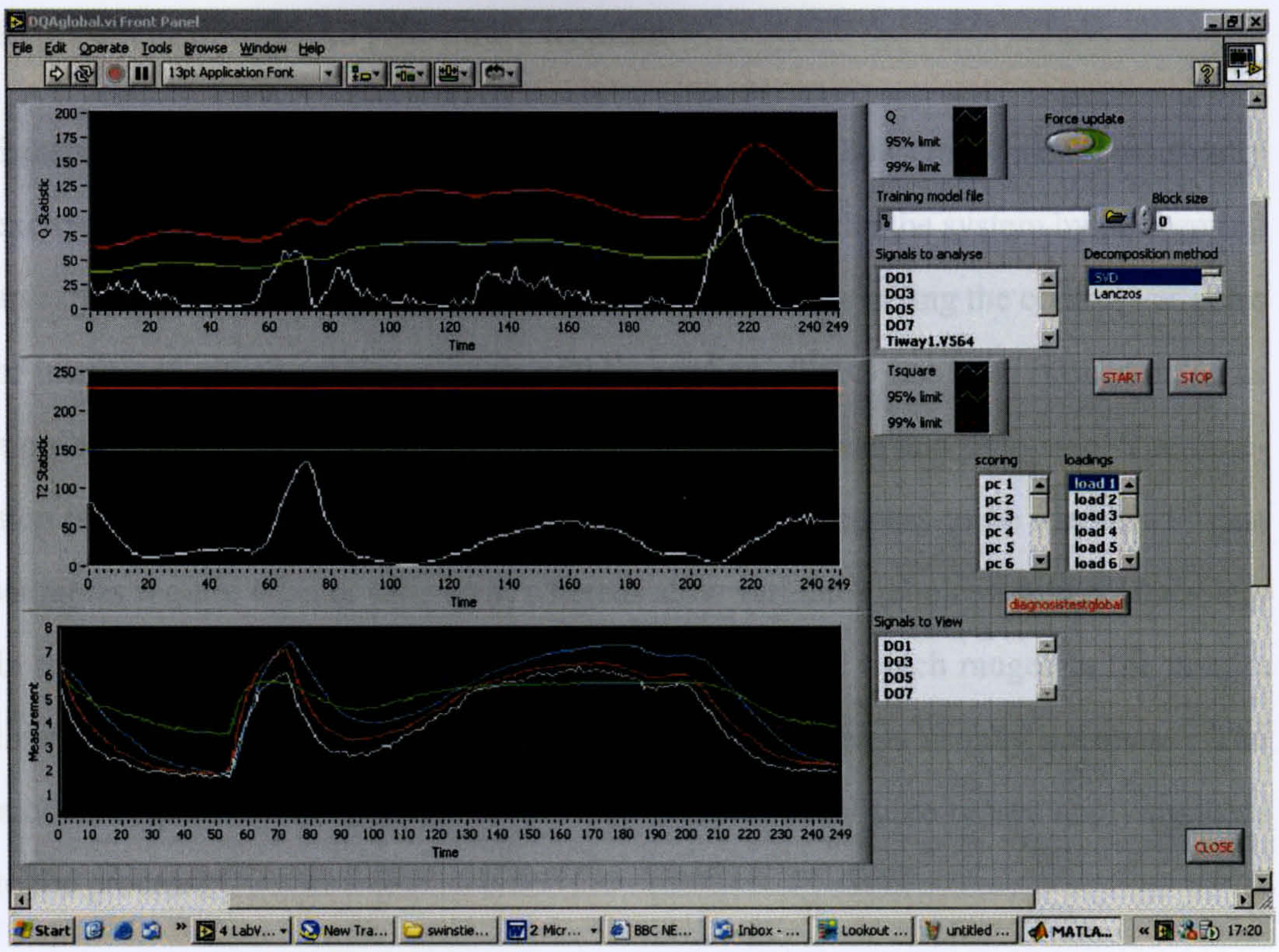

Figure 7.3: User interface of the data quality assessment unit.

matrix also provides the loads and scores. The examination of these vectors allows to recover information regarding which measurement in the process is contributing more extensively to the fault. Even more, by using more advanced algorithms not implemented in this software, it is possible to identify the type of fault.

Finally, Figure (7.3) gives a view of the graphical user frontend of the unit.

7.2.3 Identification unit

The identification unit allows to perform the following functions,

1. Experiment design and real-time execution

2. Subspace identification

3. Model analysis and simulation

Experiment design and real-time execution

Subspace identification requires that the data employed shows the system persistently excited. In order to obtain such data, it is necessary to excite the system by performing an experiment. The design of the experiment consists in selecting the control variables of interest to which probing signals will be applied. PRBS has been implemented as a probing signal, which can be modified in amplitude, mean and time-step. The time-step is selected according to the response time of the process to be identified. Fast variables require a faster time-step, whereas slow variables require a higher time-step. In the activated sludge process, oxygen is a fast variable which ranges in the minute time horizon, therefore a time step of about 2-3 minutes is usually adequate. The module also allows to simulate the probing signals in real-time before applying them to the process.

With the probing signals ready, the experiment can be run in real-time. The software database will automatically record all the signals (inputs and outputs). Once sufficient data has been collected, the data is exported from the database into a normal text file. This is not an automatic process, since there might be the need to select only a subset of the recorded data. Usually these experiments can take up to 15 hours in fast variables as oxygen, time during which many unexpected events can occur. Therefore the need of manually selecting appropriate data.

Subspace identification

Two subspace identification methods have been implemented: (a) robust N4SID 'SV' (b) robust N4SID 'CVA'. Both algorithms have been programmed in MATLAB and compiled into C shared libraries (DLLs). The process of identifying a model is performed by following the next steps:

1. Select data file

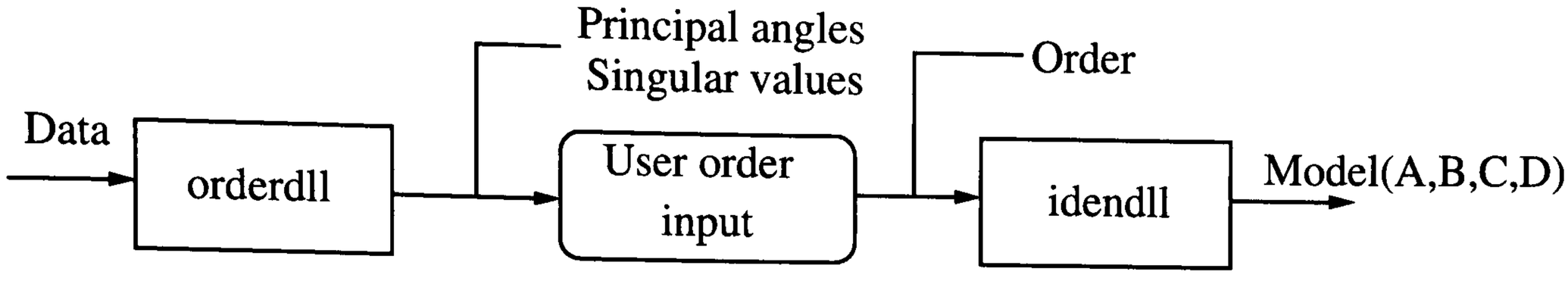

Figure 7.4: Subspace identification implementation

2. Choose subspace algorithm: SV or CVA

3. Select number of block-rows in Hankel matrices

4. Perform data decomposition

5. Estimate system order

6. Identify model

The original implementation code of the two subspace identification algorithms is from the toolbox provided in (van Overschee and De Moor, 1996*b*) and freely available in the following ftp server:

ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/vanoverschee/book/subfun/

Some modifications to the code have been necessary in order to correctly interface the algorithm into LabVIEW. The core of the implementation of this module consists of two software blocks executed in sequence, as presented in Figure (7.4).

The blocks orderdll.dll and idendll.dll are dynamic link libraries compiled from MATLAB. The first library receives data read from the text file, and performs the decomposition step necessary in subspace identification algorithms. The library output are the singular values if the selected algorithm is SV, and the principal angles if the selected algorithm is CVA. The LabVIEW implementation then gives a graphical representation of the singular values or the principal angles. The user is then able to estimate the order of the model to identify. This is passed into the second dll which calculates the model.

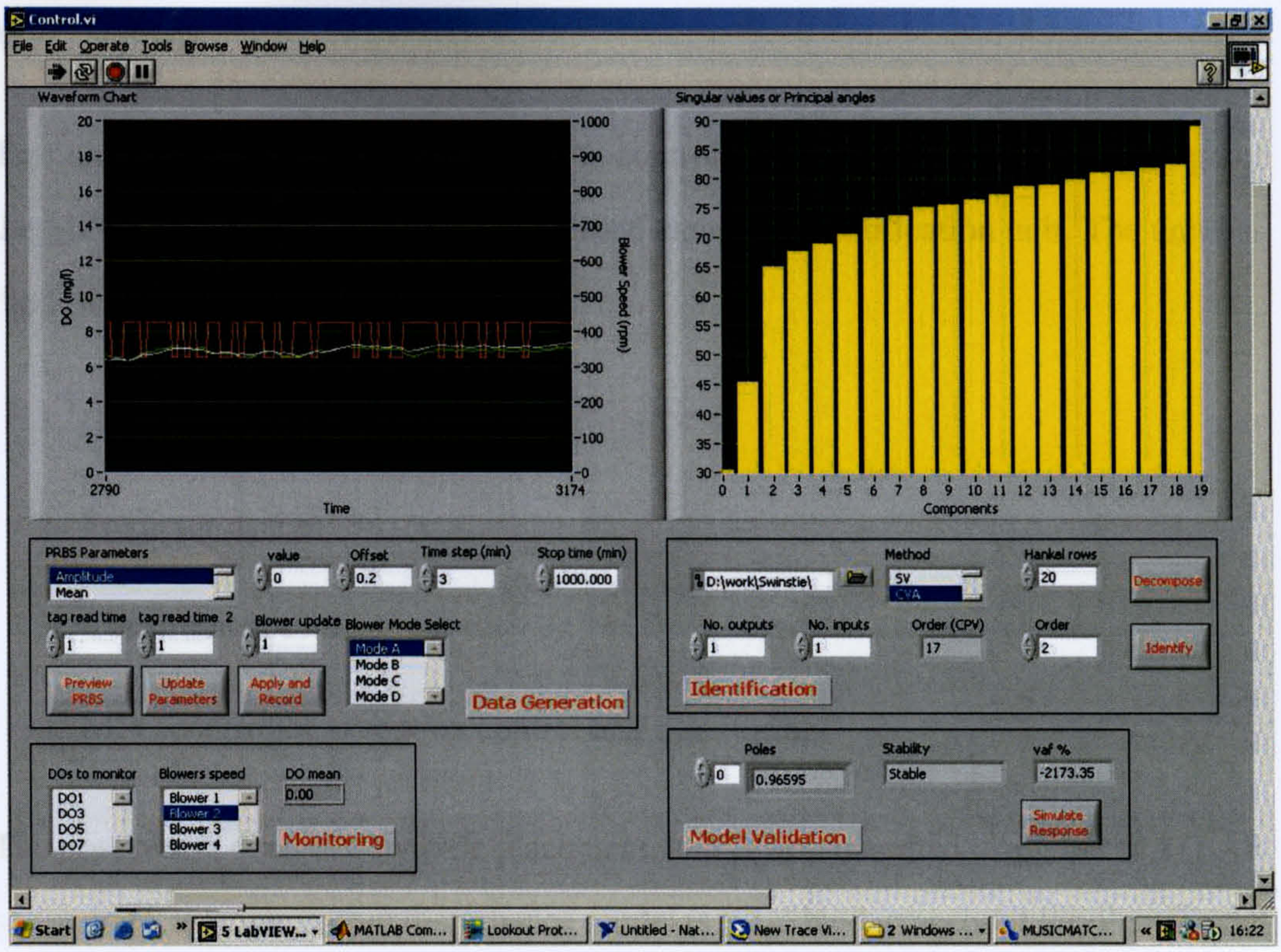

Figure 7.5: User interface of the identification screen.

Model analysis and simulation

Once the order is selected and the model identified, the module will give an indication concerning the stability. The three possible indications are: (a) stable, (b) unstable (c) marginally stable.

The model simulation is performed using the same data used for the identification. The response of the model is presented in a graph which also includes the input and output signals used in the identification. Also, the *vaf* coefficient is calculated to give an indication of the quality of the model. These functions are totally implemented using LabVIEW function blocks. Figure (7.5) presents the graphical user frontend of this unit.

MPC control

This unit implements a hierarchical MPC controller. The unit comprises the design of the MPC controller given a model identified using the identification unit. The unit has the following modules:

1. Observer design
2. Predictor design
3. Constraint specifications
4. MPC controller execution control and monitoring

The observer is designed by pole placement, and is implemented using the MATLAB script node. Also an output disturbance model is employed to account for plant-model mismatch. This method has been described in Chapter 3.

The predictor module calculates the matrices Ψ, Γ, and Θ of the standard MPC formulation presented in Chapter 4. The module also allows the inclusion of constraints if required. This modules is implemented using a DLL compiled from MATLAB.

Once the observer and the predictor have been designed, the unit allows the real-time exection of the controller in the plant. During the real-time execution the system allows the fine tuning of the MPC controller, by modifying horizons, and weights.

The numerical implementation of the on-line optimisation employs a least-square approach as formulated in Chapter 4.

Figure (7.6) shows the user graphical frontend of the MPC unit.

7.2.4 The HMI unit

Even though, in itself the HMI is not something really new, the original contribution made by this reported research is to synthesise all the complex information from the

Figure 7.6: User interface of the MPC controller.

data assessment and control units in a way that can be readily understood by wastewater technical staff; staff who may not be control engineering experts. Figure (7.7) shows a schematic from a HMI designed specifically for the test plant at Scottish Water.

7.3 Real-time identification of dissolved oxygen

This section shows results obtained by the real-time execution of the identification module. Some of the results obtained demonstrate that dissolved-oxygen can be modelled by low-order models, and that this model can be used in a model predictive controller.

Several identification runs where performed at Swinstie, by exciting the plant with a PRBS signal to identify a closed-loop model. To do so, several limitations had to be overcome to make the experiments possible. First of all, the aeration capacity in

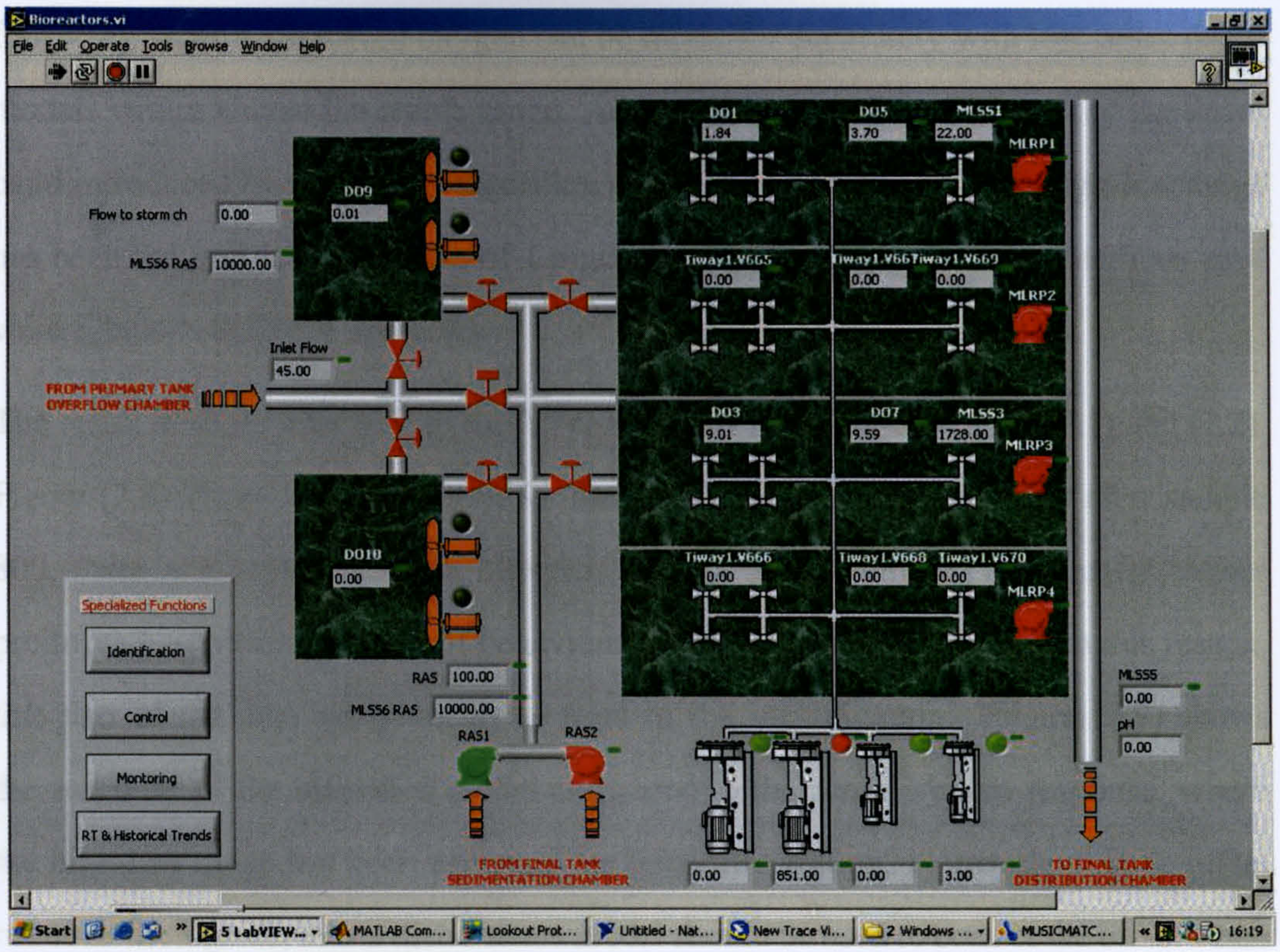

Figure 7.7: Swinstie plant HMI screen.

Swinstie is over designed. The aeration system is composed by 4 blowers: two of low capacity and two of high capacity. The plant normally requires a small blower to pump sufficient air into the system. However, since the plant is configured to operate only at half its design load, even the small blower at its slowest speed over-aerates the plant. The effect of this over-aeration is that the dissolved oxygen levels are almost always over the specified range of 1.8 - 2.2 mg/l.

A second limitation in the blower actuator system has to do with the blower update speed. The small blowers update their speed (increase or decrease) at a rate of 10 minutes, and the big blowers at a rate of 1 minute.

Due to these two limitations, the only possible solution, which did not demand the high cost of PLC re-programming, was to perform all the experiments at a high dissolved oxygen level using a big blower. This approach, however, does not invalidate the results, since as proposed in chapters before and as will be demonstrated in the

following section, dissolved oxygen can be modelled accurately with low order linear models within almost the whole range. Additionally, to reduce the effect of the dead-band introduced by the control algorithm programmed in the PLC, the high-low range has been reduced to a minimum of 1 mg/l. The subspace identification methods used were robust-N4SID SV and CVA.

The result of these tests lead to the 2^{nd}order linear model of equations (7.1) and (7.2). Figure (7.8) shows the trends used for the identification. Notice that just after sample 800, there is a switch between blowers, and the update time of the smaller blower produces a significant change of behaviour in the system response. Due to this reason, this part of the data set is not to be used in the identification. Figure (7.9) shows the response of the identified model compared to the actual system response, where the high-low range has been averaged for better illustration. Notice there is an offset between the signals, thus a disturbance model should be used to correct the offset when designing the predictor.

$$x(k+1) = \begin{bmatrix} 0.97788 & -0.04357 \\ 0.03658 & 0.84292 \end{bmatrix} x(k) + \begin{bmatrix} -0.04629 \\ 0.1124 \end{bmatrix} r(k) \qquad (7.1)$$

$$y(k) = \begin{bmatrix} -0.38536 & -0.18981 \end{bmatrix} x(k) + [-0.00067]\, r(k) \qquad (7.2)$$

7.4 MPC control of dissolved oxygen in Swinstie WWTP

The design and evaluation of a MPC controller for dissolved oxygen in Swinstie WWTP is presented in this section. The MPC controller is designed using the model identified in the previous section. The controller uses the state-space formulation reviewed in Chapter 4. The obtained results show a degree of improvement compared with the controller implemented in the plant PLC. The section is organised in such a way that

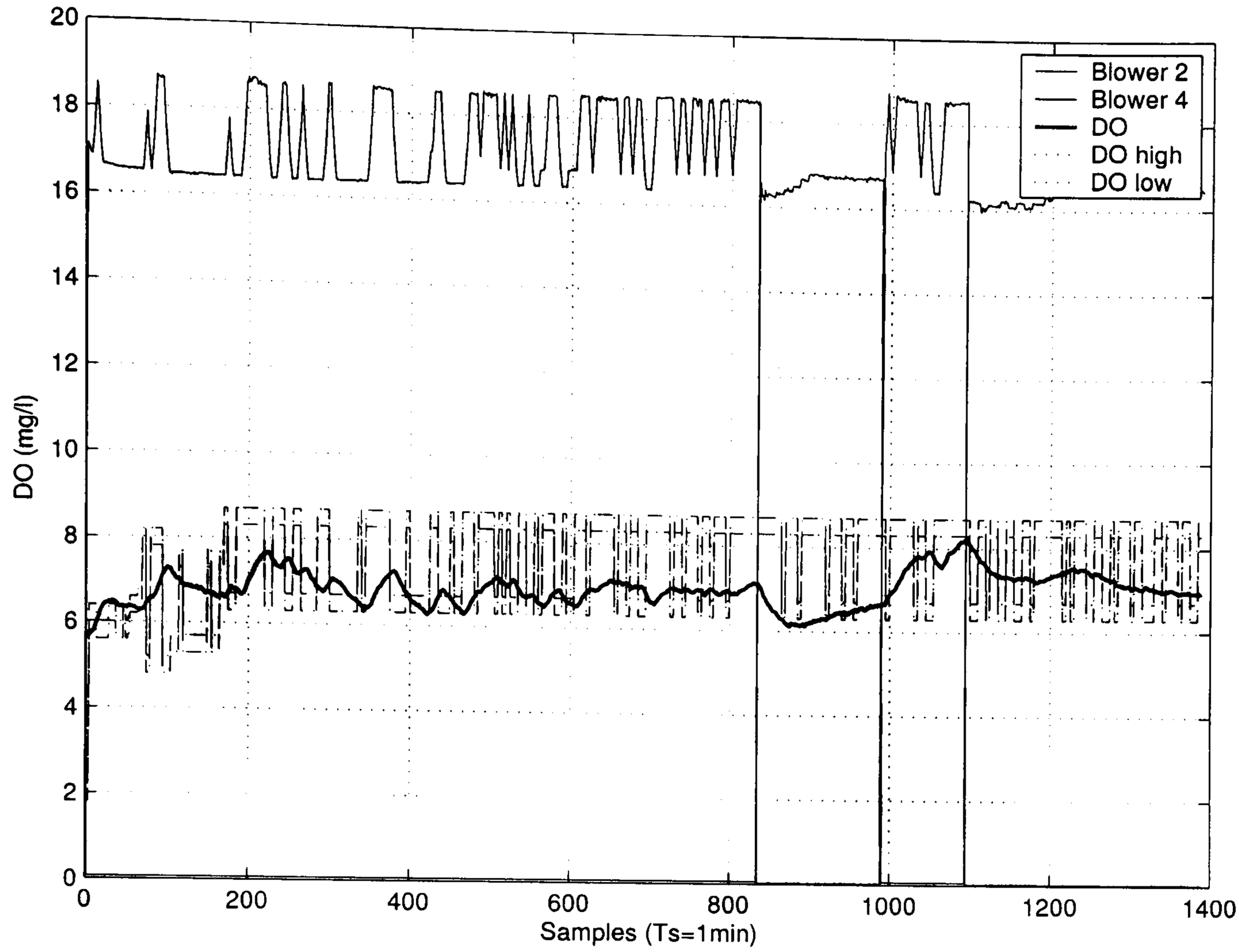

Figure 7.8: Identification experiment in Swinstie WWTP.

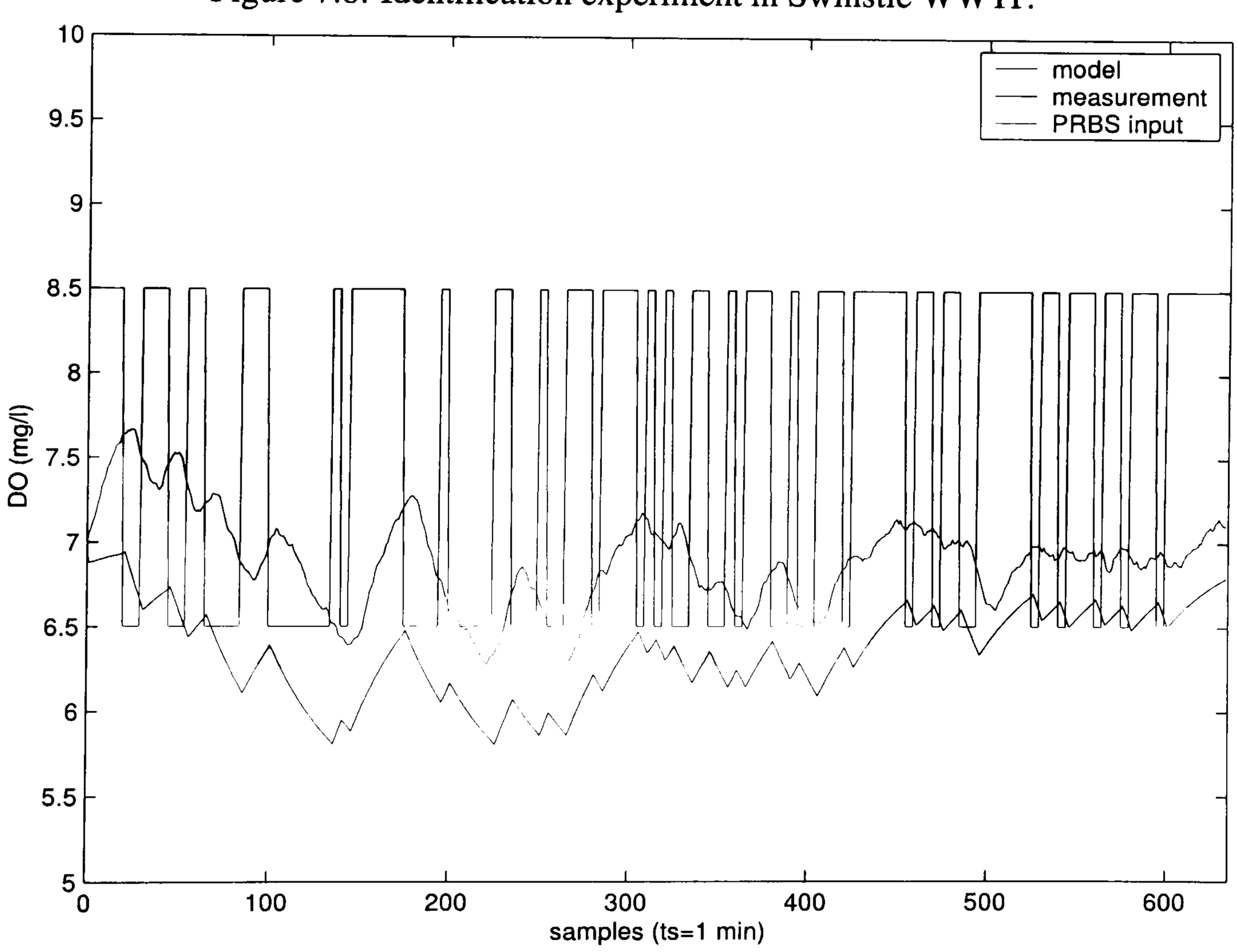

Figure 7.9: Dissolved oxygen model validation in Swinstie.

Table 7.2: Prediction and optimisation parameters.

H_p	H_u	$\mathscr{Q}$	$\mathscr{R}$
50	20	3	0.5

the design of the MPC controller is presented first, followed by results showing the improvement gained using the MPC controller.

7.4.1 MPC design

The first component to be designed for the MPC is the observer. The state observer follows the same formulation as in Chapter 3. A constant output correction has been introduced to compensate for unmeasured disturbances and plant-model mismatch. The observer was designed by placing the closed-loop poles as follows,

$$Poles = \begin{bmatrix} 0.7 & 0.6 & 0.5 \end{bmatrix}^T \tag{7.3}$$

Figure (7.10) shows the observer response when initialised. It can be seen that the observer requires around 25 minutes to converge to the actual measurement. Also, due to its large overshoot, it is of vital importance to allow sufficient time for the observer to converge before activating the MPC controller. This is just a precaution to avoid the controller producing unnecessary large control actions.

Figure (7.11) presents an interesting result regarding the linearity of dissolved oxygen. Notice in this figure that even though the oxygen level drops sharply, the observer is still able to track reasonably well the trajectory.

The predictor and optimisation parameters are presented in Table (7.2). These values have been obtained by carefully tuning the controller in real-time operation.

One of the main advantages of using a predictive controller is the facility to include constraints in the optimisation process. It is in this way that the physical limitations arising from the actuators (air compressors) are included when solving the optimisa-

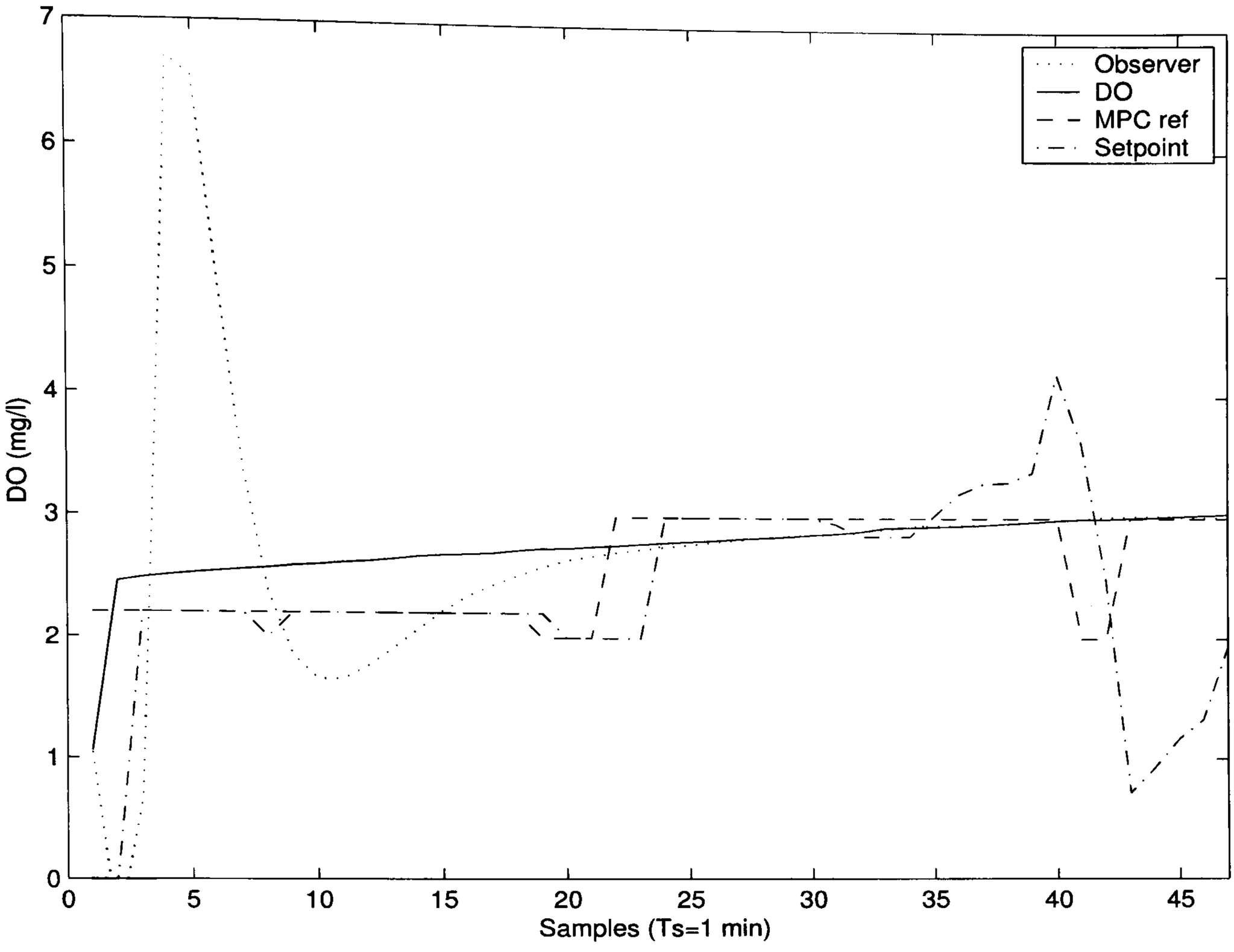

Figure 7.10: Dissolved oxygen observer response in Swinstie WWTP.

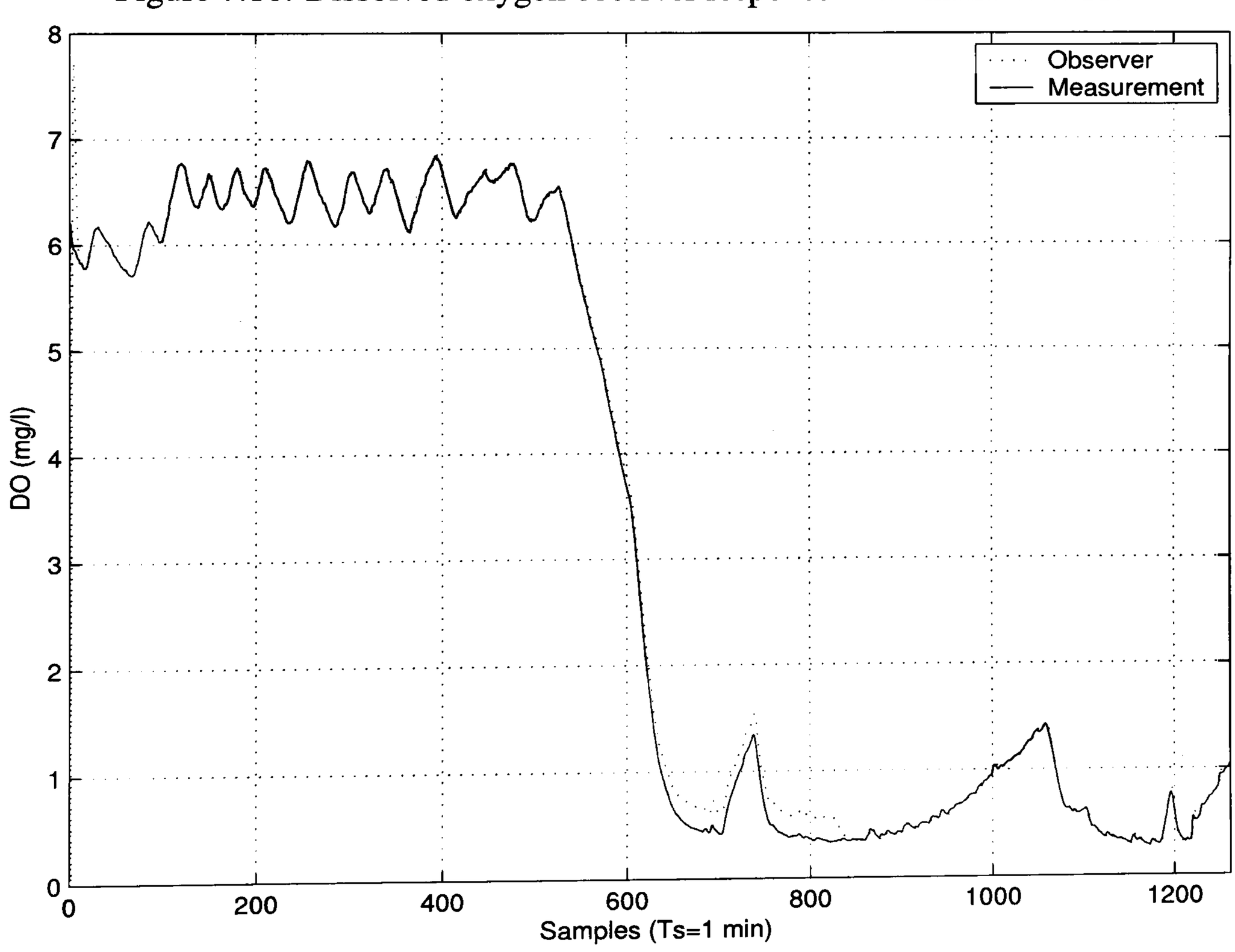

Figure 7.11: Dissolved oxygen observer response in Swinstie WWTP.

tion. Constraints also allow the inclusion of operation conditions that are necessary for the process to work. Thus constraints allow limits to be imposed over variables which in practice cannot go under or over certain limits, as for example the oxygen concentration cannot be less than zero.

The optimisation algorithm employed in the implementation of the MPC controller is a constrained least-squares. The use of this algorithm to solve the QP problem guarantees a robust numerical implementation. The implemented code also contains a protection mechanisms for infeasible problems, by using the last known optimisation solution.

7.4.2 Dissolved Oxygen control

The MPC controller implemented in Swinstie WWTP presents some degree of improvement over the existing controller. More tests are required in order to adequately tune the system and obtain a significant benefit. In general, the use of the MPC controller in Swinstie is limited to the plant actuators, since there is no possibility of performing this control at acceptable oxygen levels of around 2-3 mg/l. This limitation arises from the limited speed range of the blowers and the absence of a more sophisticated control structure (e.g. pressure loop, and motorised valves). However, just for the purpose of demonstration, the MPC has been set to run with high dissolved oxygen setpoints.

Figure (7.12) shows a step response of the system operating with the MPC. Notice as well in this figure, that the MPC is activated only after the observer has converged. Figure (7.13) shows a comparison between the control algorithm in the PLC operating by its own, and the system operating with the MPC.

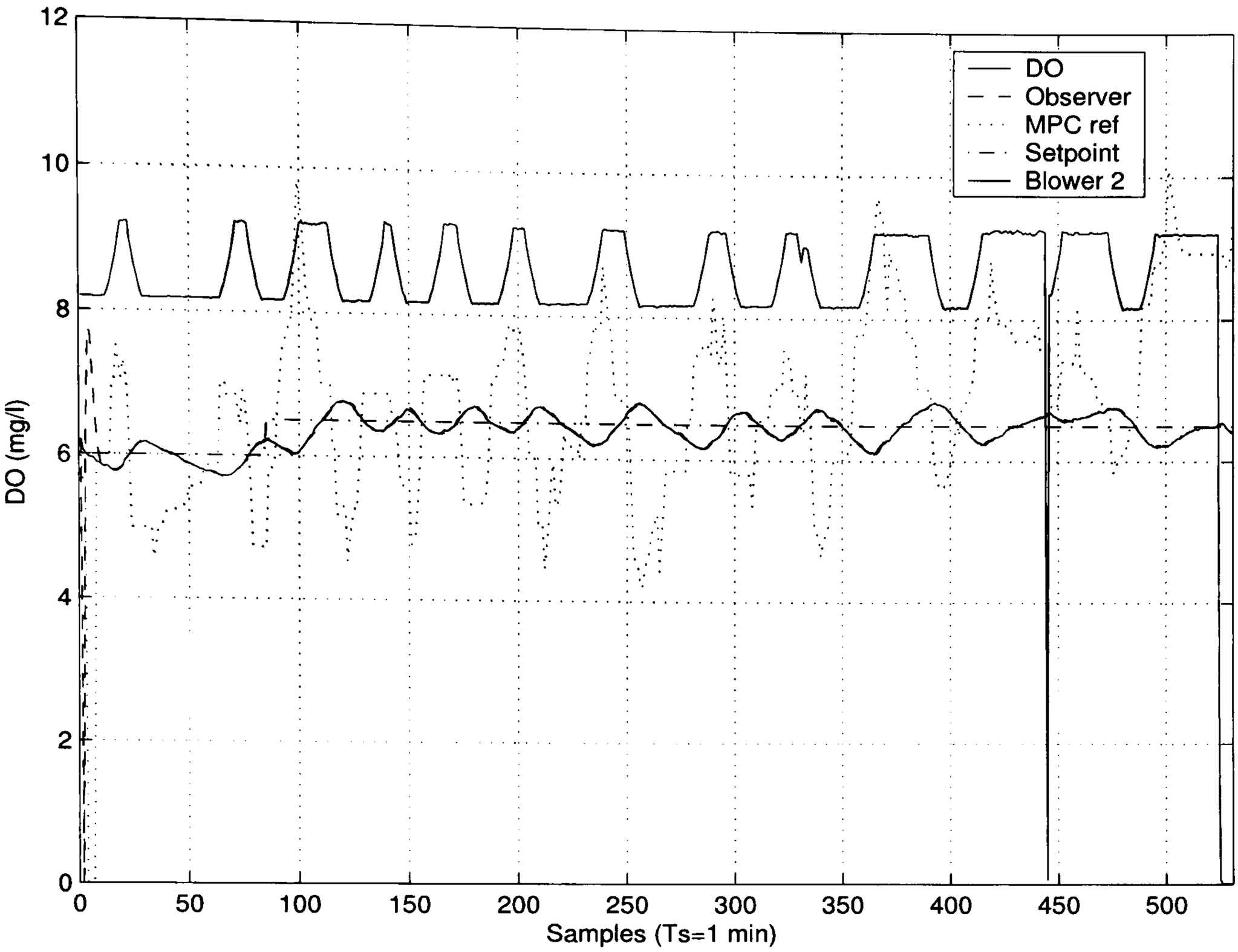

Figure 7.12: Step change response when using MPC.

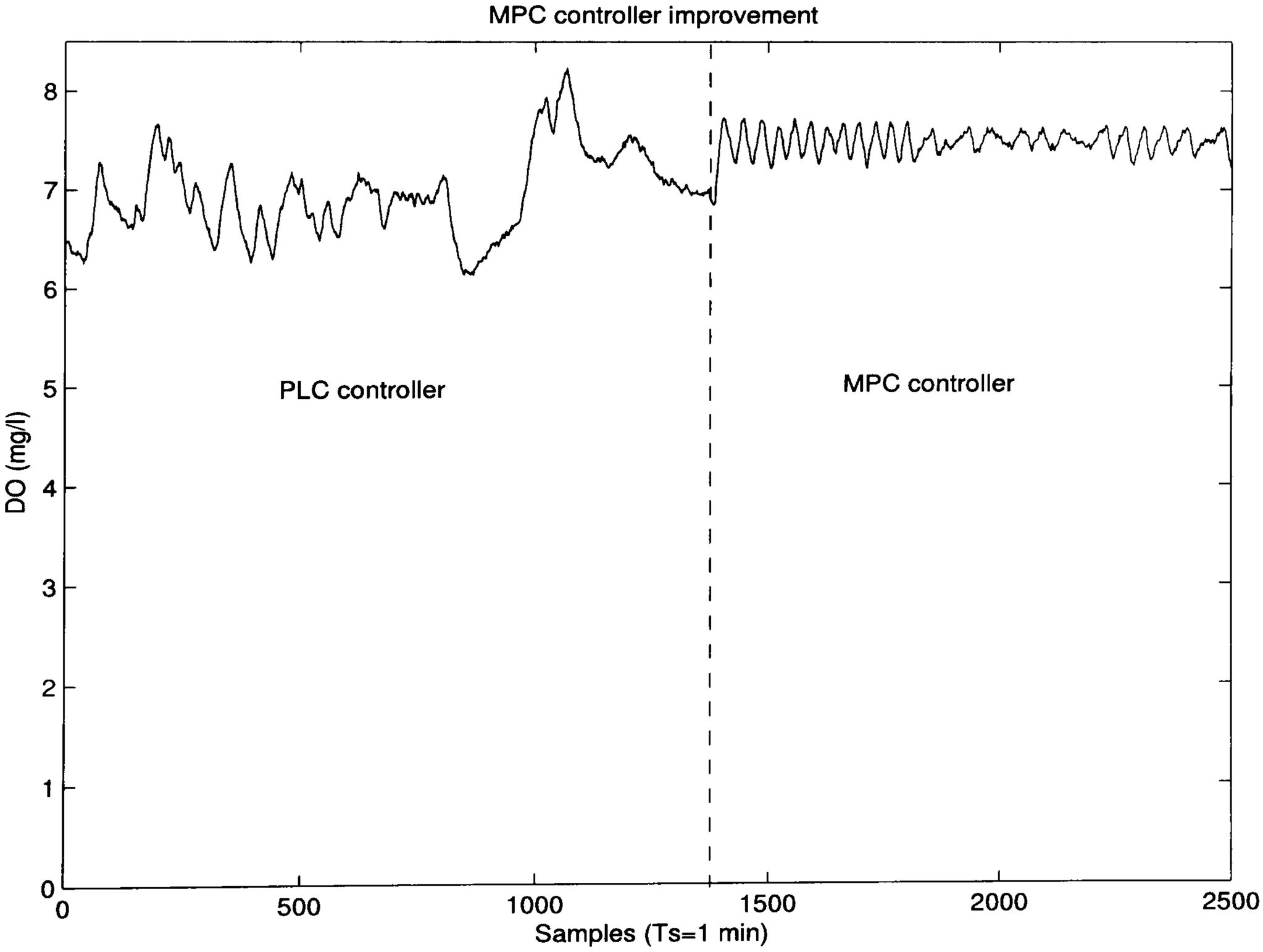

Figure 7.13: Variance reduction when using a MPC for dissolved oxygen control.

7.5 Summary

This chapter has presented the development and implementation of a control software platform, by the author, for advanced process control and process monitoring. The development of the software has used several tehcnologies to implement the different algorithms. MATLAB and LabVIEW have been employed to develop the software. The integration of code with the two software has been possible using: dynamic link libraries (DLLs) and Activex.

The interfacing of the software with the PLC in the plant has been implemented using OPC technology. OPC technology allows a fast, and efficient communication with the PLC.

The platform has been succesfully tested in full-scale in Swinstie WWTP. The test performed in real-time are: (a) Subspace identification, (b) MPC controller design (c) Process monitoring. The tests results have also corraborated previous obtained by simulation. Some of the most important convey the linearity of dissolved oxygen under the saturation level; therefore allowing the assumption of simple first or second order model approximations consistent for the rest of the work presented in this thesis.

On the control side, the chapter contains results which show the design of a real-time model predictive controller for dissolved oxygen in Swinstie WWTP. The results obtained show a considerable improvement over the existing controller. Further, the real-time implementation of this hierarchical control structure has proven valid for the purpose of performance improvement and cost reduction.

In conclusion, the development of the software tool, and the execution in real-time of identification, MPC control and process monitoring has added a very consistent component to the studies presented in the previous chapters. Moreover, much of the results presented have corroborated previous results obtained by simulation. Also, the software is open for more process industry applications and not limited to the wastewater treatment industry. The integration of MATLAB programmed algorithms into Lab-

VIEW has been performed using several technologies so the code could be executed in real-time. Some disadvantages of this approach is that MATLAB requires additional high-level programming (C), so it can be interfaced by standard C types. The code executions is however much faster than in the MATLAB environment.

Chapter 8

Conclusions and further work

The aim of the SMAC project is to design a system which is capable of embracing all possible conditions in a wastewater treatment plant by,

1. Integrating information from the different processes in the plant and the sewer system.
2. Utilizing the multivariable nature of the process.
3. Employing models to predict plant behaviour.
4. Efficiently rejecting disturbances.
5. Using advanced process control.

Therefore the main objective of this thesis has been to use a data-driven approach to:

1. Obtain simple linear models of dissolved-oxygen and nutrient removal for control purposes.
2. Explore the use model predictive control for the control of dissolved oxygen and nutrients.

3. Enhance the performance of low-level control systems by developing model-free and data-driven approaches for tuning.

4. Measure the performance of control-loops for tuning.

5. The use of identification algorithms, and model predictive control in real-time in a full-scale wastewater treatment plant.

The work presented in this thesis has achieved all these objectives. The obtained results indicate that there are potential applications in the use of data-driven algorithms for identification and control in wastewater treatment. However, it has also revealed that the problem can become complex, especially in the case of nutrients. Nutrients exhibit a non-linear behaviour, which leads to a difficult control problem to solve. This thesis has provided some initial insight into the potential of using linear-models for the prediction of nutrient dynamics with reasonably good results; however, the problem is far from solved.

The use of advanced process control strategies in a hierarchical structure would not be efficient if the low level control-loops do not perform well. This thesis has also explored the use of data-driven methods and model-free techniques to tune PID-type controllers. A new data-driven tuning algorithm for multivariable restricted-structure control systems, based in subspace identification has been one of the main achivements of this research.

The study of an LQG tuning method for PID controllers, lead also to a simple and efficient algorithm to tune SISO control loop. Further, the development of this method lead to explicit equations which could be used for loop performance monitoring.

One of the main contributions of this thesis has been the implementation of subspace identification, model predictive control and process monitoring algorithms in real-time in a full-scale wastewater treatment plant. The implementation of these algorithms has been possible through the design of a software platform in LabVIEW. Most of

the algorithms were originally programmed in MATLAB, and extensively used in the thesis for simulation. The software developed has raised interest for industrial use by Scottish Water, with possibilities of commercial exploitation.

In summary, this thesis has presented a comprehensive study in the fields of identification, control design, tuning of controllers and real-time implementation applied to the activated sludge wastewater treatment process. Further, the thesis covers a wide spectrum of results including theoretical, simulation and full-scale plant implementation and testing.

The following sections will discuss the results obtained in this thesis and will provide future lines of work.

8.1 Summary of achievements from the research

The main aims and achievements of the research of this thesis are discussed below:

1. **Identification of models for control purposes of dissolved oxygen and nitrogen removal in an activated sludge wastewater treatment plant under continuous aeration**. A comprehensive study of identification of suitable models for dissolved oxygen and nitrogen removal has been performed. The study has been performed employing a WWTP simulation model and real-plant data. Simulations results convey that dissolved oxygen exhibits a linear behaviour in almost all the range below the saturation point. The linear behaviour of dissolved oxygen has been corroborated by performing several identification exercises over data compiled from a full-scale WWTP in Denmark (Helsingor Kommune), and by performing on-line experiments in a WWTP in Scotland (Swinstie). **This work has been published as Sánchez and Katebi (2003); Sánchez (2002*a*, 2001).**

2. **Development of a model for control purposes for nitrogen removal in an alternating aerobic-anoxic wastewater treatment plant.** A linear model for nitrogen removal in an alternating aerobic-anoxic treatment plant has been developed. The model has been obtained by simplification and reduction of the ASM No.1 model. The novel feature, embedded in this model is its switching characteristic between the anoxic and aerobic phase. The model input is only the switching duty cycle and the output are mainly the nitrate and ammonia concentrations in the basin. **This work has been published as Sánchez *et al.* (2002).**

3. **MPC controller design for dissolved-oxygen and nutrient removal under continuous aeration.** The design of a hierarchical controller for nitrogen removal has been fully developed. The controller is composed of three control levels composed of: single SISO PI control loops for dissolved oxygen, a multivariable MPC driving the PI controllers in the three aeration basins, and a multivariable MPC controller, driving the dissolved oxygen setpoints and the internal recirculation flow to control nitrogen concentrations in the effluent and in the anoxic zone. The method improves nitrogen removal by controlling different processes in different time-scales. **This work has been published as Sánchez and Katebi (2003);Sánchez (2003*c*);Sánchez (2003*a*);Sánchez (2003*b*);Sánchez (2002*c*).**

4. **MPC controller formulation for alternating aerobic-anoxic wastewater treatment plant.** A new MPC controller formulation for the control of alternating aerobic-anoxic treatment plant has been developed. The formulation employs a switching model derived from the ASM model for the anoxic and aerobic phase. The controller employs the zero frequency component of the predicted outputs to calculate the optimal switching point. **This work has been published as Sánchez *et al.* (2002).**

5. **Implementation of advanced process control in real-time in a full-scale WWTP.**

The design and testing of a MPC controller in a full-scale wastewater treatment plant has been presented in this thesis. The design has been performed employing a software tool developed in LabVIEW and MATLAB\SIMULINK, and which allows the tuning of a MPC controller in real-time. It is shown in this thesis that the use of a MPC controller, in a hierarchical structure, improves considerably the performance of the system. Due to the hierarchical structure, its implementation can be performed in a more efficient computational machine (i.e. PC), without any risk to system integrity and producing a significant economical benefit. In addition, subspace identification algorithms have also been implemented for their use in this real-time environment. **This work has been published as Wade and Sánchez (2004);Sánchez *et al.* (2004*a*);Wade *et al.* (2004).**

6. **A new formulation for IFT and explicit solution of an optimal restricted structure LQG problem with applications to loop monitoring.** A new formulation for continuous-time deterministic systems of IFT has been developed in this thesis. The method allows the calculation of the parameters of a multivariable PID control system, by performing a series of successive experiments over the process and without carrying any explicit identification. A novel solution to a restricted structure LQG control problem is also presented. The solution of the problem leads to explicit formulas for the calculation of a controller parameter which is of the PID type. The value of the cost function can also be explicitly calculated, thus a restricted structure LQG benchmark is proposed for loop monitoring purposes. **This work has been published as Johnson and Sánchez (2003); Mahathanakiet *et al.* (2002); Sánchez (2004, 2002*b*).**

7. **A new tuning algorithm for multivariable restricted structure controllers.** A new tuning algorithm for multivariable restricted structure controllers using subspace identification is developed. The method employs input-output data

and initial knowledge of the controller structure and parameters to calculate the new optimal parameters based on a finite horizon LQG criterion minimisation. The method guarantees that the resulting closed-loop system will be stable if the optimisation converges. **This work has been published as Sánchez *et al.* (2004*b*, 2003*b*).**

8. **The development of a software platform for the testing of advanced process control and data quality management.** A software to test advanced process control algorithms as MPC and data quality management has been fully developed. The software is programmed using LabVIEW and MATLAB\SIMULINK. The main interface employs LabVIEW to communicate with the user and the PLC or SCADA. MATLAB has been used as programming language to implement the different algorithms. This code has later been compiled into C shared libraries, which are called by the main program in LabVIEW. All the experimental work performed in Swinstie WWTP has employed this software platform. **This work has been published as Sánchez *et al.* (2003*a*).**

8.2 Future work

Future work can be divided in the following areas:

- Identification
- Switching MPC control
- Multivariable loop-monitoring
- Convergence of optimisation problems

Identification

The problem of identification of models for nutrients is still an open issue. This thesis has presented a possible alternative to obtain a model for nutrients by employing subspace identification. The major problem however, lies in the highly non-linear characteristics of the processes. The identification of multiple models could be an alternative. Another important consideration is the excitation of the system for identification. From the results provided in this thesis, it appears that it is very difficult to sufficiently excite the process by just using dissolved oxygen levels and the recirculation flow rate. These are however, the main control handles in any activated sludge treatment process. Consequently, a more exhaustive study to determine the most efficient ways of exciting the process for identification purposes is a good line of research.

Switching MPC control

Intermittent aeration is a control approach which has demonstrated to be efficient in economical terms. Many of the treatment plants operating with this scheme (i.e. Helsingor WWTP) calculate the phase lengths only based on nutrient levels in a suboptimal way and employing a feedback structure only. This thesis has presented a control approach in which the influent flow is used as a feedforward signal; however, many assumptions have been made such as the number of available measurements. Also, the control methodology is only based on the mean of the predictions of the switching model. Consequently, research focusing on the design of filters which can provide accurate estimates of the state variables for a switching model and considerations to include maximum effluent concentrations in the predictions is strongly recommended.

Multivariable loop-monitoring

The measure of optimal performance is a subject of major research. Finding out when a process needs to be re-tuned has been a question difficult to answer, specially in multivariable processes. It is common in industry to employ rules of thumb and well known indexes (i.e. ISE, ITAE) which can give some indication of performance degradation. These however, are usually misleading due to unfair comparisons or because they only indicate local optimality and do not consider the multivariable nature of the process. The problem with the existent multivariable benchmarks is their high degree of complexity. Therefore, the development of simple numerically easy to use and understand benchmarks for multivariable processes is an interesting field.

Convergence of optimisation algorithms

This thesis has presented the development of several tuning algorithms with different characteristics. The most novel method however, is presented in Chapter 6, where a subspace identification framework is employed to calculate the parameters of a multivariable restricted-structure controller based only on closed-data and knowledge of the operating controller. The algorithm drives the process directly and seamlessly from the data collection towards the optimal parameters. However, the method employs a numerical optimisation of a nonlinear constrained function, for which there is no guarantee of convergence. Therefore an interesting line of research could be the determination of necessary and sufficient conditions to guarantee convergence.

References

Abramowitz, M. and Stegun, I.A., (Eds.) (1972). *Handbook of mathematical functions.* Dover publications.

Alex, J., R. Tschepetzki and U. Jumar (2002). Predictive control of nitrogen removal in WWTPs using parsimonious models. In: 15^{th} *IFAC Triennial World Congress*. Vol. Q Modelling and Control of Agricultural, Biological and Chemical Systems. 21-26 July. Barcelona, Spain.

Anderson, J.S., H. Kim, T.J. McAvoy and O.J. Hao (2000). Control of an alternating aerobic-anoxic activated sludge system. part 1: Development of a linearization based modelling approach. *Control Engineering Practice* **8**, 271–278.

Arnold, E. and S. Dietze (2001). Nonlinear moving horizon state estimation of an activated sludge model. In: *Proc. of the 9th IFAC/IFORS/IMACS/IFIP Symposium on Large Scale Systems:Theory & Applications*. 18-20 July. Bucharest, Romania.

Åström, K.J. and T. Hägglund (1985). Method and an apparatus in tuning a PID regulator. US Patent No. 4549123.

Bechmann, H. (1999). Modelling of Wastewater Systems. PhD thesis. Technical University of Denmark.

Carstensen, J. (1996). Identification of wastewater processes. PhD thesis. Technical University of Denmark (DTU).

Chiang, L., E. Russell and R. Braatz (2001). *Fault detection and diagnosis in industrial systems*. Advanced textbooks in control and signal processing. Springer-Verlag, London.

Chui, N.L.C. and J.M. Maciejowski (1996). Realization of stable models with subspace methods. *Automatica* **32**(11), p.1587–1595.

Copp, J., (Ed.) (2002). *COST Action 624 - The COST Simulation Benchmark: Description and Simulation Manual*. European Commission - European cooperation in the field of scientific and technical research.

Crowe, J. and M.A. Johnson (1998). New approaches to non-parametric identification for control applications. In: *Preprints IFAC Workshop on Adaptive Systems in Control and Signal Processing*. 26-28 August. Glasgow, Scotland. pp. 309–314.

Crowe, J. and M.A. Johnson (1999). A new non-identification procedure for online controller tuning. In: *American Control Conference*. 2-4 June. San Diego, USA. pp. 3337–3341.

Crowe, J. and M.A. Johnson (2000). Automated PI controller tuning using a phase locked loop identifier module. In: *IEEE International conference on Industrial Electronics, Control and Instrumentation, IECON 2000*. 22-28 October. Nagoya, Japan.

Crowe, J., M.A. Johnson and M.J. Grimble (2003). PID parameter cycling to tune industrial controllers - a new model-free approach. In: *Proc. of the 13th IFAC Symposium on System Identification*. 27-29 August. Rotterdam, The Netherlands.

De Moor, B., J. Vandewalle, L. Vandenberghe and P.V. Mieghem (1988). A geometrical strategy for the identification of state space models of linear multivariable systems with singular value decomposition. In: *Proc. IFAC 88*. Beijing. pp. 700–704.

Favoreel, W. and B. De Moor (1998). SPC : Subspace predictive control. Technical Report 98-49. Departement Elecktrotechniek - Katholieke Universiteit Leuven.

Favoreel, W., B. De Moor and P. van Overschee (2000). Subspace state space system identification for industrial processes. *Journal of process control* **10**(2-3), 149–155.

Favoreel, W., B. De Moor, M. Gevers and P. Van Overschee (1998). Closed loop model-free subspace-based LQG -design. Technical Report ESAT-SISTA/TR 1998-108. Departement Elecktrotechniek - Katholieke Universiteit Leuven.

Giovanini, L. and J. Marchetti (1999). Shaping time-domain responses with discrete controllers. *Ind. Eng. Chem. Res.* **38**, 4777–4789.

Godfrey, K. (1993). *Perturbation signals for system identification*. Prentice-Hall. Hertfordshire - UK.

Grimble, M.J. (2002). Restricted structure generalised minimum variance performance assessment and tuning. Technical Report ICC 191. ICC, University of Strathclyde. Scotland, UK.

Grimble, M.J. and A.W. Ordys (2001). Non-linear predictive control for manufacturing and robotics applications. In: *IEEE Conference on Methods and Models in Automation and Robotics*. 28-31 August. Miedzyzdroje, Poland. pp. 579–592.

Grimble, M.J. and M.A. Johnson (1988). *Optimal Control and Stochastic Estimation*. Vol. 1 and 2. John Wiley and Sons. Chichester, UK.

Gujer, W., M. Henze, T. Mino and M.C.M. Van Loosdrecht (1999). Activated sludge model no.3. *Wat. Sci. Tech.* **39**(1), 183–193.

Harris, T.J. (1989). Assessment of control loop performance. *Can. J. Chem. Eng.* **67**, 856–861.

Henze, M. (1997). *Wastewater Treatment - Biological and Chemical Processes*. Chap. 3. Basic Biological Processes, pp. 55–111. 2^{nd} ed.. Springer-Verlag. Berlin.

Henze, M., C.P.L. Grady, W. Gujer, G.v.R. Marais and T. Matsuo (1987). Activated sludge model no.1. Technical report. IAWQ Scientific and Technical Report No.1.

Henze, M., W. Gujer, T. Mino, T. Matsuo, M.C. Wentzel and G.v.R. Marais (1995). Activated sludge model no.2. Technical report. IAWQ Scientific and Technical Report No.3.

Henze, M., W. Gujer, T. Mino, T. Matsuo, M.C. Wentzel, G.v.R. Marais and M.C.M. Van Loosdrecht (1999). Activated sludge model no.2d - asm2d. *Wat. Sci. Tech.* **39**(1), 165–182.

Hjalmarsson, H. (1999). Efficient tuning of linear multivariable controllers using iterative feedback tuning. *Int. J. Adapt. Control Signal Process.* **13**(7), 553–572.

Hjalmarsson, H., S. Gunnarsson and M. Gevers (1994). A convergent iterative restricted complexity control design scheme. In: *33rd IEEE Conference on Decision and Control*. Vol. 2. 14-16 December. Lake Buena Vista, Florida, USA. pp. 1735–1740.

Hjalmarsson, H., S. Gunnarsson and M. Gevers (1995). Optimality and sub-optimality of iterative identification and control design schemes. In: *American control conference*. Vol. 4. 21-23 June. Seattle, Washington, USA. pp. 2559–2563.

Hjalmarsson, H., S. Gunnarsson and M. Gevers (1998). Iterative feedback tuning: Theory and applications. *IEEE Control Systems Society Magazine* pp. 26–41.

Ho, B. and R. Kalman (1966). Efficient construction of linear state variable models from input/output functions. *Regelungstechnik* **14**, 545–548.

Ho, W.K., Y. Hong, A. Hansson, H. Hjalmarsson and J.W. Deng (2003). Relay auto-tuning of PID controllers using iterative feedback tuning. *Automatica* **39**(1), 149–157.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Edu. Psychol.* **24**, 417–441,498–520.

Huang, B. and L. Shah (1999). *Performance assessment of control loops*. Springer Verlag London. ISBN 1-85233-639-0.

Huang, J. and H. Hao (1996). Alternating aerobic-anoxic process for nitrogen removal: Dynamic modelling. *Wat. Env. Res.* **68**, 94–102.

Ingildsen, P. (2002). Realising Full-Scale Control in Wastewater Treatment Systems using in-situ Nutrient Sensors. PhD thesis. Department of Industrial Electrical Engineering and Automation, Lund University.

Jeppsson, U. (1995). A simplified control-oriented model of the activated sludge process. *Mathematical Modelling of Systems* **1**(1), 3–16.

Jeppsson, U. (1996). Modelling Aspects of Wastewater Treatment Processes. PhD thesis. IEA - Lund Institute of Technology, Lund University.

Johnson, M.A. and A. Sánchez (2003). Process control loop tuning and monitoring using LQG optimality with applications in wastewater treatment plant. In: *IEEE 4th International Conference on Control & Automation ICCA'03*. 10-12 June. Montreal, Canada. pp. 84–90.

Kadali, R., B. Huang and A. Rossiter (2003). A data driven approach to predictive controller design. *Control Engineering Practice* **11**(3), 261–278.

Katebi, M.R. (2001). H_∞ state estimation in activated sludge processes. In: *Proc. of the 9th IFAC/IFORS/IMACS/IFIP Symposium on Large Scale Systems:Theory & Applications*. Bucharest.

Katebi, M.R. and M.H. Moradi (2001). Predictive PID controllers. *IEE Proc.-Control Theory Appl.* **148**(6), 478–487.

Katebi, M.R., M.A. Johnson and J. Wilkie (1999). *Control and Instrumentation of Wastewater Treatment Plants*. Advances in Industrial Control Monograph Series. Springer Verlag. London.

Kim, H., T.J. McAvoy, J.S. Anderson and O.J. Hao (2000). Control of an alternating aerobic-anoxic activated sludge system. part 2: Optimisation using a linearized model. *Control Engineering Practice* **8**, 279–289.

Kung, S. (1978). A new identification and model reduction algorithm via singular value decomposition. In: *Proc. 12th Asilomar Conf. on Circuits, Systems and Computers*. pp. 705–714.

Larimore, W.E. (1990). Canonical variate analysis in identification, filtering and adaptive control. In: *Proc. 29th Conf. on Decision and Control CDC'90*. Honolulu. pp. 596–604.

Li, W., H.H. Yue, S.Valle-Cervantes and S. Joe Qin (2000). Recursive PCA for adaptive process monitoring. *Journal of process control* **10**(5), 471–486.

Lindberg, C.F. (1997). Control and Estimation Strategies Applied to the Activated Sludge Process. PhD thesis. Uppsala University.

Ljung, L. (1987). *System Identification - Theory for the User*. Prentice Hall. Englewood Cliffs.

Maciejowski, J.M. (2001). *Predictive Control with Constraints*. Prentice Hall. England.

Mahathanakiet, K., M.A. Johnson, A. Sanchez and M. Wade (2002). Iterative feedback tuning and an application to wastewater treatment plant. In: *Asian Control Conference*. 25-27 September. Singapore.

Marinaki, M. and M. Papageorgiou (2002). Technology review report. Technical Report 23-15. Technical University of Crete.

Mayne, D.Q. (2001). Control of constrained dynamic systems. *European Journal of Control* **7**, 87–99.

Mayne, D.Q., J.B. Rawlings, C.V. Rao and P.O.M. Scokaert (2000). Constrained model predictive control: stability and optimality. *Automatica* **36**(6), 789–814.

Metcalf and Eddy, (Eds.) (1991). *Wastewater Engineering: Treatment, Disposal and Reuse*. Chap. 12, pp. 779–795. Civil Series Engineering. third ed.. McGraw-Hill International Editions. Singapore.

Moonen, M., B. De Moor, L. Vandenberghe and J. Vandewalle (1989). On- and off-line identification of linear state space models. *Int. J. Control* **49**, 219–232.

Moradi, M.H., M.R. Katebi and M.A. Johnson (2002). The MIMO predictive PID controller design. *Asian Journal of Control* **4**(4), 452–463.

Nielsen, M.K. (2001). Control of wastewater systems in practice. Technical Report 3. IWA, Instrumentation, Control and Automatisation specialist group.

Nielsen, M.K., (Ed.) (2002). *Deliverable 8: Knowledge Extraction*. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Nielsen, M.K., H. Bechmann and M. Henze (2000). Modelling and test of aeration tank settling (ATS). *Wat.Sci.Tech.* **41**(9), 179–184.

Nielsen, M.K., J. Carstensen and P. Harremoës (1996). Combined control of sewer and treatment plant during rainstorm. *Wat.Sci.Tech.* **34**(3), 181–187.

Nielsen, T.M., D. Thornberg and B. Hoog (2002). Biological phosphorus removal by periodic anaerobic release. *Water Intelligence Online* **1**, 100–112. Original paper from ICA 2001 Malmö - Sweden.

Ordys, A.W. and M.J. Grimble (2001). Predictive control design for systems with state-dependent nonlinearities. In: *SIAM Conference on Control and its Applications*. 9-13 July. San Diego, California.

Puta, H., G. Reichl and R. Franke (1999). Model based optimisation of a wastewater treatment plant. In: *Proc. of the European Control Conference ECC*. 31 August - 3 September. Karlsruhe, Germany.

Robbins, H. and S. Munro (1951). A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407.

Ruscio, D.D. (1997*a*). Model based predictive control: An extended state-space approach. In: *Proc. of the* 36^{th} *Conference on Decision and Control*. San Diego, CA. pp. 3210–3217.

Ruscio, D.D. (1997*b*). Model predictive control and identification: A linear state-space approach. In: *Proc. of the* 36^{th} *Conference on Decision and Control*. San Diego, CA. pp. 3202–3209.

Ruscio, D.D. (2000). *Subspace System Identification: Theory and Applications*. 6th ed.. Telemark Institute of Technology. Porsgrunn, Norway. Lecture Notes.

Sánchez, A. (2001). *Deliverable 5: Definition of System Performance Assessment Criteria and Definition of Models for Monitoring and Control*. Chap. 5 Definition of system performance criteria, and selection of models for monitoring and control. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Sánchez, A. (2002*a*). *Deliverable 6: Algorithms for System Monitoring*. Chap. Models for estimation and control: model for Fast Layer. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Sánchez, A. (2002*b*). *Deliverable 7: Measures and Algorithms for risk and situation assessment*. Chap. 7 Report on measures and algorithms for risk and situation assessment. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Sánchez, A. (2002*c*). *Deliverable 9: Coordination Control System Architecture and Design*. Chap. 5 Fast Control Layer. SMAC - Smart Control of Wastewater Treatment Systems.

Sánchez, A. (2003*a*). *Deliverable 11: Internal WWTP Flow Rate Control Functions, Algorithms, Design*. Chap. Development of a MPC for the Fast Control Layer (several sections). SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Sánchez, A. (2003*b*). *Deliverable 12: External WWTP Flow Rate Control Functions, Algorithms, Design*. Chap. 3 Data-based loop controller tuning and multivariable dissolved oxygen control. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Sánchez, A. (2004). *Deliverable 10: Report on Operational Planning Procedures*. Chap. 4 Operational planning procedures at the Swinstie test site. SMAC - Smart Control of Wastewater Treatment Systems.

Sánchez, A. and M.R. Katebi (2003). Predictive control of dissolved oxygen in an activated sludge wastewater treatment plant. In: *European Control Conference ECC 2003*. 1-4 September. Cambridge, UK.

Sánchez, A., (Ed.) (2003*c*). *Deliverable 13: SMArt Control System Design, Algorithm, and Software Report*. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Sánchez, A., M. Wade and M.R. Katebi (2004*a*). On real-time control and process monitoring of wastewater treatment plants. Real-time control. *Trans. Inst. Measurement and Control* **Submitted**, 0–0.

Sánchez, A., M.R. Katebi and M.A. Johnson (2002). Optimal control of an alternating aerobic-anoxic wastewater treatment plant. In: *Proc. of the* 15^{th} *Triennial*

IFAC World Congress. Vol. Q Modelling and Control of Agricultural, Biological and Chemical Systems. 21-26 July. Barcelona, Spain.

Sánchez, A., M.R. Katebi and M.A. Johnson (2003*a*). Design and implementation of a control platform for the testing of advanced control systems and data quality management in the wastewater industry. In: *Proc. of the IEEE* 4^{th} *International Conference on Control & Automation ICCA'03*. 10-12 June. Montreal, Canada. pp. 68–74.

Sánchez, A., M.R. Katebi and M.A. Johnson (2003*b*). Subspace identification based PID control tuning. In: *Proc. of the* 13^{th} *IFAC Symposium on System Identification*. 27-29 August. Rotterdam, The Netherlands.

Sánchez, A., M.R. Katebi and M.A. Johnson (2004*b*). A tuning algorithm for multivariable restricted structure control systems using subspace identification. *Int. J. Adapt. Control Signal Process.* **accepted for publication**, 0–0.

Sotomayor, O. and C. Garcia (2002). Model-based predictive control of a pre-denitrification plant: a linear state-space model approach. In: 15^{th} *IFAC Triennial World Congress*. Vol. Q Modelling and Control of Agricultural, Biological and Chemical Systems. 21-26 July. Barcelona, Spain.

Sotomayor, O.A.Z., S.W. Park and C. Garcia (2003). Multivariable identification of an activated sludge process with subspace-based algorithms. *Control Engineering Practice* **11**(8), p. 961–969.

Takcs, I., G.G. Patry and D. Nolasco (1991). A dynamic model of the clarification thickening process. *Wat. Res.* **25**(10), 1263–1271.

Tulleken, H.J.A.F. (1990). Generalized binary noise test-signal. *Automatica* **26**(1), 37–49.

Uduehi, D., A. Ordys and M.J. Grimble (2002). Multivariable PID controller design using online generalised predictive control optimisation. In: *Proc. of the 2002 IEEE*

International Conference on Control Applications. Vol. 1. 18-20 September. Glasgow. pp. 272–277.

van Overschee, P. and B. De Moor (1994). N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* **30**, 75–93.

van Overschee, P. and B. De Moor (1995). A unifying theorem for three subspace system identification algorithms. *Automatica* **31**(12), 1853–1864.

van Overschee, P. and B. De Moor (1996*a*). Closed-loop subspace system identification. Technical Report ESAT-SISTA/TR 1996-521. Departement Elecktrotechniek - Katholieke Universiteit Leuven. Submitted for publication in IFAC.

van Overschee, P. and B. De Moor (1996*b*). *Subspace Identification for Linear Systems*. Kluwer Academic Publishers. USA.

Verhaegen, M. (1991). A novel non-iterative MIMO state-space model identification technique. In: *IFAC/IFORS Symposium on Identification and System Parameter Estimation*. Budapest. pp. 1453–1458.

Verhaegen, M. (1993). Application of a subspace model identification technique to identify LTI systems operating in closed-loop. *Automatica* **29**(4), 1027–1040.

Verhaegen, M. (1994). Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica* **30**, 61–74.

Viberg, M., B. Ottersten, B. Wahlberg and L. Ljung (1993). Performance of subspace-based system identification methods.. In: *Proc. IFAC 93*. Vol. 7. Sydney, Australia. pp. 369–372.

Vrecko, D., N. Hvala and B. Carlsson (2003). Feedforward-feedback control of an activated sludge process - a simulation study. *Wat.Sci.Tech.* **47**(12), 19–26.

Wade, M. and Sánchez, A., (Eds.) (2004). *Deliverable 17: Prototype participant - Scottish Water*. SMAC - Smart Control of Wastewater Treatment Systems. http://www.smac.dk.

Wade, M., M.R. katebi and A. Sánchez (2004). On real-time control and process monitoring of wastewater treatment plants. Real-time process monitoring. *Trans. Inst. Measurement and Control* **submited**, 0–0.

Wade, M.J. (2004). Process monitoring and knowledge extraction in wastewater treatment plants. PhD thesis. University of Strathclyde. Glasgow.

Wilson, F. (1981). *Design Calculations in Wastewater Treatment*. Spon. London.

Yu, C.C. (1999). *Autotuning of PID controllers: Relay Feedback Approach*. Springer Verlag London.

Yuan, Z., J. Keller and P. Lant (2001). Optimization and control of nitrogen removal in activated sludge processes: A review of recent developments. Technical Report 2. IWA,Instrumentation, Control and Automatisation specialist group.

Zieger, H. and A. McEwen (1974). Approximate linear realizations of given dimension via Ho's algorithm. *IEEE Trans. Autom. Control* **19**(2), 153–163.

Ziegler, J.G. and N.B. Nichols (1942). Optimum setting for automatic controllers. *Transactions of the ASME* **64**, 759–768.

Appendix A

How to generate a DLL from a MATLAB function script to run in LabVIEW

1. Basic Requirements

   (a) Make sure you have:

      i. MATLAB compiler

      ii. C/C++ compiler. Most people can afford to buy MS Visual C/C++ so why not try (eg: Open Watcom http://www.openwatcom.org/)

   (b) Make sure your MATLAB compiler and C/C++ compiler are running: Try mathworks technical note 1621 (http://www.mathworks.com/support/tech-notes/1600/1621.shtml). This note should give you enough information about how to setup your MATLAB compiler and C/C++ compiler.

2. A simple example:

   The problem of generating a dll from MATLAB to work in LabVIEW is that the MATLAB compiler requires special data types (e.g. mxArray). It is therefore

required to write a special 'wrapper function' which interfaces the MATLAB generated source code with a more 'standard' ANSI C code. The following example illustrates the procedure. The code following has been implemented using MATLAB 5.3, MATLAB Compiler 2.0, Open Watcom 1.0, and LabVIEW 6.1.

(a) Write the following m-code function and save as foo.m:

function y = foo(x)

y = 2*x

(b) Compile into 'C' code: Type the following command in the MATLAB prompt:

>>mcc -t -L C -W lib:foolib -h foo.m

This will generate the following files:

foo.c: contains the implementation of foo (Mfoo) and the interfacing functions (mlfFoo, mlxFoo)

foo.h: contains the prototypes of mlfFoo and mlxFoo

foolib.c: contains the implementations of foolibInitialize and foolibTermintate, necessary to initialize and terminate mlfFoo

foolib.h: contains the prototypes of mlxFoo, foolibInitialize and foolibTerminate.

Foolib.exports: contains the symbols to export in the dll.

(c) Create foo_wrapper.c. This is the wrapper function which will allow the use of the C implementation of foo.m. The code is listed and commented below.

```
/* This file foo_wrapper.c */
#include "matlab.h"
#include "foodll.h"
#include "matrix.h"
```

```
//main wrapper function definition
double wrapper_main(double *in1){
//declare variable to deliver result double out;
//Create two pointers of mxArray type to store inputs and outputs
mxArray *in1_ptr, *out1_ptr;
//Allocate input pointer to a 1 by 1 double, real matrix
in1_ptr = mxCreateDoubleMatrix(1,1,mxREAL);
//Move the data from the input to the pointer
fill(mxGetPr(in1_ptr),in1,1);
//Initialise foo implementation
foolibInitialize();
//Pass values to mlfFoo and receive in mxArray type variable
out1_ptr = mlfFoo(in1_ptr);
//Terminate foo implementation
foolibTerminate();
//Move from mxArray type to double type
fill(in1,mxGetPr(out1_ptr),1);
//move data to output variable
out = *in1;
//Return value
return(out); }
void fill(double *out, double *in, int size){
//This function moves data from one type to another
int i;
for(i=0;i<size;i++)
out[i] = in[i]; }
```

(d) Create foodll.h which contains prototypes of functions in foo_wrapper.c. The code is listed below.

```
//This file foodll.h
double wrapper_main(double *in1);
void fill(double *out, double *in, int size);
```

(e) Add entry point to foolib.exports: add the following line to the file foolib.exports:

```
wrapper_main
```

(f) Build using the following command in MATLAB:

```
>>mbuild -link shared foo_wrapper.c foo.c foolib.c foolib.exports
```

(g) You have now generated a file called foo_wrapper.dll, which contains the function wrapper_main.c among others. Use this function making sure the input (arguments) and output (return value) are of 8-bit double type. Also, notice that you should pass a pointer to the value and not the value itself to the function.

Appendix B

Recursive principal component analysis

B.1 Recursive correlation matrix update

The following algorithm is taken from Li *et al.* (2000). Let $X_1^0 \in \mathbb{R}^{\eta_1 \times m}$ be a matrix of raw data, where m is the number of measured variables, η_1 is an initial data block size. Also $\sigma_{i,j}$ is the standard deviation of column j of the initial data block. The recursive update of the correlation matrix algorithm with forgetting factor μ can be summarised as follows,

1. Initialisation

$$b_1 = \left(X_1^0\right)^T \cdot 1_{\eta_1}$$

$$\Sigma_1 = diag\left(\sigma_{11}, \cdots, \sigma_{1m}\right)$$

$$X_1 = \left[X_1^0 - 1_{\eta_1} b_1^T\right] \Sigma_1^{-1}$$

$$R_1 = \frac{1}{\eta_1 - 1} X_1^T X_1$$

2. Recursive update

$$X_{k+1}^0 = \begin{bmatrix} X_k^0 \\ X_{\eta_{k+1}}^0 \end{bmatrix}$$

$$b_{k+1} = \mu b_k + (1-\mu)\frac{1}{\eta_{k+1}}\left(X^0_{\eta_{k+1}}\right)^T \cdot 1_{\eta_{k+1}}$$

$$\Delta b_{k+1} = b_{k+1} - b_k$$

$$\sigma^2_{k+1,i} = \mu\left(\sigma^2_{k,i} + \Delta b^2_{k+1}(i)\right) + (1-\mu)\frac{1}{\eta_{k+1}} \times \left\| X^0_{\eta_{k+1}}(:,i) - 1_{\eta_{k+1}} b_{k+1}(i) \right\|^2$$

$$\Sigma_{k+1} = diag\left(\sigma_{k+1,1}, \cdots, \sigma_{k+1,m}\right)$$

$$X_{\eta_{k+1}} = \left[X^0_{\eta_{k+1}} - 1_{\eta_{k+1}} b^T_{\eta_{k+1}}\right]\Sigma^{-1}_{k+1}$$

$$X_{k+1} = \begin{bmatrix} X_k \Sigma_k \Sigma^{-1}_{k+1} - 1_k \Delta b^T_{k+1} \Sigma^{-1}_{k+1} \\ X_{\eta_{k+1}} \end{bmatrix}$$

$$R_{k+1} = \mu\Sigma^{-1}_{k+1}\left(\Sigma_k R_k \Sigma_k + \Delta b_{k+1}\Delta b^T_{k+1}\right)\Sigma^{-1}_{k+1} + (1-\mu)\frac{1}{\eta_{k+1}} X^T_{\eta_{k+1}} X_{\eta_{k+1}}$$

B.2 Hotelling's T^2 statistic

Let $X \in \mathbb{R}^{n\times m}$ be a data training set consisting of m variables and n measurements for each variable. The covariance matrix of the training set is,

$$R = \frac{1}{n-1}X^T X$$

An eigenvalue decomposition of the matrix R,

$$R = V\Lambda V^T$$

Let,

$$z = \Lambda^{-1/2} V^T x$$

The Hotelling's T^2 statistic is given by,

$$T^2 = z^T z$$

B.3 Q statistic

The Q statistic is defined as,

$$Q = r^T r$$

where,

$$r = (I - PP^T)x$$

and P is the loading vector of the SVD decomposition.

Appendix C

Tag mapping in Swinstie WWTP

Table C.1: Tag mapping

Tag	Register	Description	Tag	Register	Description
DO1	V568	DO meter 1	DOH	V1057	DO high setpoint
DO2	V665	DO meter 2	DOL	V1058	DO low setpoint
DO3	V569	DO meter 3	DOerror%	V1059	DO error % allowance
DO4	V666	DO meter 4	BLOWupdate	V1068	Blower update time
DO5	V570	DO meter 5	BLOW1run	C00873	Blower 1 status
DO6	V667	DO meter 6	BLOW2run	C00951	Blower 2 status
DO7	V571	DO meter 7	BLOW3run	C00876	Blower 3 status
DO8	V668	DO meter 8	BLOW4run	C00954	Blower 4 status
DO9	V572	DO meter 9	MLRP1run	C0855	Recirculation pump 1
DO10	V573	DO meter 10	MLRP2run	C0858	Recirculation pump 2
MLSS1	V574	MLSS meter 1	MLRP3run	C0861	Recirculation pump 3
MLSS2	V669	MLSS meter 2	MLRP4run	C0864	Recirculation pump 4
MLSS3	V575	MLSS meter 3	FM1	V503	Flow meter 1
MLSS4	V670	MLSS meter 4	FM2	V504	Flow meter 2
MLSS5	V576	MLSS meter 5	FM3	V564	Flow meter 3
MLSS6	V671	MLSS meter 6	FM4	V662	Flow meter 4
BLOWER1	V560	Blower 1 speed	FM5	V663	Flow meter 5
BLOWER2	V660	Blower 2 speed	RAS1	C0963	Return sludge pump 1
BLOWER3	V561	Blower 3 speed	RAS2	C0966	Return sludge pump 2
BLOWER4	V661	Blower 4 speed			