Methodology for the optimization of wastewater treatment plant control laws based on modeling and multi-objective genetic algorithms

Benoit Beraud

UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC

T H E S E

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Génie des procédés

Ecole Doctorale *:* Science des Procédés-Science des Aliments

présentée et soutenue publiquement

par

Benoît BERAUD

le 19 mars 2009

_____

Méthodologie d'optimisation du contrôle/commande des usines de traitement des eaux résiduaires urbaines basée sur la modélisation et les algorithmes génétiques multi-objectifs

_____

JURY

Pr. Alain GRASMICK	
Professeur, Université Montpellier II	Examinateur
Pr. Willy GUJER	
Professeur, EAWAG	Président
M. Cyrille LEMOINE	
Directeur de Programme, Veolia Environnement-CRE	Examinateur
Dr. Marie-Noëlle PONS	
Directeur de recherche, CNRS-ENSIC-INPL	Rapporteur
Pr. Mathieu SPERANDIO	
Directeur de recherche, INSA Toulouse	Rapporteur
Dr. Jean-Philippe STEYER	
Directeur de Recherche, INRA Narbonne	Directeur de Thèse
Dr. Imre TAKACS	
Expert en modélisation, Envirosim	Examinateur

Année : N°:

Cette thèse a été réalisée en partenariat entre le Laboratoire de Biotechnologie de l'Environnement de l'INRA et le Centre de Recherche sur l'Eau de Veolia Environnement.

INRA, UR50, Laboratoire de Biotechnologie de l'Environnement
Avenue des Etangs
F-11000 NARBONNE

Centre de Recherche sur l'Eau
Chemin de la Digue - BP 76
F-78603 MAISONS LAFFITTE

Cette thèse a été financièrement supportée par le Centre de Recherche sur l'Eau de Veolia Environnement et la bourse CIFRE de l'Association Nationale de la Recherche et de la Technologie n° 40/2006.

Remerciements

Je souhaite tout d'abord adresser mes sincères remerciements et toute ma reconnaissance à mon directeur de thèse, Dr. Jean-Philippe Steyer. Son aide infaillible, ses conseils avisés, son soutien, sa qualité de visionnaire et sa conception de la recherche industrielle m'ont permis de réaliser ce travail dans d'excellentes conditions. Il a notamment su rester le garant de la qualité de la recherche effectuée lors de cette thèse et a su m'ouvrir d'innombrables portes sur la recherche mondiale.

Je souhaite également exprimer ma profonde gratitude à M. Cyrille Lemoine, encadrant industriel de cette thèse. Sa connaissance des défis industriels, son ouverture au travail universitaire, sa confiance indéfectible et son soutien de tous les jours m'ont permis de réaliser ce travail dans d'incomparables conditions cognitives et matérielles. Je remercie également MM. Hervé Suty et Jean Cantet qui ont accepté de m'accueillir au sein de leurs équipes et m'ont ainsi ouvert les portes de la recherche industrielle.

Je remercie également MM. Eric Latrille (INRA-LBE Narbonne), Cyril Printemps-Vacquier (Veolia DT St-Maurice), Krist Gernaey (Technical University of Denmark), Christian Rosen (VA-Ingenjörerna, Sweden) et Cristian Trelea (AgroParisTech) qui ont participés très activement à cette thèse au travers des multiples comités de pilotage. Ils m'ont apporté un regard toujours plus neuf sur mes travaux. Je leur suis très reconnaissant des questions et débats qu'ils ont su susciter et qui ont permis d'améliorer sans cesse mes travaux. Je les remercie également pour leur disponibilité tout au long de ma thèse lors de mes divers questionnements et interrogations.

Mes sincères remerciements à toutes les personnes du CRE qui ont été impliquées dans ce travail, que ce soit par une aide ponctuelle ou par une implication plus conséquente dans le projet. Pour n'en citer que quelques uns, bien que la liste soit beaucoup plus longue, je remercie MM. Julien Chabrol, Pascal Boisson, Julie Jimenez, Chrystelle Ayache, Nicolas David, Olivier Daniel et Gildas Manic.

Merci également à MM. Philippe Duverllie, Magalie Denis, Stéphane Mer, Anne Godard et Mathieu Rossard de la Direction Technique Régionale d'Arras et à MM. Jean-Michel, Laurent et Lionel de l'usine de traitement de Cambrai. Leur accord et leur aide ont permis la mise en place de toute la partie applicative sur site réel sans laquelle l'intérêt industriel et scientifique de cette étude aurait été grandement diminué.

Je souhaite enfin remercier tous mes collègues du CRE où j'ai effectué la majorité de mon travail. Merci pour votre bonne humeur, votre passion et votre énergie sans lesquelles je ne serais surement pas parvenu à réaliser ce travail de manière aussi complète. Un merci également tout particulier à Mlle Sophie Vaudran pour son assistance administrative, sa bienveillance et sa disponibilité de tout instant. Merci également à toute l'équipe du LBE pour avoir toujours su m'accueillir à merveille lors de mes différents déplacements à Narbonne.

Table of contents

Chapter 3 - Multiobjective optimization with genetic algorithms

Chapter 4 - Optimization methodology development on a literature case study

Chapter 5 - Application of the methodology to Cambrai WWTP

Chapter 6 - Conclusion- contributions and perspectives for future development

Table des matières

Chapitre 3 – L'optimisation multi-objectifs à l'aide d'algorithmes génétiques

Chapitre 4 – Développement de la méthodologie d'optimisation sur un cas d'étude de la littérature

List of figures

Figure 5.1: Geographical location of Cambrai..... 147

Figure 5.2: Physical layout of the WWTP and arrangements of individual processes 149

Figure 5.3: Inputs and outputs of the storm basin model..... 153

List of tables

Nomenclature

ADM	anaerobic digestion model
AMSTAR	aeration module of STAR
ASM	activated sludge model
ASU	activated sludge unit
ATV	Abwasser Technische Vereinigung
BSM	benchmark simulation model
BOD5	biological oxygen demand at five days
COD	chemical oxygen demand
CSO	combined sewer overflow
DO	dissolved oxygen
DWID	dry weather input dataset
CSIRO	Australian commonwealth scientific and research organization
GA	genetic algorithm
EPA	environmental protection agency
IAWQ	international association on water quality
ISS	inorganic suspended solids
IWA	international water association
LCFA	long chain fatty acids
LP	linear programming
MILP	mixed-integer linear programming
MINLP	mixed-integer non-linear programming
MOGA	multiobjective genetic algorithm
NGL	total nitrogen
NLP	non-linear programming
NPGA	niched Pareto genetic algorithm
NSGA	non-dominated sorting genetic algorithm
ORP	oxidation-reduction potential
PAES	Pareto archived evolution strategy
PE	population equivalent
PESA	Pareto envelope-based selection algorithm
PFC	predictive functional controller
PI	proportional-integral
PSO	particle swarm optimization
RWID	rain weather input dataset
SA	simulated annealing
SABAL	sequenced aeration based on ammonia levels
SNDN	simultaneous nitrification/denitrification
SPEA	strength-Pareto evolutionary algorithm
STAR	superior tuning and reporting
TN	total nitrogen
TKN	total Kjeldahl nitrogen
TS	tabu search
TSS	total suspended solids
VSS	volatile suspended solids
WWTP	wastewater treatment plant

List of publications

Part of the work presented in this thesis have already been published as shown below.

National Conferences

Beraud B., Lemoine C., Steyer J.P., Latrille E. (2007). Optimization of a control law for simultaneous nitrification/denitrification by means of a multiobjective genetic algorithm. 5ème édition des journées Sciences et Technologies de l'Information et de la Communication pour l'Environnement (STIC 2007), Lyon, France, 13-15 Nov. 2007, 8pp.

International Conferences

Beraud B., Steyer J.P., Lemoine C., Gernaey K.V. (2007). Model-based generation of continuous influent data from daily mean measurements available at industrial scale. Autmonet 2007, IWA, Gent, Belgium, 5-7 Sept. 2007, 8 pp.

Beraud B., Steyer J.P., Lemoine C., Latrille E., Manic G., Printemps-Vacquier C. (2007). Towards a global multi objective optimization of wastewater treatment plant based on modeling and genetic algorithms. Watermatex 2007, Washington DC, USA, 7-9 Mai 2007, 8pp.

Beraud B., Steyer J.P., Lemoine C. , Latrille E. (2008). Optimization of WWTP control by means of multi-objective genetic algorithms and sensitivity analysis. 18th European Symposium on Computer Aided Process Engineering, ESCAPE 18, Lyon, France, 1-4 June 2008, 8 pp.

Journals

Beraud B., Steyer J.P., Lemoine C. , Latrille E., Manic G., Printemps-Vacquier C. (2007). Towards a global multi objective optimization of wastewater treatment plant based on modeling and genetic algorithms. Water Science and Technology, 56(9), pp. 109-116.

Beraud B., Steyer J.P., Lemoine C. , Latrille E. (2008). Optimization of WWTP control by means of multi-objective genetic algorithms and sensitivity analysis. Computer Aided Chemical Engineering, 25, 2008, pp. 539-544.

Books

Beraud B., Lemoine C., Steyer J.P. (accepted). Multiobjective Genetic Algorithms for the Optimisation of Wastewater Treatment Processes. In Nicoletti M.C., Jain L.C. (Eds.). Computational Intelligent Techniques for Bioprocess Modelling, Supervision and Control. Studies in Computational Intelligence Springer-Verlag, Germany (accepted), 34 pp.

Thesis outline

This thesis is divided into 6 chapters.

Chapter 1 introduces the thesis with the description of the context and challenges identified, the solutions already existing in the literature and the objectives of the thesis, taking into account the existing work.

Then, chapters 2 and 3 present the theoretical background of the thesis while chapter 4 and 5 detail the main new contributions of this work to research.

More specifically, chapter 2 focuses on the description of main wastewater treatment plant processes as well as their modeling. Most common aeration strategies for wastewater treatment plants are detailed in this chapter. Finally, literature case studies available for the development of the thesis are described, with a description of the main objectives to consider in these applications as well as an illustration of typical control laws behavior in one of these models.

Chapter 3 presents the theory of multiobjective genetic algorithms. The first section of this chapter is a general description of genetic algorithms, their key principles and operations. Then, the interest of the multiobjective approach proposed in this thesis is explained, as well as the genetic algorithm chosen for this study.

Chapter 4 presents the development of the optimization methodology for wastewater treatment plant control law optimization. This development is based on the Benchmark Simulation Model 1 (BSM1) and the full application of the methodology on this literature case study is also presented in this chapter.

Chapter 5 enlarges the scope of the thesis with an application of the methodology on the real wastewater treatment plant of Cambrai, located in the north of France. The challenges of this application and preparatory work are first detailed, followed by the application of the optimization methodology itself.

Finally, chapter 6 concludes this thesis. The contributions of this thesis to scientific research are first summarized, followed by the limitations and perspectives of the work presented. Details about three of these perspectives are finally given in this chapter, followed by a conclusion of future research needs.

Structure de la thèse

Cette thèse est divisée en 6 chapitres

Le 1er chapitre introduit cette thèse en présentant son contexte et les défis identifiés, les solutions déjà existantes dans la littérature scientifique ainsi que les objectifs qui ont été défini pour cette thèse pour donner suite aux études réalisées à ce jour.

Ensuite, les chapitres 2 et 3 présentent les fondements théoriques de cette thèse tandis que les chapitres 4 et 5 présentent les principales contributions de ce travail du point de vue de la recherche académique et appliquée.

Le chapitre 2 présente plus particulièrement la description des principaux procédés utilisés dans les stations d'épuration d'eaux usées ainsi que leur modélisation. Les principales lois de commande de l'aération des procédés à boues activées sont ensuite décrites. Les cas d'étude disponibles dans la littérature et pouvant servir au développement de cette thèse sont ensuite présentés, ainsi que les principaux objectifs à considérer pour l'évaluation des performances d'une usine d'épuration. Enfin, une illustration du fonctionnement de deux lois de contrôle sur le cas d'étude sélectionné vient clore ce chapitre.

Le chapitre 3 présente la théorie des algorithmes génétiques multi-objectifs. La première section de ce chapitre contient une description générale des algorithmes génétiques, de leurs principales caractéristiques et de leur fonctionnement. L'intérêt de l'approche multi-objectif proposée dans cette thèse est ensuite expliqué, suivi par une explication du fonctionnement détaillé de l'algorithme sélectionné pour cette thèse.

Le développement de la méthodologie pour l'optimisation des lois de commandes des stations d'épuration à boues activées est ensuite présenté dans le chapitre 4. Ce développement est basé sur le « Benchmark Simulation Model 1 », cas d'étude largement étudié dans la littérature. Le protocole de la méthodologie est tout d'abord détaillé, suivi par une application complète sur le cas d'étude considéré.

Le chapitre 5 élargit le champ d'investigation de cette thèse en présentant l'application de la méthodologie d'optimisation sur le cas réel de la station d'épuration de Cambrai, située au Nord de la France. Les défis de cette application réelle ainsi que le travail préparatoire sont tout d'abord présentés, suivis par l'application de la méthodologie pour la comparaison de deux lois de commande de l'aération de bassins de boues activées.

Finalement, le chapitre 6 conclut cette thèse. Les contributions de ce travail à la recherche scientifique sont tout d'abord résumées, suivi par les limitations et perspectives de cette thèse. Trois perspectives sont plus particulièrement détaillées. Enfin, une conclusion sur les futurs travaux de recherche nécessaires est présentée.

Chapter 1 - Introduction

Chapitre 1 - Introduction

1.1 Purpose

Continuous human population growth and its concentration in cities as well as the industrialization have dramatically increased water demand and water pollution over the last decades and centuries. Solutions have been proposed with wastewater treatment plants that remove pollution from waste flows before reintroducing them in the natural environments such as rivers, lakes or seas. Such treatment is important regarding the preservation of the ecosystems but also because a great amount of drinking water is pumped from these same rivers, especially in urban areas. For instance, in France, 38% of drinking water is produced with surface water (IFEN, 2006). It should also be noticed that preserving the ecosystems of rivers, lakes and seas might preserve their natural capacities of removing small loads of pollutants (even if some pollutants are persistent and cannot be removed by natural ecosystems).

Drinking water is becoming less available throughout the world (Cosgrove, 2000), not only in developing countries in warm regions but also in developed countries such as Australia. In fact, recent studies are indicating that changes in water availability in the future will be the consequence of population growth and industrialization more than the consequences of climate change. In such dry areas, there will be a great demand for the direct reuse of cleaned water directly from the wastewater treatment plant (WWTP). For instance, CSIRO researchers in Australia are working on aquifer storage and recovery of WWTP cleaned water and Singapore's Public Utilities Board is producing ‘NEWater’ from the local WWTP. ‘NEWater’ is a “cleaned” and potable water, mostly used by industries as fresh process water but it can also be drunk by humans. For now, reuse of cleaned wastewater is mainly directed to industrial applications needing huge amount of process water but its use for drinking water may be extended in the future.

All these considerations are putting high pressure on the enhancement of the wastewater treatment industries. Main solutions consist in using multi-barrier treatment lines or/and advanced processes to ensure efficient treatment under all conditions. Examples of new technologies are the biofilters and membrane bioreactors that enhance the quality of the treatment with the same or even smaller plant size. However, these solutions can induce higher costs of wastewater treatment and there is considerable pressure from the populations

(and therefore the politicians) to not increase these costs, mainly due to a lack of information about future drinking water availability challenges.

Another solution to enhance wastewater treatment consists in optimizing the current operation and process sizes by implementing advanced control laws that will enable better monitoring of the process operation.

The first problem in this area is the availability of reliable sensors for the measurement of the process state variables. Due to continuous advances in instrumentation over the last decades, the first reliable and cheap solutions are now available on the market for main pollutant measurements (Jeppsson et al., 2002; Ingildsen and Olsson, 2002; Ingildsen and Wendelboe, 2003; Kaelin *et al.*, 2008).

The second problem in this area is the development of adequate control laws. This is mainly a theoretical problem of the development of control laws that are in accordance with the available measurements and their reliability. Many solutions are also available in this area nowadays (Olsson *et al.*, 2005; Olsson and Jeppsson, 2006).

The third and remaining problem about the optimization of a WWTP operating conditions is the choice of the most adequate control law according to the WWTP specific size, incoming loads and constraints of the regulation on the treatment performances, as well as its optimal settings. No universal answer to this problem can be provided but, as will be emphasized in this thesis, efficient solutions can be proposed.

Moreover, this third problem cannot be solved for each specific case study. Apart from the huge cost of each test, the main reason is that long evaluation of each control law is required for the real process. This is due to the required adaptation of the biomass to the new operations that usually takes between one and three months. The problem with this long evaluation is the continuous change of the incoming load. This change is induced by seasonal variations on an annual basis as well as population and industry changes on a multi-annual basis. It implies bias in the evaluation of each control law that cannot be corrected easily. Another point against the full scale evaluation and comparison of control laws is the difficulties and reluctance usually encountered by practitioners when testing new parameters on a real process since the legislation must always be respected.

The answer to this problem of optimizing a given WWTP functioning can therefore only be based on theoretical studies (in simulations), that will be then verified on the real process. This is possible thanks to the huge amount of optimization techniques available, as well as the models of WWTP processes that have become more and more reliable since their start in the eighties. Solutions have already been proposed in the scientific community but some remaining challenges for their real implementation will now be described.

1.2 Challenges and solutions already proposed in the literature

The theoretical optimization of WWTP control laws is challenged by two main problems: (i) the continuous non-steady state operation of the WWTP and (ii) the huge amount of often contradictory objectives that have to be considered in practice.

The first problem is the consequence of continuous and huge variations in incoming pollutant loads. This problem has already been partially addressed in the literature (further details will be provided in the following) but some improvements can be brought by considering more complex variations than the typical daily variations usually considered.

The objectives of the second problem include effluent quality, energy consumption, and sludge production. These objectives are difficult to take into account together. However, the consideration of a single one is not reliable either. For instance, we can say that the main objective of a WWTP is to have the best effluent quality possible, regardless of energy consumption. This however, is not true because energy consumption also has a big environmental and economical impact (e.g. indirect CO_2 emissions, pollution taxes, etc.). As a consequence, the goal of a WWTP is more to provide effluent quality and energy consumption in accordance with a global environmental and economical impact point of view, with energy consumption also compatible. However, as these two objectives are sometimes opposed, a compromise has to be found and is still sought by the optimization scientific community.

Mainly, two kinds of solutions are proposed in the literature to provide an optimization tool applied to bioprocesses (Sakizlis *et al.*, 2004).

In a first category of optimization attempts (Steyer and Harmand, 2000; Vera *et al.*, 2003), the authors are considering two objectives: one is the economical optimality of the process, the

other takes into account the dynamical performances around a nominal point. These attempts are based only on steady state computations that are complemented with a disturbance analysis.

In Steyer and Harmand (2000), a controller is optimized for the nitrification process. The objective is to minimize the "noise to signal" ratio for sensors by manipulating the parameters of the control law under constraints of operational performances. In Vera *et al.* (2003), the optimization of the design and control law of a WWTP bioprocess is performed. The objective is to minimize the error of the controller and the investment costs corresponding to the chosen design. A step perturbation is considered for the evaluation of the controller performance, the goal being to reject this perturbation to stay as close as possible to a fixed set point. Constraints are added to ensure a proper operation of the process.

The main weakness of these two optimization attempts is that they only take into account the rejection of a single perturbation. The extension of the results to continuously changing perturbations that arise at the WWTP inlet is not considered and may be at risk.

In the second category of optimization attempts, authors focus only on a single economical performance index which allows the use of dynamic simulations for the optimization. Operational objectives are transformed into constraints that have to be satisfied. Such techniques have the drawback to not provide clear insight about the trade-off occurring between the economical and the operational objectives. Examples are the optimization of (i) aeration control (Balku and Berber, 2006; Fikar *et al.*, 2005; Holenda *et al.*, 2007) and (ii) overall plant design and control (Rivas *et al.*, 2008).

In Balku and Berber (2006), the objective is to minimize energy consumption by manipulating the aeration sequences under constraints on effluent quality. Dynamic simulations are performed for each candidate solution. The main drawback of this method is that the optimal aeration profile found for the perturbations used in the simulation may not be adequate for other perturbations. As the control law acts in open-loop (i.e. no measurements of incoming pollution or current process states are made), no correction is possible. Another weakness is that the impact of the optimal aeration profiles on the long term is not assessed and may lead to problems such as the washout of bacteria from the system (Chachuat *et al.*, 2005).

In Fikar *et al.* (2005), the objective is to minimize the mean total nitrogen concentration in the effluent by manipulating aeration sequences. This objective is meant to also minimize the aeration energy and to be the most critical objective of the studied case. A Non Linear Programming (NLP) solver dedicated to dynamic optimization is therefore used. To address the potential risk of microbial washout, a constraint is added to ensure that initial and final states are equal, these state values being included in the optimization variables. A control law based on measurement of nitrate and oxygen is derived from the optimal aeration profile found. This allows closed-loop functioning and helps to reject disturbances. The main limitation is that only one day of dry weather is considered for the simulations. This limits the optimality of the solution as rain weather is not considered.

In Holenda *et al.* (2007), the objective is to minimize effluent quality, considering that aeration energy will be optimized at the same time. A genetic algorithm is used and 10 days of dynamic simulations are performed for each candidate solution but the same daily variations are used for every simulated day. Once more, the main limitation is that the solution is specific to the perturbation considered. Very good performance is indeed obtained for this specific case but may degrade quickly when the perturbation changes, which is always the case in practical WWTPs and this is not evaluated. Another limitation is the choice of the initial states. They are assumed to be related to the number of cycles per day. This assumption has very limited accuracy. For instance, the initial states will obviously not be the same if the aeration is on 10% or 90% of time in average, even if the number of cycles is identical in both cases.

In Rivas *et al.* (2008), steady-state and dynamic optimizations of the design and operation of a complex WWTP layout are performed. Different optimizations are performed in sequence with various objectives and manipulated variables. This allows the authors to have an insight into the potential performance. The main limitation is that all objectives and all potentially manipulated variables are not considered at the same time. The underlying assumption is that the problems are unassociated, which may not always be the case in practice. Another limitation is that no clear vision of the compromises between the conflicting objectives is possible.

To conclude, none of the studies found in the literature address all the challenges identified for optimization of the WWTP control laws. Promising solutions have however been

proposed and they will serve as a basis for the development of a methodology addressing all challenges.

1.3 Objectives of the thesis and solution proposed

This thesis attempts to develop a methodology for the optimization of WWTP control laws solving the two challenges identified in the previous subsection.

The three main questions that this thesis attempts to answer are:

- Which optimization technique is suitable for the optimization of WWTP control laws
- How to define and to handle multiple and relevant objectives
- How to compare the various WWTP control laws.

In order to make the work of this thesis practical and to open good perspectives, further questions are addressed:

- How to perform optimizations that will provide reliable results in a reasonable computing time?
- Is there a methodology that can be easily used on other optimization problems in the water field (i.e. sewer and drinking water networks, drinking water production, etc.)?

Constraints corresponding to the challenges identified include:

- Complete information about the trade-offs induced by the various objectives of the optimization needs to be provided at the end of the optimization.
- Robustness of the control law with regard to yearly WWTP inlet variations must be addressed.
- The study must be based on existing models and simulators to allow further use of this work.

The solution proposed in this thesis to address all these questions and constraints will now be described in general terms.

Currently, two main categories of algorithms are used in the field of optimization: (i) techniques based on the theory of Linear, Non-Linear and Mixed Integer Programming

(LP, NLP MILP, MINLP) and (ii) techniques based on metaheuristics like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Tabu Search (TS), Simulated Annealing (SA), etc.

The three main differences between these two techniques are:

- the first ones have the capability to provide the exact solution to the problem while the second ones only try to approach the solution as fast as possible, usually without providing its exact value;
- the second ones are capable of finding the overall solution to the problem even when multiple and sharp local minima are present, whereas the first ones usually require multiple runs with different starting points to find the overall solution;
- the fact that the first ones may require many evaluations of the function objectives as they require information about the values of the derivatives at the current point while the second ones surely require many evaluations.

In our case, we did not focus too much on the search of exact solutions as there are already many uncertainties present in the models used and in the real processes so that the parameterization of WWTP control laws does not need high accuracy on the solutions (usually only two significant digits are required).

The second point concerns the practical applicability in our problems. The coupling of typical non-linear processes used in WWTPs as well as the complex link between parameters to act on and objectives to optimize will generally induce non-convex optimization problems and therefore many local optima.

As for the third point, the exact formula of the objective functions derivatives regarding a change in the parameter values will never be available in our case studies. Approximate derivatives must therefore be computed and almost the same number of objective evaluations will be performed with both algorithm categories.

In conclusion, the two first points support use of metaheuristics while the third one does not provide any useful information. The decision was therefore made to use a metaheuristic for the development of the optimization methodology presented in this thesis.

The choice then had to be made among existing metaheuristics. Very recent applications of the four techniques mentioned above in the field of process and control law design are available in the literature but do not provide any clear indication of which is most suitable: (i) application of GAs to WWTP processes in Balku *et al.* (2006) and Holenda *et al.* (2007), (ii) application of PSO to wastewater collection network in Izquierdo *et al.* (2008), (iii) application of TS to WWTP and typical product modification process in Exler *et al.* (2008) and (iv) application of SA to chemicals production in Halim *et al.* (2008).

Apart from its applicability to the studied field, another constraint on the metaheuristic to choose is that it should be able to solve multiobjective problems. Adaptations of the genetic algorithms were found to be the best solution to tackle such problems (Abraham and Jain, 2005) because of its working procedure, as well as the very little knowledge required about the problem to be solved (no need of information on the convexity or continuity of the set of best trade-offs).

The methodology developed in this thesis is therefore based on the combination of a genetic algorithm for multiobjective optimization with existing WWTP models. Details of this are given in chapter 4.

The next two chapters of this thesis focus on background material about (i) WWTP modeling (chapter 2) and (ii) multiobjective genetic algorithms (chapter 3). The development and application of the methodology is presented in a case study in chapter 4. The application of a real case study is covered in chapter 5. Finally, chapter 6 concludes this thesis with a summary of the contributions and a brief overview of the perspectives for future developments in the studied area.

Chapter 2 - Current situation in WWTP control, modeling and simulation

Chapitre 2 – État des lieux de la commande, de la modélisation et de la simulation des stations d'épuration

Choice of optimization methodology must be based on the characteristics of the system studied. However the diverse processes used for the treatment of wastewater in conventional activated sludge systems, their non-linearity and the high level of disturbances induced by incoming wastewater typically limit the range of reliable optimization techniques. This section will highlight these characteristics together with a description of the typical processes encountered in WWTPs. The state-of-the-art models of WWTP secondary treatment are detailed in a second section. A third part of this chapter focuses on typical control laws used to control the key process considered in our application: aeration of the activated sludge unit. Finally, the benchmark simulation models, two frameworks for unbiased comparison of WWTP control laws, are presented in the last section. An illustration of the behavior of two control laws that will be optimized on the benchmark simulation models is finally proposed.

2.1 Typical characteristics of a municipal WWTP

WWTPs, as any real life systems, have many specificities which have to be considered before choosing an optimization technique. These particularities are detailed in this section, first considering the characteristics of the influent going into our system and causing major disturbances. Typical processes encountered in WWTPs to treat this influent are then presented.

2.1.1 Influent characteristics

Wastewater is the collection of all reject waters from households and small industries not equipped with their own treatment plant. It is composed of many pollutants. The three main ones that can be treated in conventional processes are organic carbon, nitrogen and phosphorus. Nitrogen and phosphorus are responsible for ecosystem eutrophication by favoring the algae development. This eutrophication may indeed lead to the destruction of the ecosystem by reducing incoming solar flux and decrease of oxygen concentrations.

This wastewater pollution is usually referred to in terms of chemical oxygen demand (COD), total suspended solids (TSS), total nitrogen (TN or NGL), Total Kjeldahl Nitrogen (TKN), ammonia (S_{NH}) and phosphorus (S_{PO4}). Chemical oxygen demand is a measurement of the quantity of oxygen required to oxidize all organic pollutants. It is a widely accepted measurement of organic carbon pollution. The concentration of total suspended solids

represents the amount of particulate compounds. Total nitrogen is the total mass of nitrogen for all kind of molecules while the Kjeldahl measurement does not take nitrate and nitrite (oxidized nitrogen forms) into account.

The most important characteristic of wastewater is that it is strongly influenced by daily and weekly variations induced by human activity. Long-term modifications also occur. They may be seasonal events like holidays or long-term demographic changes (both inducing either an increase or decrease in the population).

Typical variations in the influent flow rate at the inlet of the wastewater treatment plant are illustrated in Figure 2.1 (the flow rate corresponds to typical daily production per inhabitant in the absence of industry, obtained with the influent generator of Gernaey *et al.*, 2006b).

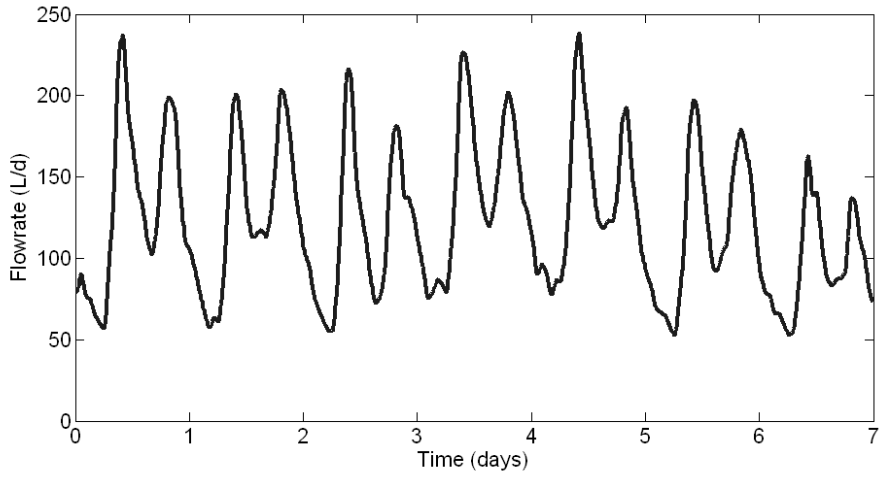

Figure 2.1: Typical variations of influent flow rate

Similar variations occur for pollutant concentrations. The pattern of these variations is influenced by:

- sewer topology,

- condition of the pipe (infiltration, holes, etc.),

- rainfall events in case of combined sewer (collecting rain and wastewater),

- occasional disturbances from households (oil releases, mainly during weekends due to do-it-yourself paint works, etc.) or industries (washout of some process tanks, etc.).

These variations induce large disturbances of wastewater treatment plant processes. Traditionally, these disturbances were lowered using high volume processes. Nowadays, more compact systems are constructed and advanced control laws and actuators capable of handling these disturbances are required.

2.1.2 Main processes involved

For treatment of typical wastewater, primary and secondary treatments are essential in all WWTPs. The primary treatment consists in screening and settling heavy particle compounds (e.g. sand, gravels) and floating greases. The secondary treatment consists in removing organic carbon, nitrogen and sometimes phosphorus. The processes of this secondary treatment are usually based on biological treatment with activated sludge units (anaerobic, anoxic and aerobic) and a physical separation with settlers/clarifiers.

Many physical configurations of biological treatment units are used, ranging from the most extensive activated sludge tanks to the most compact ones (fixed-culture processes, either with fixed or moving media). Many layouts of these basic processes are used in order to have maximum pollutant treatment with minimum energy consumption and minimum tank volume. The main constraints influencing these different designs are local legislation, the influent characteristics and the experience and expertise of the company designing the plant.

The settling and clarification operations are included at the end of the secondary treatment in a single process with two outputs. From one standpoint, this can be viewed as clarification, since one of the outputs should contain water without particulates compounds. From another standpoint, this can be viewed as settling or thickening, since the other output should contain only particulate compounds with a very little amount of clean water.

Phosphate treatment can be included in the primary or secondary treatment. This can be a physico-chemical treatment based on chemical precipitation or a biological treatment with specific bacteria. The biological treatment is sometimes difficult to control while the physico-chemical treatment requires reactants. Both require additional tank volume.

Treatment of wasted sludge from primary and secondary treatment is a key issue nowadays. The most common practices used to provide sludge as fertilizer to farmers or dispose the sludge in landfills after thickening and drying to remove as much water as possible. More

often now, the sludge is incinerated to produce something similar to sand that can then be used as a material (e.g. for road construction). This change is mainly linked to the increase in micro-pollutants in raw water and its sanitary impact. These processes are however energy-consuming and they release many different compounds and greenhouse gases that increase treatment cost. This is why another process is becoming popular: anaerobic digestion. This process reduces the quantity of sludge produced by the WWTP and produces methane, which can be easily transformed into electricity, reducing both the environmental impact of the WWTP and its energy bill.

When the ecosystem of the receiving body is very sensitive and there are stringent legislative constraints, or more commonly when the treated water is intended to be directly reused as process water or irrigation water, final treatment is included. This consists in removing all remaining micro-organisms and pathogens from the secondary treatment using filtration (membrane, sand, etc.) and disinfection with chlorine, ozone and/or ultraviolet light.

2.2 Main models of WWTPs

The main processes of the secondary treatment will now be detailed in this section. Optimization of primary treatment and final filtration and disinfection of water is not considered in this thesis and they will hence be omitted.

2.2.1 Activated sludge units

The activated sludge process was discovered in 1913 in the United-Kingdom with experiments on the treatment of wastewater in a draw-and-fill reactor. First, wastewater was filled into a reactor and oxygen was supplied. Pollution was hence removed from the water. Then, a phase of settling was performed. Finally, clean water was removed and sludge was kept in the reactor. These actions were repeated many times. The experiments showed that the water quality was enhanced cycle after cycle. Scientists first thought that the sludge was activated (in a manner similar to activated carbon) and the process was therefore named the activated sludge process. The name of the process remained even after it was realized that there was no activation of the sludge but a concentration of bacteria and a selection of the best-performing organisms.

Three main pollutant removal processes occur in an activated sludge process: oxidation of organic compounds with oxygen as electron acceptor, transformation of ammonia into nitrate (also named *nitrification*) and oxidation of organic compounds with nitrate as electron acceptor, occurring when no more oxygen is available as electron acceptor (also named *denitrification* as it includes the transformation of nitrate into gaseous nitrogen).

The first two processes require the presence of oxygen while the last one requires its absence. This is the main reason why activated sludge units are usually intermittently aerated. Recent advances tend to promote advanced control laws that could lead to simultaneous reactions by providing the exact amount of oxygen required (Olsson *et al.*, 2005; Lemoine and Grelier, 2008, Thauré *et al.*, 2008). Details of these control laws are given in section 2.3.

Over the last twenty years, four dynamic models for activated sludge processes have been published by the IWA (International Water Association), formerly IAWQ (International Association on Water Quality). These models are known in the literature as the ASM models (i.e. activated sludge models ASM1, ASM2, ASM2d and ASM3). They are summarized in a report of the IWA by Henze *et al.* (2000). They are now the models most widely used to represent the behavior of activated sludge processes, even if many minor deviations have been proposed in the literature.

All four models are based on the same principle of description. They are macroscopic models considering the bacteria consortium in its whole, and even if individual cells have different characteristics, only their mean characteristics are considered. The functioning of the activated sludge tank is described through a fractionation of the mixed liquor (activated sludge and wastewater) into different compounds (substrates, particulates compounds and organisms) and through a number of processes which describe the transfers between the different compounds/organisms, based on mass conservation equations. The reaction rates of these different processes depend on the concentration of compounds/organisms in the mixed liquor by means of Monod activation terms.

As originally suggested by Willi Gujer to the Task Group that produced ASM1[1], the models can be summarized in a table named the "Petersen matrix", where the components are represented in columns and the processes in rows, the stoichiometry between the various components for each processes is represented in the table and the process kinetics are represented in the last right column of the table. Table 2.1 illustrates the ASM1 kinetic model.

ASM1 is the first model developed by Henze *et al.* (1987). The main processes occurring in an activated sludge reactor are included in this model (i.e. oxidation of organic compounds, nitrification and denitrification) as well as the decay of the biomass, the ammonification of soluble organic nitrogen and the hydrolysis of the particulate substrates. The fractionation of the mixed liquor is based on 13 components: particulate and soluble substrate (X_S and S_S), particulate and soluble inert compounds (X_I and S_I), autotrophic and heterotrophic organisms (X_A and X_H), particulate inert compounds arising from biomass decay (X_P), oxygen (S_O), ammonium and ammonia nitrogen (S_{NH}), nitrate and nitrite (S_{NO}), soluble and particulate organic nitrogen (S_{ND} and X_{ND}) and alkalinity (S_{ALK}). ASM1 is the simplest model of the four IWA but some important processes were recognized as missing.

In ASM2 (Henze *et al.*, 1995) and ASM2d (Henze *et al.*, 1999), the processes of biological phosphorus removal and chemical phosphorus coagulation were been added. This induces a finer fractionation of the mixed liquor with new components corresponding to these processes: soluble phosphorus (S_{PO4}), phosphorus accumulating organisms (X_{PAO}) and their internal cell storage of organics and poly-phosphate (X_{PHA} and X_{PP}), metal-hydroxides (X_{MeOH}) and metal-phosphate (X_{MeP}). Soluble organic substrate (S_S) is divided into fermentable substrate (S_F) and fermentation products (S_A), with a new fermentation process describing the transition between these two compounds. The effect of temperature on the model parameters is also included in this model. The sole difference between ASM2 and ASM2d is the denitrification capacity of phosphorus accumulating organisms that exists in the ASM2d alone. In these two models, the main drawback than with ASM1 is an increased number of parameters.

[1] No formal reference has been found on this topic but it has been confirmed by Willi Gujer itself.

Table 2.1: Matrix representation of ASM1 showing processes, components, process kinetics and stoichiometry for carbon oxidation, nitrification and denitrification, based on processes of growth and decay of bacteria, hydrolyses and ammonification.

Component i → / Process j ↓	1 S_I	2 S_S	3 X_I	4 X_S	5 X_{BH}	6 X_{BA}	7 X_P	8 S_O	9 S_{NO}	10 S_{NH}	11 S_{ND}	12 X_{ND}	13 S_{ALK}	Process rate, ρ_j [M.L^{-3}.T^{-1}]
1 Aerobic growth of heterotrophs		$-\frac{1}{Y_H}$			1			$-\frac{1-Y_H}{Y_H}$		$-i_{XB}$			$-\frac{i_{XB}}{14}$	$\mu_{mH}\left(\frac{S_S}{K_S+S_S}\cdot\frac{S_O}{K_{OH}+S_O}\right)X_{BH}$
2 Anoxic growth of heterotrophs		$-\frac{1}{Y_H}$							$-\frac{1-Y_H}{2.86\,Y_H}$	$-i_{XB}$			$\frac{1-Y_H}{14\cdot 2.86\cdot Y_H}-\frac{i_{XB}}{14}$	$\mu_{mH}\left(\frac{S_S}{K_S+S_S}\cdot\frac{K_{OH}}{K_{OH}+S_O}\cdot\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{BH}$
3 Aerobic growth of autotrophs						1		$-\frac{4.57-Y_A}{Y_A}$	$\frac{1}{Y_A}$	$-i_{XB}-\frac{1}{Y_A}$			$-\frac{i_{XB}}{14}-\frac{1}{7\,Y_A}$	$\mu_{mA}\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\cdot\frac{S_O}{K_{OA}+S_O}\right)X_{BA}$
4 'Decay' of heterotrophs				$1-f_P$	–1		f_P					$i_{XB}-f_P\,i_{XP}$		$b_H X_{BH}$
5 'Decay' of autotrophs				$1-f_P$		–1	f_P					$i_{XB}-f_P\,i_{XP}$		$b_A X_{BA}$
6 Ammonification of soluble organic nitrogen										1	–1		$\frac{1}{14}$	$k_a S_{ND} X_{BH}$
7 'Hydrolysis' of entrapped organics		1		–1										$k_h\frac{X_S/X_{BH}}{K_X+X_S/X_{BH}}\left(\frac{S_O}{K_{OH}+S_O}+\eta_h\frac{K_{OH}}{K_{OH}+S_O}\frac{S_{NO}}{K_{NO}+S_{NO}}\right)X_{BH}$
8 'Hydrolysis' of entrapped organic nitrogen											1	–1		$\rho_7\left(X_{ND}/X_S\right)$
Observed Conversion Rates [M.L^{-3}.T^{-1}]	$r_i=\sum_j \nu_{ij}\rho_j$													
	Soluble inert organic matter [M(COD).L^{-3}]	Readily biodegradable substrate [M(COD).L^{-3}]	Particulate inert organic matter (g(COD).m^{-3})	Slowly biodegradable substrate [M(COD).L^{-3}]	Active heterotrophic biomass [M(COD).L^{-3}]	Active autotrophic biomass [M(COD).L^{-3}]	Particulate products arising from biomass decay [M(COD).L^{-3}]	Oxygen (negative COD) [M(-COD).L^{-3}]	Nitrate and nitrite nitrogen [M(N).L^{-3}]	NH_4^+ and NH_3 nitrogen [M(N).L^{-3}]	Soluble biodegradable organic nitrogen [M(N).L^{-3}]	Particulate biodegradable organic nitrogen [M(N).L^{-3}]	Alkalinity – Molar units	

Stoichiometric parameters :

Heterotrophic yield: Y_H

Autotrophic yield: Y_A

Fraction of biomass yielding particulate products: f_P

Mass N / Mass COD in biomass: i_{XB}

Mass N / Mass COD in products from biomass: i_{XP}

Kinetic Parameters:

Heterotrophic growth and decay: μ_{mH}, K_S, K_{OH}, K_{NO}, b_H

Autotrophic growth and decay: μ_{mA}, K_{OA}, K_{NH}, b_A

Correction factor for anoxic growth of heterotrophs: η_g

Ammonification: k_a

Hydrolysis: k_h, K_X

Correction factor for anoxic hydrolysis: η_h

ASM3 (Gujer *et al.*, 1999) is a modification of the ASM1. The concept of death/lysis is replaced with endogenous respiration and internal cell storage of heterotrophic bacteria is added. A new component corresponding to this storage (X_{STO}) is added to the fractionation. This is meant to allow for a better estimation of the active biomass with a separation between autotrophic and heterotrophic biomasses. This separation can also make the calibration of ASM models easier. The addition of internal cell storage allows better representation of batch configurations (draw-and-fill reactors). Soluble and particulate organic nitrogen are removed from the fractionation as well as the corresponding hydrolysis and ammonification processes. This is related to the difficulty of measuring these two components, as well as the difficulty of estimating the associated kinetics. In this model, they are directly considered as soluble ammonia. Finally, inert compound X_I and inert compound arising from biomass decay X_P are grouped in a single component X_I.

ASM1, ASM2d and ASM3 processes are illustrated in Figure 2.2 and Figure 2.3. When "S_O/S_{NO}" is written in a block, it means that either oxygen or nitrate can be used as the electron acceptor for the reaction. The ASM2 schema is identical to that of ASM2d except that only oxygen is considered as an electron acceptor for phosphorus accumulating organisms in ASM2.

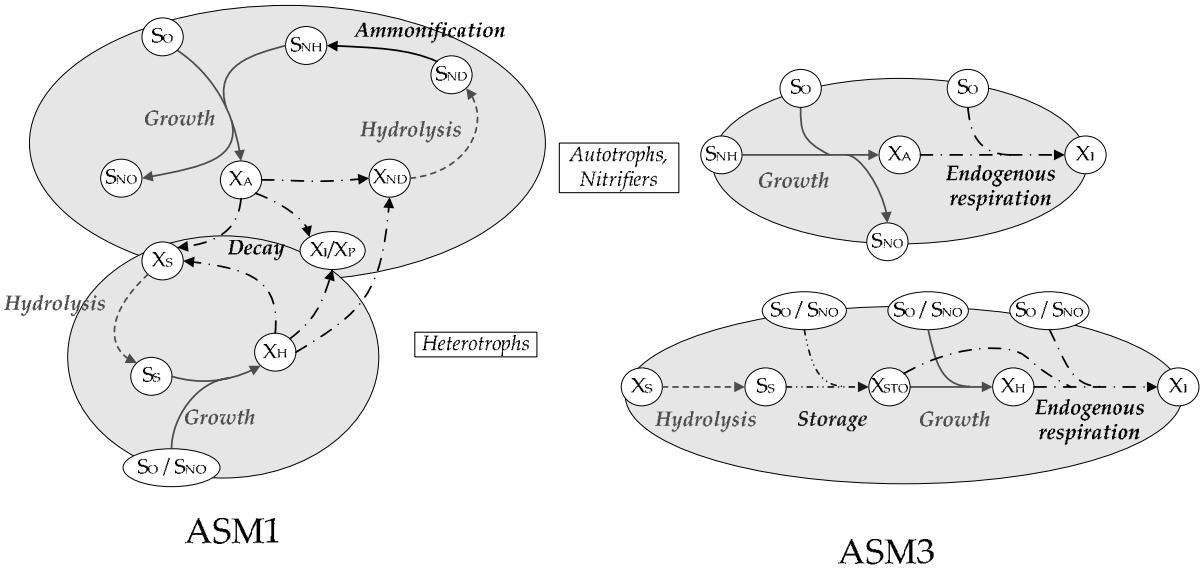

Figure 2.2: Main processes of ASM1 and ASM3 (adapted from Henze *et al.*, 2000)

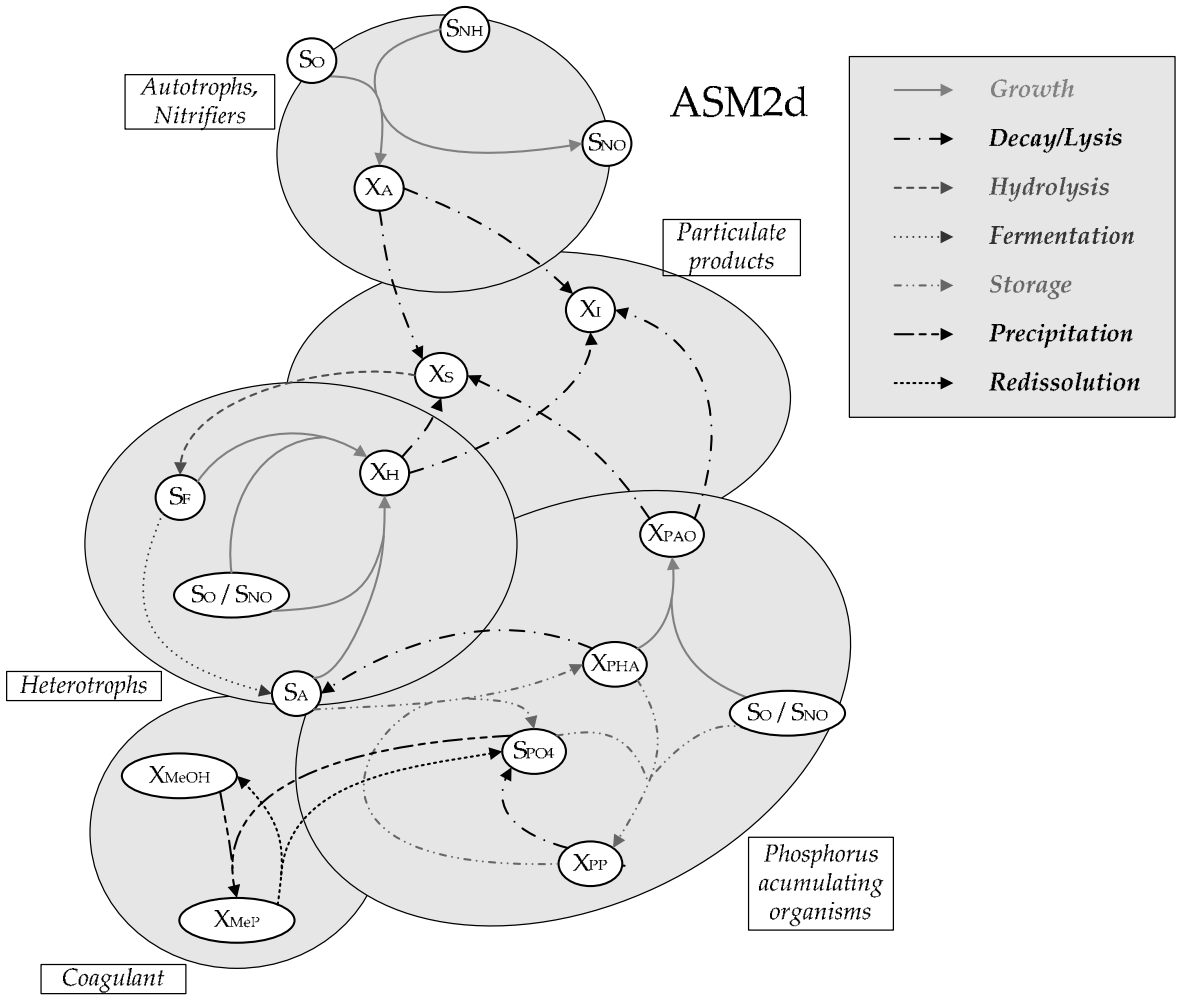

Figure 2.3: Main ASM2d processes

Finally, a new model was developed and presented by Printemps (2004) that is meant to better describe sludge production by considering the difference between the volatile suspended solids (VSS) and the inorganic suspended solids (ISS). This model is named ASMAnjou and is a direct adaptation of ASM1 with a finer fractionation of the influent for suspended solids and the inclusion of a slowly biodegradable soluble substrate that first has to be hydrolyzed before being biodegraded. This dynamic model is an adaptation of the SIMBAD model previously developed (Lesouëf, 1990; Gilles and Pellas, 2000).

All these models have adaptations that take into account the impact of the mixed liquor temperature on the process kinetics. This may be particularly important since biomass activity is quite limited when water temperature drops below 15°C (and such temperatures are typically encountered in winter in Europe).

One key issue in the use of all these models is that we cannot measure all the compounds of the wastewater directly and they have to be expressed as a fraction of other overall parameters such as COD or TSS. The other main issue is the need to calibrate the parameters of these models to each individual case study, mainly due to the varying performance of the bacteria consortium under different operating conditions.

2.2.2 Clarifiers

The other main wastewater treatment process required with activated sludge processes is clarifying of water. This is used at the end of the secondary treatment to separate soluble compounds that go to the receiving body from particulates compounds (including the biomass) that are re-circulated in the activated sludge units or withdrawn when in excess. Many different kinds of models are available for settlers, ranging from the simplest 0-Dimensionnal models to the most complicated 3-Dimensionnals models.

0-Dimensionnal Conceptual Models

Conceptual 0-Dimensionnal models are used for the design of settling tanks and is based on the overflow rate concept introduced by Hazen (1995). However, the traditional Hazen model is applicable only to primary settling tanks. More recently, Krebs *et al.* (2000) introduced a conceptual model for secondary settling tanks, which is based on a linear solids concentration profile in the solids blanket. The main limitation of these models is that they cannot easily be applied for control and operation of a WWTP; they are suitable only for its design.

0-Dimensionnal Flux Models

The limiting flux theory is based on the consideration that the settling velocity of particles is related to their current concentration. This is linked with the vertical gradient of concentration which is always present. As concentrations of particles below a given point are higher than the concentration of particles at this point, maximum possible velocity is limited. The classic flux theory was introduced by Kynch (1952) and returns a hindered solids settling velocity. The main limitation of these models is that the theory only provides an underflow solids concentration.

1-Dimensionnal Layered Flux Models

1-Dimensionnal layered models are based on discretization of the settler in a set of horizontal layers. In each layer, the concentration of sludge is assumed to be constant and only the transfers between the layers is described. The theory of the limiting flux is used in these models. Two fluxes therefore cause the sludge concentration to change (see Figure 2.4). First, a settling flux represents the settling of the sludge using the hindered solids settling velocity of the limiting flux theory. Second, a transport flux (or advection flux) takes into account the water velocity across layers. Many models are available in this area and the review of Grijspeerdt et al. (1995) showed that the one of Takács et al. (1991) is the most reliable for dynamic simulations of secondary settlers.

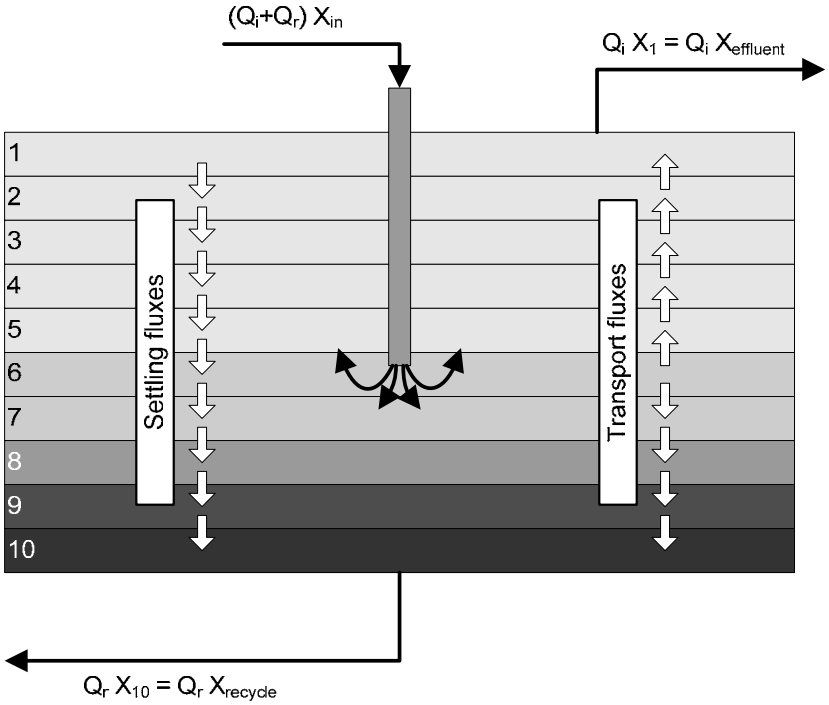

Figure 2.4: 1-Dimensionnal layered models of clarifiers

Adaptations of this model have been proposed to include the biological reactions occurring in secondary clarifiers (Gernaey *et al.*, 2006a). Two solutions are studied: inclusion of an activated sludge model for each layer of the secondary clarifier or addition of a model for depletion of oxygen and a defined fraction of nitrate in the underflow flux (these reactions mostly occur at the bottom of the clarifier where there is enough activated sludge). In this

work, the second solution was found to be the best compromise between computation time and model accuracy.

The main advantages of the layered models are their limiting computer complexity and their possible coupling with activated sludge models. Their main limitation is that they have to be calibrated with a more complex model (2D or 3D). The choice of the number of layer has a particular impact on the parameters of the model. Another limitation is the low accuracy of the estimation of particle concentrations in the effluent.

1-Dimensionnal Advection-Dispersion Models

Instead of using a limited number of layers for settling tank modeling, the 1D advection-dispersion equation can be solved numerically. This partial differential equation is (Ekama *et al.*, 1997):

$$\frac{\partial X}{\partial t} + u\frac{\partial X}{\partial y} + \frac{\partial v_S X}{\partial y} - D_C \frac{\partial^2 X}{\partial y^2} = 0 \qquad \text{(II.1)}$$

where X is the solids concentration (in $g.m^{-3}$), y the vertical coordinate (in m), u the bulk liquid velocity (in $m.s^{-1}$), v_S the settling velocity of solids (in $m.s^{-1}$) and D_C is the dispersion coefficient (in $m^2.s^{-1}$). The main limitation of this model is the required solving of a partial differential equation, which usually avoids its combination with activated sludge models for dynamic simulations (mainly due to simulation software issues).

2 and 3-Dimensionnal Models.

In a real plant, there are many factors influencing the performance and the capacity of the settling tank. For example, many boundary and flow conditions cannot be reflected in 1D models. There are four categories of unconsidered influences (De Clercq, 2003):

- geometry, e.g. shape of the basin, inlet and outlet arrangements, and baffles
- flow, e.g. density effects causing non-uniform velocity profiles, potentially resulting in turbulence and/or short-circuits from the inlet to the outlet
- solids removal mechanism, which results in many unsteady effects
- environmental conditions, e.g. wind shear, air and inlet water temperature

The prediction of the settling tank performance with 0D and 1D models is therefore a matter of calibration. Hence, more advanced models are needed. 2D and 3D models have the potential to describe the internal flow patterns. Their application is mainly related to the evaluation of internal design changes such as the addition of baffles, to the simulation of rectangular settlers and to the calibration of 0D and 1D models (Weiss *et al.*, 2006). The disadvantage of these models is their very high computational demand and that they cannot therefore be used for control evaluation purposes.

2.2.3 Digesters

Primary and secondary treatments of WWTP produce sludge in large quantities. In order to reduce the cost and the environmental impact of this sludge production, an efficient treatment should reduce this quantity and/or produce energy. Anaerobic digestion is the perfect process to this end. It is a process which can sustain high loading rates with a low sludge production. It also produces biogas which can replace fossil fuel sources for the production of energy. It has therefore a positive impact on greenhouse gas production and is becoming very popular.

Nowadays, anaerobic digestion is, in fact, one of the oldest biological process used for the treatment of organic matter from food and beverage processing. It is now largely used in several industrial wastewater treatments (chemistry, pulp and paper). The need for a model of this process was publicly identified in 1997 at the 8th IAWQ Anaerobic Digestion Conference in Sendai, Japan. A task group was formed and, as a result of this work, Anaerobic Digestion Model No. 1 (ADM1) was published (Batstone *et al.*, 2002). It is now a widely recognized dynamic model of anaerobic digesters.

ADM1 includes biochemical processes (catalyzed by intra- and extra-cellular enzymes produced by available micro-organisms) as well as physico-chemical processes (ions association/dissociation and gas-liquid transfer). Precipitation is not included in the model. The flow of organic compounds in the model as well as the various processes are represented in Figure 2.5. For clarity, the various micro-organisms involved are not represented.

The first process is the disintegration of composite compounds into inerts (i.e. compounds not degraded during digestion), carbohydrates, proteins and lipids. This step is largely non-biological. The second is the enzymatic hydrolysis of carbohydrates, proteins and lipids into

respectively monosaccharides, amino acids and long chain fatty acids (LCFA). The third step is the action of acidogens, which degrade monosaccharides and amino acids into acetic, propionic, butyric, valeric acids and hydrogen. The fourth is the action of acetogens which degrade propionic, butyric, valeric acids and LCFA into hydrogen and acetic acids. Finally, two methanogenic groups degrade the acetic acid and the hydrogen into methane. All these reactions are first order kinetics, regulated by Monod-term related to the available substrate and inhibitions of pH, hydrogen and free ammonia.

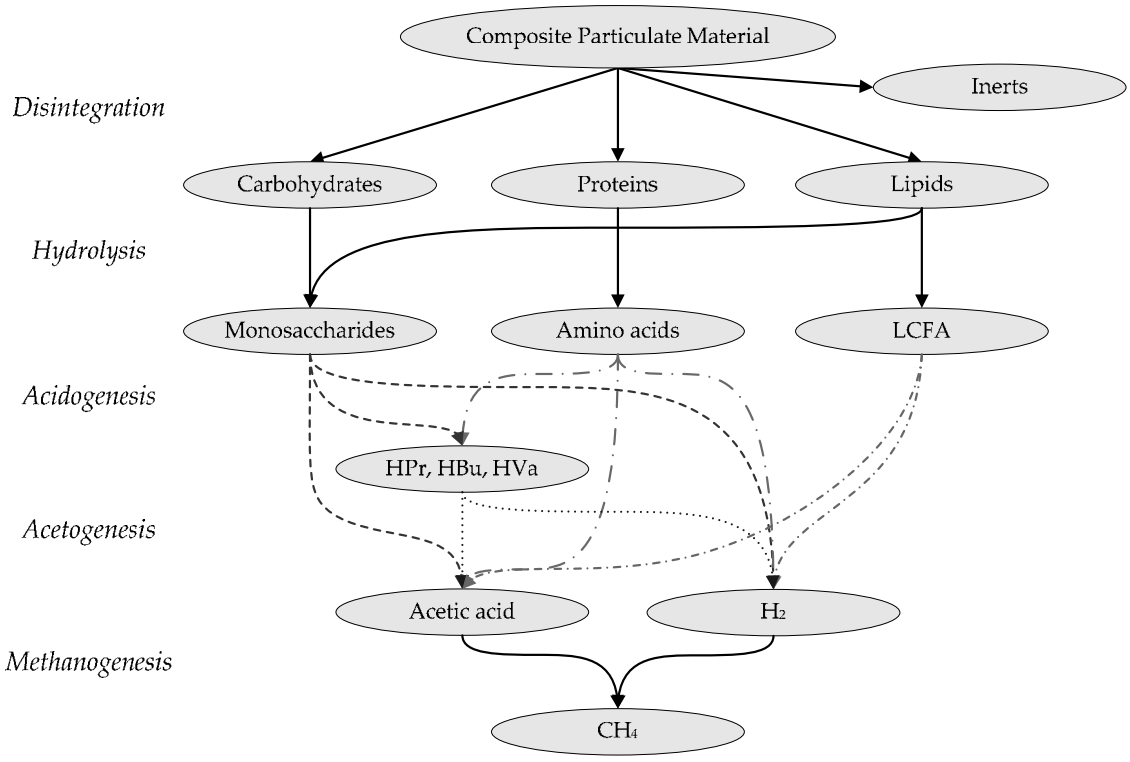

Figure 2.5: Flux of organic compounds in ADM1 (from Batstone *et al.*, 2002)

The reactions not represented on this figure (but included in ADM1) are death of the microorganisms (first order kinetics), liquid-gas transfers, production/consumption of carbon dioxide and acid base reactions (pH, free ammonia and carbon dioxide).

2.2.4 Plant-wide models

Combining all the models previously exposed (i.e. activated sludge units, settler and anaerobic digester) is becoming popular in order to gain more insight into the different couplings involved in a WWTP using all these processes. However, due to the different

fractionations of wastewater in ASM models and in ADM1, a direct link between models is not possible.

Two approaches have been proposed to address this coupling. The “supermodel” approach (Seco *et al.*, 2004; Jones and Takács, 2004) consists in developing a common fractionation for both ASM and ADM models. Traditional models have then to be modified to work with this new fractionation. This approach is convenient as no transformations are required between the various unit process models. Its main drawback is a lack of flexibility to add or to remove some components of the fractionation depending on the local case study. For instance, if a new ADM model is developed with a new fractionation, a new “super-fractionation” is required and all models (e.g. ASM, settler) will have to be modified accordingly. Another drawback is that resulting models are very big and many states are used in each unit process models even if they are not required.

An extension of this supermodel approach is proposed in Grau *et al.* (2007). Instead of composing *a priori* the required fractionation, a list of possible transformations and components is provided. When a WWTP model is conceived, each unit process model is built with the corresponding transformation of the previous list. The whole model WWTP is then built and interfaces are automatically constructed, based on the selected list of transformations. The advantage of this technique is that charge balance can be considered in the whole WWTP model instead of calculating it only in the interfaces as proposed in the second approach below. The main drawback of this technique is that very advanced model knowledge is required in order to be able to choose the required transformations. This drawback may however be easily solved in the future with an advanced computer aided model building tool.

The second approach to address the coupling is based on the use of existing models and the concept of interface transformations that convert one fractionation to another one (Copp *et al.*, 2003; Zaher *et al.*, 2007; Nopens *et al.*, 2009). These transformations are based on equations of continuity of the mass and charge balances. Attention has been limited to the ASM1-ADM1 and ADM1-ASM1 interfaces so far. This technique has the advantage of using existing models without any change. Its main drawbacks are the very complex conception of these transformations and the fact that they are devoted to specific interfaces.

Plant-wide modeling is hence still the subject of discussions in the scientific community and no consensus on the simplest modeling approach has been found so far. For now, the approach used by individual researchers is still the subject of personal preferences and also modeling software capabilities (depending on the software used, not all approaches may be available).

2.2.5 The Oxidation-Reduction Potential

The oxidation-reduction potential (ORP) is often used as a measurement of the state of the activated sludge (Meijer, 2004). It represents the ability of water compounds to accept or donate electrons. Its measurement is based on the use of two half-cells (electrode + aqueous solution), one having precise characteristics, used as a reference, the other one being unknown. In activated sludge treatments, one half-cell is the activated sludge mixed liquor. Platinum is typically used as an electrode in such case in order to ensure reliable measurements for continuous operation of the sensor in the activated sludge, which is an aggressive environment. The other half-cell is usually the reference couple H^+/H_2.

The theory of ORP measurements is clearly defined for aqueous systems with two species: the ORP is the result of the two half reactions of oxidation and reduction occurring. Considering a half-equation for a single couple (or half-cell):

$$x \cdot Ox + n \cdot e^- \leftrightarrow y \cdot Red \tag{II.2}$$

The Nernst equation describes the potential corresponding to this couple:

$$E = E^0 + \frac{R \cdot T}{F \cdot n} \ln \frac{A_{ox}}{A_{red}} \tag{II.3}$$

where E is the half-cell potential, E_0 is the standard potential when both concentrations are at their equilibrium, n is the number of electrons exchanged, R is the universal gas constant (8,314472 $J.mol^{-1}.K^{-1}$), T is the absolute temperature (in K), F is the Faraday constant (96485 $C.mol^{-1}$) and A_{ox} and A_{red} are the activities of the relevant chemical species. For perfect solutions at low concentrations, the activities tend to equal the concentrations of the species.

When two half-cells are combined as a cell, the reaction occurring can be formulated as:

$$a \cdot Ox_1 + b \cdot Red_2 \rightarrow a \cdot Red_1 + b \cdot Ox_2 \quad \text{(II.4)}$$

Indices 1 and 2 reference the two half-cells and *a* and *b* are the stoichiometric coefficients of the reaction. The potential measured *E* between the two cells is then:

$$E = \frac{b \cdot E_1 + a \cdot E_2}{a + b} \quad \text{(II.5)}$$

where E_1 and E_2 are the potentials of half-cells 1 and 2 (Eq. II-3), and *a* and *b* are the stoichiometric coefficients of previous equation II.4.

In the case of the activated sludge, many half couples are present in the mixed liquor. The ORP is therefore the net result of all half-reactions of all compounds of the activated sludge.

Attention should be put also on the fact that the equation for ORP prediction in pure systems is based on the assumption that the system is in a chemical equilibrium. However, in the case of continuous activated sludge systems, this is never the case. The measurement is in fact affected by the electron exchange density of the platinum electrode, which is linked to the ease with which electrons can be exchanged at the platinum surface. The measurement is therefore affected by the state of the surface of the electrode. Platinum is almost not affected by the activated sludge but many substances deposit on the electrode surface and disturb the electronic exchanges. Finally, the value of the ORP is influenced by the pH, which is also influenced by the biological and chemical reactions occurring in the activated sludge.

As a conclusion of all these previous considerations, the ORP cannot be considered as an absolute measurement and typical values are not available. It was however the most reliable measurement technique ten years ago and the variations of this signal are very informative: during the nitrification, the ORP increases whereas it decreases during denitrification. If the nitrate is completely depleted at the end of denitrification, a knee can be observed on the ORP curve (change of the sign of the second derivative of the ORP signal) (Wareham et al., 1993; Sasaki et al., 1993; Cecil, 2008a). First advanced control laws for the aeration of the activated sludge were therefore based on this measurement.

Simulation and optimization of these control laws would be needed, at least for purpose of comparison. An ORP measurement model would therefore be necessary based on the ASM

fractionation of mixed liquors. For now, mostly only “inverse” models of the ORP are available in the literature (estimation of ammonia and nitrate concentrations based on ORP and DO measurements for instance, Spérandio and Queinnec, 2004). Direct models of ORP developed in the literature are based only on batch processes where the end of nitrification and denitrification are reached and a knee can be observed in the ORP signals (Cecil, 2008a) or in the O_2 signal considering only nitrification (Héduit and Thévenot, 1989).

As no reliable model of ORP measurement is available, the simulation of control laws based on this parameter is very difficult and subject to many uncertainties. A proposition of a simple model will however be made in chapter 5, based on the real case study of Cambrai, in order to allow very simple simulations and performance evaluations.

2.3 Aeration control strategies for WWTP activated sludge units

The main operation that needs precise and dynamic control in a WWTP is the aeration of the activated sludge unit(s). This is due to the two reactions of nitrification and denitrification that typically occur in the same tank, but the first one needs oxygen (aerobic conditions) while the second one needs its absence (anoxic or anaerobic conditions). Depending on the plant layout, treatment objectives and instrumentation available, different types of controller can be used. The main control laws used in WWTPs will be presented in the following sections.

2.3.1 Simple control based on time

The simplest way to control the aeration is based on time. Aeration phases are defined based on a daily timetable. Additional timetables are typically used for weekends and for periods of high or low loads (holidays, etc.). This kind of control is robust since no specific control law nor instrumentation is required but it is not robust in terms of performance, as no measurement of the real incoming load and process state is performed.

Such a system is applied on old extensive processes. In these WWTPs, the water residence time in the activated sludge units is long, usually close to 24 hours. This induces small impact of the inlet variation on the concentrations in the activated sludge units. When changes occur, they are slow and the control can be adapted by operators on a daily basis. This control is therefore also reliable in term of performance on these extensive processes.

Such a system does not need a high level of expertise for the plant manager who merely needs to adjust the timetable according to the variation of the water quality daily measured in the outlet for the self-control procedure of WWTPs.

2.3.2 Classic ORP control

The simplest way to control the aeration with an ORP measurement is to use two levels of ORP for the start and the stop of the aeration (Charpentier, 1998). The aeration starts when the ORP is lower than the low level and stops when the ORP is higher than the high level. Levels can even be manually adjusted to reflect changes in the incoming wastewater or in the process conditions.

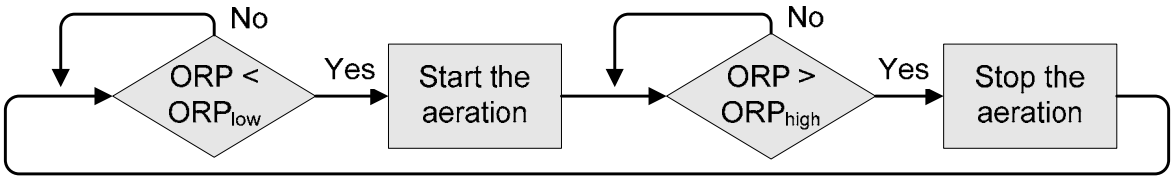

Figure 2.6: Flowchart of the ORP controller

During the aeration phase, a second control is typically used to ensure correct oxygen concentration. This is meant to provide an adequate balance between the nitrification rate and the energy cost. This also allows one to control the necessary air flow rate and to avoid too much aeration during the night or too low aeration during the day. A typical set point value used in WWTPs is 2 $g.m^{-3}$, based on operational experience. This oxygen control may be required in order to avoid problems of bulking or foaming, mainly due to filamentous bacteria development at low oxygen levels. Proportional-Integral (PI) controller is used for this control.

2.3.3 Regul'N©

A solution named Regul'N© (Charpentier 1992, Charpentier et al. 1998) is available that allows automatic adjustment of the ORP levels. This solution is based on an additional measurement of the ammonia concentration at the outlet of the activated sludge unit (or the outlet of the wastewater treatment plant). Adjustments of the levels are based on a moving average (usually with a window of 24 hours but this parameter can be adjusted), with only one

modification of levels per 24 hours (this parameter can also be adjusted). This solution can even be adapted to use only the daily measurements made with grab samples for the monitoring of the WWTP. The flowchart for this operation is illustrated in Figure 2.7.

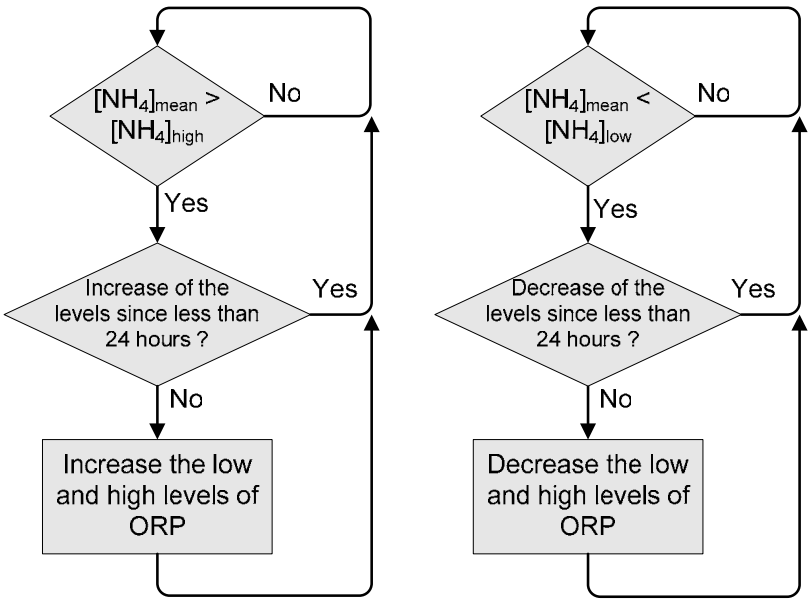

Figure 2.7: Flowchart of the adjustment of ORP levels

When the moving daily average of ammonia goes above the upper bound, both levels of ORP are increased and they are decreased when the moving daily average of ammonia goes below the lower bound. Time securities are present to avoid more than one change per 24 hours. The value of the ORP levels modifications is subject to parameterization on each WWTP. The use of a moving daily average allows the use of sensors with low reliability. The oxygen is also controlled during the aeration phase when the air production system is adequate for this control. From practical experience, this solution is recognized to be more robust than fixed levels.

2.3.4 Control based on levels of NH_4/NO_3 concentrations

Due to the recent progress in instrumentation, reliable and cheap sensors for in-situ measurements of ammonia and nitrate are now available (Jeppsson et al., 2002; Ingildsen and Olsson, 2002; Ingildsen and Wendelboe, 2003; Kaelin *et al.*, 2008). They typically require less than one calibration per month, less than one cleaning per week and cost less than 5,000 euros (with transmission equipments included).

Control laws based on these measurements are therefore emerging and WWTP operations increasingly rely on them. The simplest control laws using these parameters are based on ammonia level (Figure 2.8) or mixed levels of ammonia and nitrate (Figure 2.9). A high ammonia level starts the aeration in both cases. In the first case, a low ammonia level is used to stop the aeration while, in the second case, a high level of nitrate is used.

Time securities are usually included to avoid phases that would be too short or too long. Oxygen is usually controlled during aeration phases just as in the ORP case. Another possibility would be to exclude oxygen control and use this control for a degraded functioning of the simultaneous nitrification/denitrification (SNDN) control law. This may be necessary when the air production system is not adequate for true SNDN. In these control schemes, the parameters that can be optimized are the levels used to start and stop aeration.

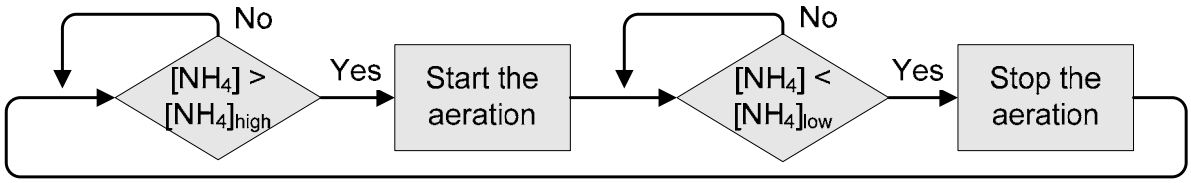

Figure 2.8: Flowchart of an NH_4 controller

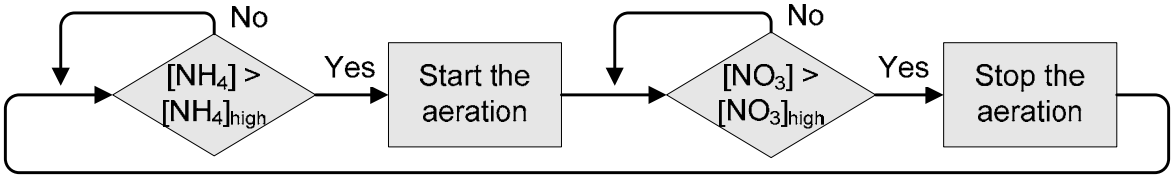

Figure 2.9: Flowchart of an NH_4 / NO_3 controller

2.3.5 STAR© / AMSTAR aeration module

More advanced control laws are also available in practice. A very efficient one is part of the STAR© system (Nielsen *et al.*, 1995; Printemps *et al.*, 2006). It provides an adequate solution to the fact that the phases should be long when there is a high amount of pollution and can be reduced when there is almost no more pollution. The basis of this controller is a phase diagram (see Figure 2.10).

This diagram has the nitrate concentration on the horizontal axis and the concentration of ammonia on the vertical axis. The concentrations are measured in the activated sludge tank. In this diagram, two curves are used to control the turning on and off of the aeration. The functioning of the phase diagram is the following: when the point corresponding to the current concentrations of nitrate and ammonia goes above the curve f_{DN}, aeration is started and nitrification takes place. Then, when the current point goes below the curve f_N, aeration is stopped and denitrification takes place.

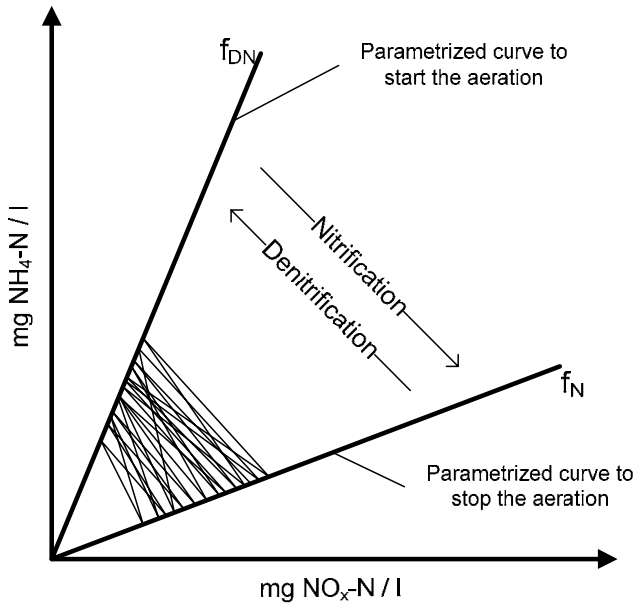

Figure 2.10: Phase diagram of the STAR© controller

Oxygen control is also typically used in this scheme during the aeration phases, just as with ORP control. An example of the combination of the phase control with this oxygen control is given in Figure 2.11.

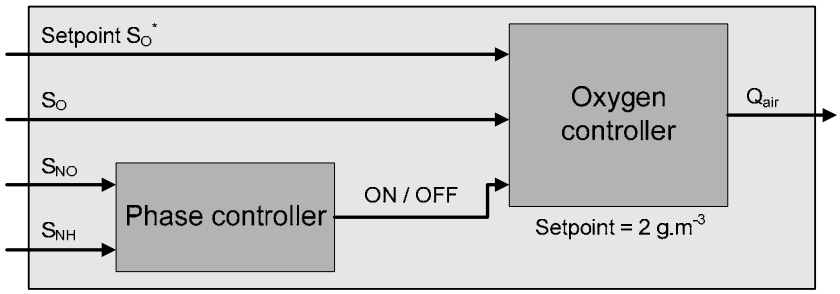

Figure 2.11: Combination of a STAR© phase controller with an oxygen controller

The parameters that can be optimized in this control law are the two curves f_N and f_{DN}. Simple lines can be used as in the example above or more advanced parameterization can be considered. If simple lines are used, their slopes as well as the point where they cross can be optimized (for instance if a plant is overloaded, the optimal crossing point is probably located at high concentrations of ammonia and maybe also nitrate).

For clarity, as the whole STAR system is not considered here but only its aeration module, this control law will henceforth be referred to as AMSTAR.

2.3.6 Control for simultaneous nitrification and denitrification

A final advanced control law based on measurement of ammonia and nitrate is simultaneous nitrification and denitrification (SNDN) (Lemoine and Grelier, 2008; Thauré *et al.*, 2008). In contrast with previous schemes which are all based on sequences of aeration and non-aeration, this scheme tries to provide continuous aeration during the whole day. This control strategy is possible because of the bacteria's adaptation to low oxygen concentrations, which allows simultaneous occurrence of both nitrification and denitrification reactions. The advantage is the reduction in the energy used for the aeration of the activated unit, as well as a more accurate control of the concentration of pollutants. This control law, however, usually requires the upgrade of the aeration system to allow variable air flow rate (even if a lower limit is always present).

The structure of the controller is described in Figure 2.12. The oxygen concentration is not used in this scheme because its very low value is not precisely detected by sensors and the measurement is not reliable anymore. Only measurements of ammonia and nitrate are thus used by two controllers in cascade. The first controller regulates the nitrate concentration (S_{NO} in $g.m^{-3}$) by manipulating a set point of ammonia (S_{NH} in $g.m^{-3}$) while the second controller regulates the ammonia level by manipulating the air flow rate (Q_{air}, in $Nm^3.d^{-1}$).

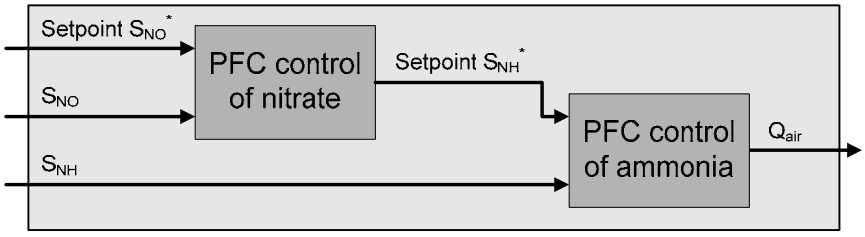

Figure 2.12: Control scheme of simultaneous nitrification/denitrification with continuous aeration

The controllers used are first-order Predictive Functional Controllers (PFCs). PFCs are an implementation of Model Predictive Controllers originally proposed by J. Richalet *et al.* (1978). They are based on an internal model of the process that ensure fast disturbance rejection and robustness around a nominal point of functioning (for more information about this topic, readers are referred to the book of Camacho and Bordons (2004)). In our case, the system behavior around the current functioning points can be approximated with first order linear models (transfer functions between nitrate and ammonia and between ammonia and oxygen transfer coefficient). This is a big simplification but it ensures simple parameterization of the control law and still provides very good performance. Even with these simple models, the parameterization on real processes is not easy as the incoming pollution is continuously disturbing the system and the gain and time constants of each transfer functions are not easy to identify.

The parameters that can be optimized in this control law are (i) the set point of nitrate and (ii) the upper and lower limits of ammonia. These two limits are necessary to avoid too high ammonia level (which may occur during daily peak of incoming load) and too low ammonia level (which are not necessary to reach due to the poor reaction efficiency at low ammonia level). These two limits will allow the nitrate level to be reduced during the night when low levels of ammonia occur and it will increase during the peak of ammonia.

2.3.7 Conclusion on control laws available in practice

The most typical control laws for the aeration of the activated sludge units have been presented in this section. In order to focus only on promising control laws that can be optimized for the long term, only the controls based on ORP or NH4 and/or NO3 measurements will be considered. Due to the very difficult modeling of ORP and the limited range of validity of such models, control laws based on this measurement can only be simulated as a reference. Their optimization would probably not be adequate due to modeling uncertainties. This study will therefore focus on controls based on measurements of ammonia and nitrate. In the first application on a literature case study in chapter 4, SNDN and AMSTAR will be optimized. In the second application on the case of Cambrai in chapter 5, a control based on levels of ammonia (SABAL, degraded SNDN due to current air production

sizing) and SNDN will be optimized. Reference performance of ORP and SABAL controls will also provided in this last application.

Literature case studies available for the methodology development that will be presented in chapter 4 as well as most promising objectives to consider will now be detailed.

2.4 Benchmark simulation models

Due to the availability and common acceptance of previously exposed WWTP models, it is possible to conduct reliable offline evaluations of the WWTP control laws presented in the next section. However, there was until recently a lack of a common WWTP design to test all control laws with the same WWTP model and evaluation criteria. Two benchmark simulation models (BSM1 and BSM2) have been developed by the scientific community to tackle this need and they are presented in this section. The performance criteria associated with these models are then detailed. Finally, typical functioning of SNDN and AMSTAR control laws on BSM1 is detailed.

2.4.1 Benchmark simulation model #1

BSM1 was developed in the framework of the COST Action 624 (Copp, 2002). This model represents the secondary treatment of a WWTP with five activated sludge units (ASUs) in series combined with a single secondary settler (Figure 2.13).

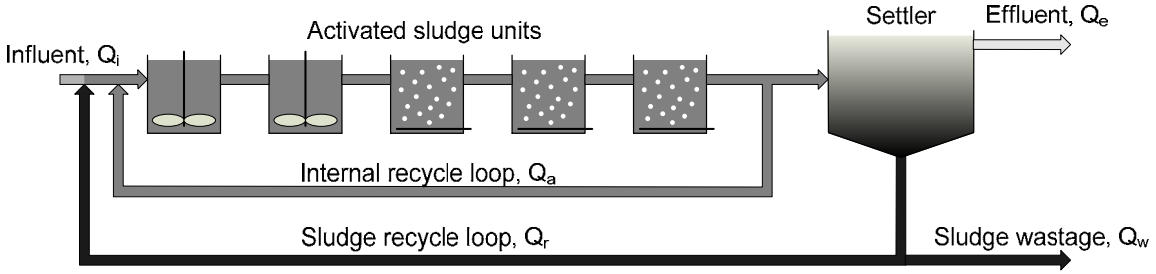

Figure 2.13: Schematic representation of "Benchmark Simulation Model 1" plant layout (Copp 2002)

The first two reactors are anoxic while the tree last ones are aerated (allowing either an aerobic or anoxic state depending on the chosen control law). The volumes of the anoxic and aerated tanks are respectively 1000 m^3 and 1333 m^3 each. The volume of the secondary settler is equal to the total volume of the activated sludge units, 6000 m^3. A nitrate internal recycle loop and a sludge recycle loop are used. The nitrate recycle loop is used to make

denitrification occur in first reactors where COD is still present in sufficient quantity. The process models used are the Activated Sludge Model 1 (Henze *et al.*, 1987) for the reactors and the model of Takács *et al.* (1991) for the secondary settler.

Three typical influent datasets are also supplied with the report of Copp and illustrated in Figure 2.14. The first influent dataset is a typical dry period of 14 days. The second one represents a period of 14 days including a period of two days of rainy weather, during which the loads remain almost the same. Only the concentrations are decreased (and the influent flow rate increased). The third file represents a period of 14 days containing a storm event of short duration but high intensity. During this event, the sewer is expected to be flushed of particulate material and an increase in TSS fluxes occurs at the beginning of the event. These three periods of data enable the evaluation of control strategy performance requiring only limited computing time.

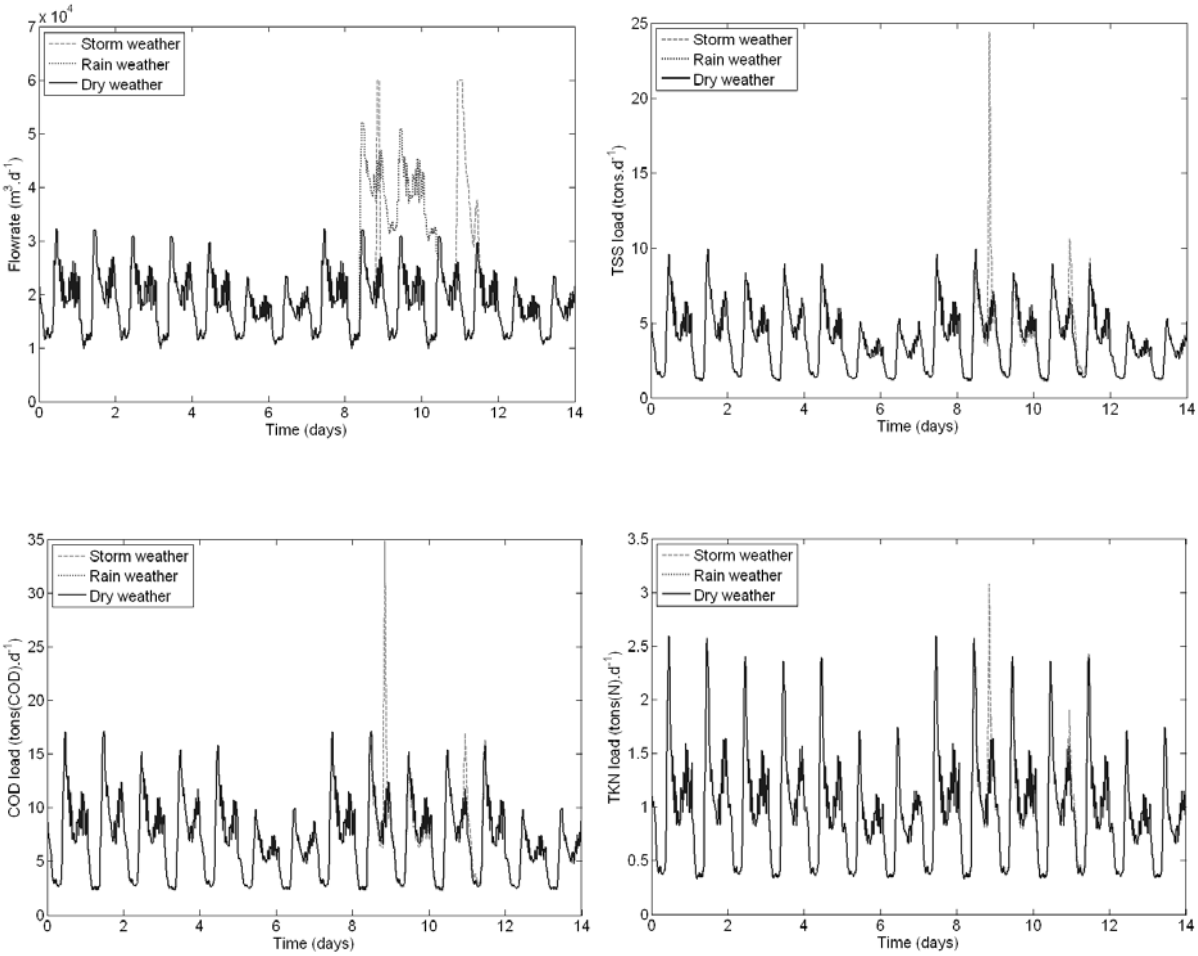

Figure 2.14: Influent datasets provided with BSM1

Average values are also provided for steady state computations that have to be performed before dynamic simulations. These values are averages of flow rate and fluxes of the dry weather dataset and are summarized in Table 2.2 (the whole ASM1 fractionation is given in BSM1 but not represented here for the sake of clarity).

Table 2.2: Average flow and loads for the stabilization period

Variable	Average value
Flow rate	18446 $m^3.d^{-1}$
TSS	3.897 $tons.d^{-1}$
COD	7.031 $tons(COD).d^{-1}$
TKN	1.004 $tons(N).d^{-1}$

The simulation procedure proposed for each control law consists in performing 100 days of steady states evaluation, followed by 14 days of dry weather dynamic dataset and finally 14 days of the chosen weather dataset. Only the last 7 days are used for performance evaluation.

After the publication of Copp's book and especially during the development of BSM2, small additions and corrections were made to BSM1. They are summarized in Rosen *et al.* (2005) considering the long-term modifications (BSM1_LT) and in the report of Alex *et al.* (2008) considering overall modifications of the framework. BSM1_LT concerns the simulation of 609 days of operations for the performance evaluation, as well as the addition of sensors and actuators models to add realism to the simulation (i.e. to avoid overly good results that would never be possible in reality). Temperature modifications are also included in the influent dataset (composed mainly of seasonal variations, but also including smaller daily variations). The activated sludge model is modified in order to take into account the impact of these temperature variations on the model kinetic parameters. The second report concerns the modification of the aeration performance criterion calculation (in order to adapt it to other volumes than BSM1 aerated tanks) and the addition of criteria for the mixing energy (required during low aeration phases) and the cost of carbon addition (used as external source of carbon for denitrification), as well as a total operating cost index (a weighted sum of previous operational criteria).

2.4.2 Benchmark simulation model #2

BSM2 was developed by the IWA Task Group on Benchmarking of Control Strategies for WWTPs (Jeppsson *et al.*, 2007 ; Nopens *et al.*, 2008). This new model is an extension of BSM1 with primary and secondary sludge treatments. The design of the activated sludge part has also been updated to fit WWTP design guidelines from both ATV (Abwassertechnische Vereinigung, German agency for wastewater treatment) and EPA (US Environmental Protection Agency). The full layout is presented in Figure 2.15. The activated sludge model has been modified to include the impact of temperature on kinetic parameters.

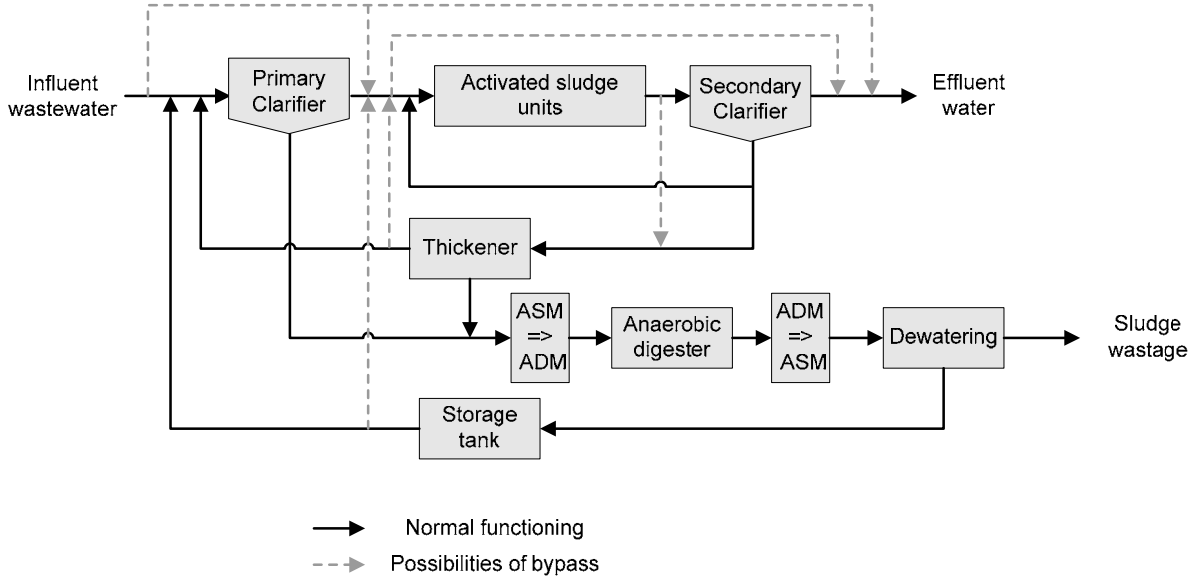

Figure 2.15: Layout of Benchmark Simulation Model 2 (from Jeppsson *et al.*, 2007)

The simulation of this model is based on datasets representing two years of operation. These datasets have been generated by means of a phenomenological model to avoid repetition of the same time series but to still have an accurate representation of events occurring at the WWTP inlet (Gernaey *et al.*, 2006b) (see Figure 2.16). The generation is based on daily, weekly and yearly profiles of flow rate and pollutants fluxes from households and industries, as well as random rain generation. All these datasets are then passed through a sewer model containing a module for first flush effect generation (i.e. the increase in TSS concentrations at the beginning of storm event due to the deposit of materials in the sewer during dry weather, even if there is not yet consensus in the scientific community about this effect). Yearly temperature variations are also included.

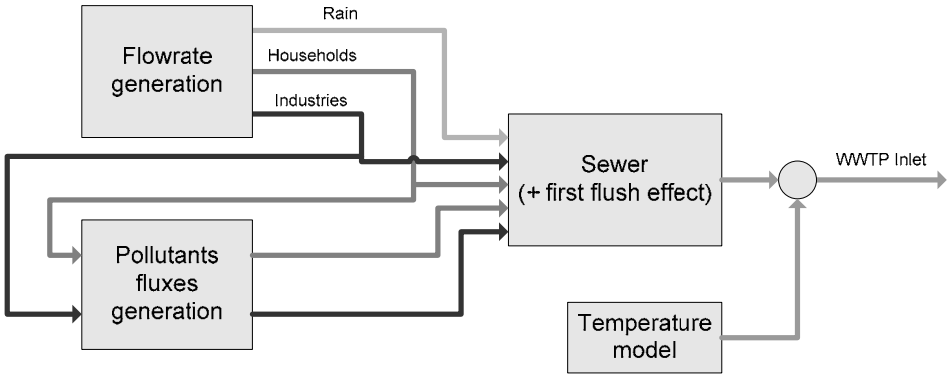

Figure 2.16: Influent generation model (from Gernaey *et al.*, 2006b)

Finally, a risk assessment module has been added to take into account various biological malfunction risks not taken into account in mass balance models (Comas *et al.*, 2006). This concerns the risk of foaming and bulking in the activated sludge, as well as the risk of rising sludge in the secondary clarifier. The causes associated with each of these risks are summarized in Table 2.3.

Table 2.3: Causes identified in the risk assessment module

Risk	Cause(s)
Bulking	Low DO in aerated basins
	Nutrient deficiency (high ratio BOD_5/N or BOD_5/P)
	Low organic loading
Foaming	Low F/M ratio
	High fraction of readily biodegradable organic matter
Rising sludge	Critical nitrate concentration in the inlet of secondary clarifier

2.4.3 Objectives to consider in BSMs

Many objectives can be considered during optimization of a WWTP and are proposed in BSMs. They are listed below, along with an additional personal criterion.

Effluent quality

The effluent quality (*EQ*, in kg(pollution unit).d^{-1}) is defined in BSM1 as a weighted sum of the effluent loads of total suspended solids (TSS_e, in g.m^{-3}), chemical oxygen demand (COD_e,

in g(COD).m^{-3}), biochemical oxygen demand (BOD_e, in g(COD).m^{-3}), total Kjeldahl nitrogen (TKN_e, in g(N).m^{-3}) and NO_x nitrogen ($S_{NO,e}$, in g(N).m^{-3}) in the effluent (Copp, 2002):

$$EQ = \frac{1}{T \cdot 1000} \cdot \int_{t_0}^{t_{7\,days}} \begin{pmatrix} 2 \cdot TSS_e(t) + 1 \cdot COD_e(t) \\ + 2 \cdot BOD_e(t) + 20 \cdot TKN_e(t) \\ + 20 \cdot S_{NO,e}(t) \end{pmatrix} \cdot Q_e(t)dt \qquad \text{(II.6)}$$

Aeration energy

Aeration energy (AE, in kWh.d^{-1}) is related to each plant design, depending on the aeration technique chosen as well as the performance of the material used. The original formula proposed in BSM1 is valid only for 1333 m^3 tanks. This computation is based on the K_La in each aerated reactor (in d^{-1}), T being the duration of the evaluation (in d) and $S_{O,sat}$ the concentration of oxygen at saturation (in g(-COD).m^{-3}):

$$AE = \frac{1}{T} \cdot \int_0^T \sum_{i=3}^{i=5} 0.4032 \cdot K_L a_i(t)^2 + 7.8408 \cdot K_L a_i(t)dt \qquad \text{(II.7)}$$

This formula has now been recently updated by the BSM2 Task Group to be simpler, to work for all size of aerated units and to better correspond to the energy really used:

$$AE = \frac{S_{O,sat}}{T \cdot 1.8 \cdot 1000} \cdot \int_0^T \sum_{aerated\ units} V_i \cdot K_L a_i(t)dt \qquad \text{(II.8)}$$

Pumping energy

Pumping energy (PE, in kWh.d^{-1}) is directly related to the integral of the sum of the internal recycle flow rate (Q_a), the return sludge flow rate (Q_r) and the waste sludge flow rate (Q_w). Flow rates are expressed in m^3.d^{-1}. The formula originally proposed in BSM1 is the following:

$$PE = \frac{0.04}{T} \cdot \int_0^T Q_a(t) + Q_r(t) + Q_w(t)dt \qquad \text{(II.9)}$$

Just as with aeration energy, it has also been recently updated for BSM2:

$$PE = \frac{1}{T} \cdot \int_0^T 0.004 \cdot Q_a(t) + 0.08 \cdot Q_r(t) + 0.05 \cdot Q_w(t)dt \qquad \text{(II.10)}$$

Sludge production to be disposed

The computation of the sludge production, *SP* (in kg.d^{-1}), proposed in BSM1, is based on the sum of the mass of solids accumulated in the system during the last week of the simulation and the mass of solids wasted from the secondary clarifier. The formula is presented in equation III.11, where *MTSS(t)* is the sum of the mass of solids in the reactors and the settler (in kg) and $TSS_w(t)$ is the concentration of solids in the wastage (in g.m^{-3}) and $Q_w(t)$ is the wastage flow rate (in m^3.d^{-1}) at time *t*:

$$SP = \frac{1}{T} \cdot \left(MTSS(T) - MTSS(0) + \frac{1}{1000} \int_0^T TSS_w(t) \cdot Q_w(t) dt \right) \quad \text{(II.11)}$$

The first two terms represent a potential increase or decrease of the mass of sludge in the system. They are corrected by the third term which is the mass of sludge withdrawn from the system.

Total sludge production

It may also be necessary to consider total sludge production (SP_{total}, in kg.d^{-1}) in order to take into account the amount of solids leaving the WWTP with the effluent (the concentration of solids in the effluent is low but the mass of sludge leaving the WWTP with the effluent is usually around 10 % of total sludge production). Total sludge production is therefore based on the sum of the sludge produced for disposal and the mass of sludge in the effluent, computed with the concentration of solids in the effluent $TSS_e(t)$ (in g.m^{-3}) and the effluent flow rate $Q_e(t)$ (in m^{-3}) at time *t*:

$$SP_{total} = SP + \frac{1}{1000 \cdot T} \cdot \int_0^T TSS_e(t) \cdot Q_e(t) dt \quad \text{(II.12)}$$

Mixing energy

The compartments in anaerobic and anoxic state must be mixed to avoid settling (which does not occur in aerobic state due to the mixing induced by the aeration). The formula proposed in BSM2 to compute this mixing energy (*ME*, in kWh.d^{-1}) is:

$$ME = \frac{24}{T} \cdot \int_0^T \sum_{i=1}^{5} \begin{Bmatrix} 0.005 \cdot V_i \ \ if \ \ K_L a_i(t) < 20 \ d^{-1} \\ 0 \ \ otherwise \end{Bmatrix} \cdot dt \qquad \text{(II.13)}$$

where V_i is the individual tank volumes (in m^3) and $K_La_i(t)$ is the instantaneous oxygen transfer coefficient (in d^{-1}).

Additional carbon source dosage

To improve denitrification, some controls of BSM2 can consider the addition of external carbon to the wastewater. This leads to an additional cost corresponding to this extra carbon source. The dosage of carbon added to the system (*EC*, in kg(COD).d^{-1}) is evaluated with the following formula, where COD_{EC} is the concentration of the external carbon source (in g(COD).m^{-3}) and $q_{EC,i}$ (in m^{-3}) is the flow rate of external carbon added to the activated sludge unit *i*:

$$EC = \frac{COD_{EC}}{T \cdot 1000} \cdot \int_0^T \left(\sum_{i=1}^{5} q_{EC,i} \right) \cdot dt \qquad \text{(II.14)}$$

Heating energy for anaerobic digester

The anaerobic digester needs to be heated for proper operation. The corresponding heating energy (HE, in kWh.d^{-1}) is estimated in BSM2 with the following formula:

$$HE = \frac{24}{T \cdot 86400} \cdot \int_O^T \rho_{H2O} \cdot c_{p,H2O} \cdot \left(Temp_{op} - Temp_{ad,in}(t) \right) \cdot Q_{ad,in}(t) \cdot dt \qquad \text{(II.15)}$$

where ρ_{H2O} is the water density (1000 kg.m^{-3}), $c_{p,H2O}$ is the water specific heat capacity (4.186 kJ.kg^{-1}), $Temp_{op}$ is the operational temperature needed in the mesophilic anaerobic digester (35 °C or 308.15 K in BSM2), $Temp_{ad,in}$ is the temperature of the anaerobic digester influent (in °C or K, the same unit as $Temp_{op}$ must be used) and $Q_{ad,in}$ is the flow rate of the anaerobic digester influent (in m^3.d^{-1}).

When multiple flows are entering the digester, the inlet temperature is estimated with their respective temperature $Temp_{ad,in,i}$ and flow rate $Q_{ad,in,i}$, based on a flow-weighted average:

$$Temp_{ad,in} = \frac{\sum_i Temp_{ad,in,i} \cdot Q_{ad,in,i}}{\sum_i Q_{ad,in,i}} \quad \text{(II.16)}$$

Net heating energy (HE^{net}, in kWh.d^{-1}) is also computed, based on the assumption that methane is used to produce electricity with a gas engine. Heat is therefore also produced and it is typically used to heat the digester. The assumption of the following formula is that 1 kg of methane produces 7 kWh of heat (in addition to the electricity produced by the gas engine). A correction is also made in the computation of net energy as it can never be negative:

$$HE^{net} = \max(0, HE - 7 \cdot MP) \quad \text{(II.17)}$$

Methane, hydrogen and carbon dioxide gases production

The methane production (MP in kg(CH_4).d^{-1}) of the anaerobic digester is evaluated in BSM2 based on the general gas law,

$$MP = \frac{p_{atm} \cdot M_{CH4}}{T \cdot R \cdot Temp_{op}} \cdot \int_O^T \frac{p_{gas,CH4}(t)}{p_{gas,tot}(t)} \cdot Q_{gas}(t) \cdot dt \quad \text{(II.18)}$$

where p_{atm} is the atmospheric pressure (1.013 bar), M_{CH4} is the atomic weight of CH_4 (16 g.mol^{-1}), R is the universal gas constant (8.3145.10^{-2} bar.m^{-3}.kmol^{-1}.K^{-1}), $Temp_{op}$ is the operational temperature of the digester (308.15 K), $p_{gas,CH4}$ is the partial pressure of the methane gas produced (in bar), $p_{gas,tot}$ is the total pressure of gas produced (in bar) and Q_{gas} (in m^3.d^{-1}) is the gas flow rate (normalized at p_{atm}).

The same computation is applied to evaluate the production of hydrogen and carbon dioxide, with the partial pressures and atomic weights of these gases.

Operating cost index

The operating cost index (OCI, without unit) is defined in BSM2 as a weighted sum of aeration energy, pumping energy, sludge production for disposal, dosage of external carbon addition, mixing energy, heating energy and methane production:

$$OCI = AE + PE + 3 \cdot SP + 3 \cdot EC + ME - 6 \cdot MP + HE^{net} \quad \text{(II.19)}$$

Effluent concentrations 95% percentiles

In order to report the worst performance on effluent concentrations, 95% percentiles of effluent concentrations are computed in BSM2 for ammonia ($S_{NH,e95}$, in g(N).m^{-3}), total nitrogen ($N_{tot,e95}$, in g(N).m^{-3}) and total suspended solids (TSS_{e95}, in g.m^{-3}).

Effluent limits violations

To quantify the violation of effluent limits, two criteria are proposed for each limit in BSM1: the percentage of total simulation time during which the limit is violated and the number of violations during the simulation.

In BSM1 and BSM2, the following limits are defined: $COD_e < 100$ g(COD).m^{-3}, $BOD_{5,e} < 10$ g(COD).m^{-3}, $TSS_e < 30$ g.m^{-3}, $N_{tot,e} < 18$ g(N).m^{-3} and $S_{NH,e} < 4$ g(N).m^{-3}.

Average concentration of pollutant in the effluent

The average concentration of a given pollutant P in the effluent on a period T is also proposed in the new BSM1 protocol that will soon be published. This criterion is computed based on the equation II.20. This criterion is added to allow the separate consideration of individual pollutants.

$$P = \frac{1}{T} \cdot \int_0^T P_e(t)dt \quad \text{(II.20)}$$

Controller performance

In order to assess each controller performance, the first criteria proposed in the BSMs are based on the controller error (e_j), which is the difference between the set point ($Z_{j,set\ point}$) and the measured value ($Z_{j,observed}$):

$$e_j = Z_{j,set\ point} - Z_{j,observed} \quad \text{(II.21)}$$

The criteria considered are the integral of the absolute error (Eq. II.22), integral of the squared error (Eq. II.23), maximum value of the absolute error (Eq. II.24) and variance of the error (Eq. II.25) .

$$IAE_j = \int_0^T \left| e_j(t) \right| dt \tag{II.22}$$

$$IAE_j = \int_0^T e_j(t)^2 \; dt \tag{II.23}$$

$$MaxDevErr_j = \max\left(\left| e_j \right| \right) \tag{II.24}$$

$$VarErr_j = \frac{\int_0^T e_j(t)^2 \; dt}{T} - \left(\frac{\int_0^T e_j(t) \; dt}{T} \right)^2 \tag{II.25}$$

Then, a second set of criteria is proposed, based on the controller impact on the manipulated variable (u_j). The maximum deviation (Eq. II.26) and the mean of this controller output are measured (Eq. II.27).

$$MaxDevU_j = \max(u_j) - \min(u_j) \tag{II.26}$$

$$MeanU_j = \frac{1}{T} \int_0^T u_j(t) \; dt \tag{II.27}$$

If uniform time steps (dt) are used in all simulations outputs as advised in the BSMs, the difference of the output between two consecutive time steps (Δu_j) can also be considered (Eq. II.28).

$$\Delta u_j = \left| u_j(t + dt) - u_j(t) \right| \tag{II.28}$$

Maximum variation (Eq. II.29) and variance (Eq. II.30) are computed.

$$MaxDev\Delta u_j = \max(\Delta u_j) \tag{II.29}$$

$$Var\Delta u_j = \frac{\int_0^T \Delta u_j(t)^2 \, dt}{T} - \left(\frac{\int_0^T \Delta u_j(t) \, dt}{T} \right)^2 \qquad \text{(II.30)}$$

Conclusion

This list is probably not exhaustive but all criteria considered in the BSMs are presented as well as a personal extension. All performance is not impacted in the same proportion by the control of the aeration which will be optimized, so only some of these criteria will be selected in the next chapters for the optimizations. Nevertheless, it was, in our opinion, important to introduce all of them in order to emphasize the multiple dimensions of the optimization problem.

2.4.4 Examples and comparison of typical operation of BSM1

In this thesis, SNDN and AMSTAR will be optimized and compared using BSM1. This section briefly introduces these two control laws with illustrations of how they operate.

The implementation chosen for these control laws on BSM1 is illustrated in Figure 2.17. Measurements for the control law are made in the last aerated tank and the same oxygen transfer coefficient computed by the control law is applied in the three aerobic tanks.

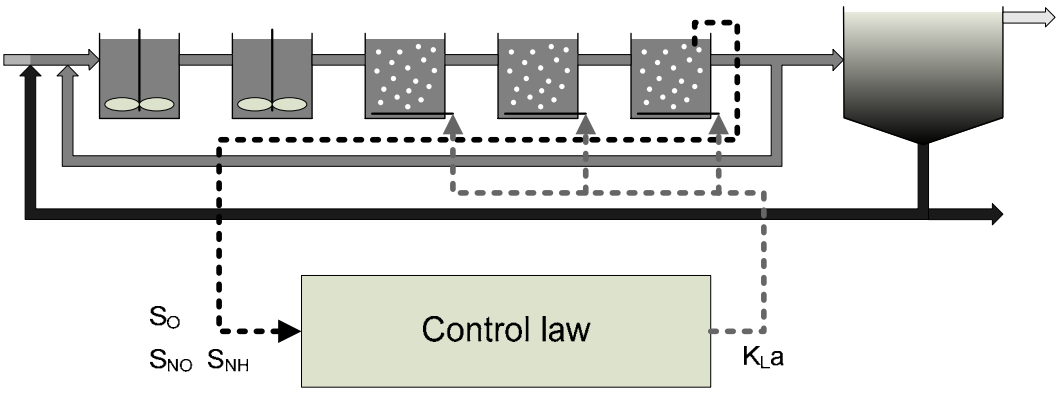

Figure 2.17: Implementation of the control law on BSM1 WWTP layout

For better operation of BSM1, two modifications are first made on aspects other than aeration control. First, the wastage flow rate is reduced to 220 $m^3.d^{-1}$ (instead of 385 $m^3.d^{-1}$). The

amount of sludge (and bacteria) in the activated sludge unit is hence increased as well as sludge residence time. In the BSM1 open loop (i.e. without any control law), the suspended solid concentration in the activated sludge is increased to 4.8 $kg.m^{-3}$ in average (instead of 3.3 $g.m^{-3}$) and the average sludge residence time is increased to 9.2 days (instead of 5.6 days). This modification allows better treatment performance with both control laws and is mandatory for a proper operation of SNDN. The flow rate is not reduced below this value of 220 $m^3.d^{-1}$ to avoid problems in the operation of the secondary settler (concentrations of suspended solids above 5 $kg.m^{-3}$ in the inlet of a conventional settler are not advised for proper operation).

The second modification is the dynamic control of the recycle flow rate to 100% of the influent flow rate. This modification allows better hydraulic velocities in the secondary settler and is broadly applied in WWTPs. This avoids the major problems of sludge rising during storm events. Such problems occur when most of the incoming flow of the settler goes to the overflow. Settling velocity of sludge is not sufficient in such cases to cancel the upward advection velocity and there is a direct bypass of the mixed liquor to the settler overflow and hence to the outlet of the WWTP.

The typical dynamic behaviors of BSM1 plant with AMSTAR and SNDN controls are illustrated in Figure 2.18. The S_{NH}, S_{NO} and S_O concentrations are depicted on the top graphs while the aeration coefficient is shown on the bottom graphs. The curves corresponding to AMSTAR are on the left, and to SNDN on the right. For these simulations, BSM1 procedure is used and the results of the second week of the BSM1 rain weather dataset are plotted on the figure. The controller's settings corresponding to this illustration are chosen to have the same treatment performance in terms of mean concentration of total nitrogen in the effluent for both control laws for sequenced or continuous aeration (12.1 $g(N).m^{-3}$).

However, if the effluent quality is almost equal, the mean concentration of ammonia in the effluent is larger with AMSTAR (6.1 $g(N).m^{-3}$) than with SNDN (5.2 $g(N).m^{-3}$). The energy consumption of AMSTAR is also higher (8647 $kWh.d^{-1}$) than the one of SNDN (8042 $kWh.d^{-1}$). In this example, the AMSTAR control scheme therefore consumes 7.5% more energy and releases 17% more ammonia than SNDN.

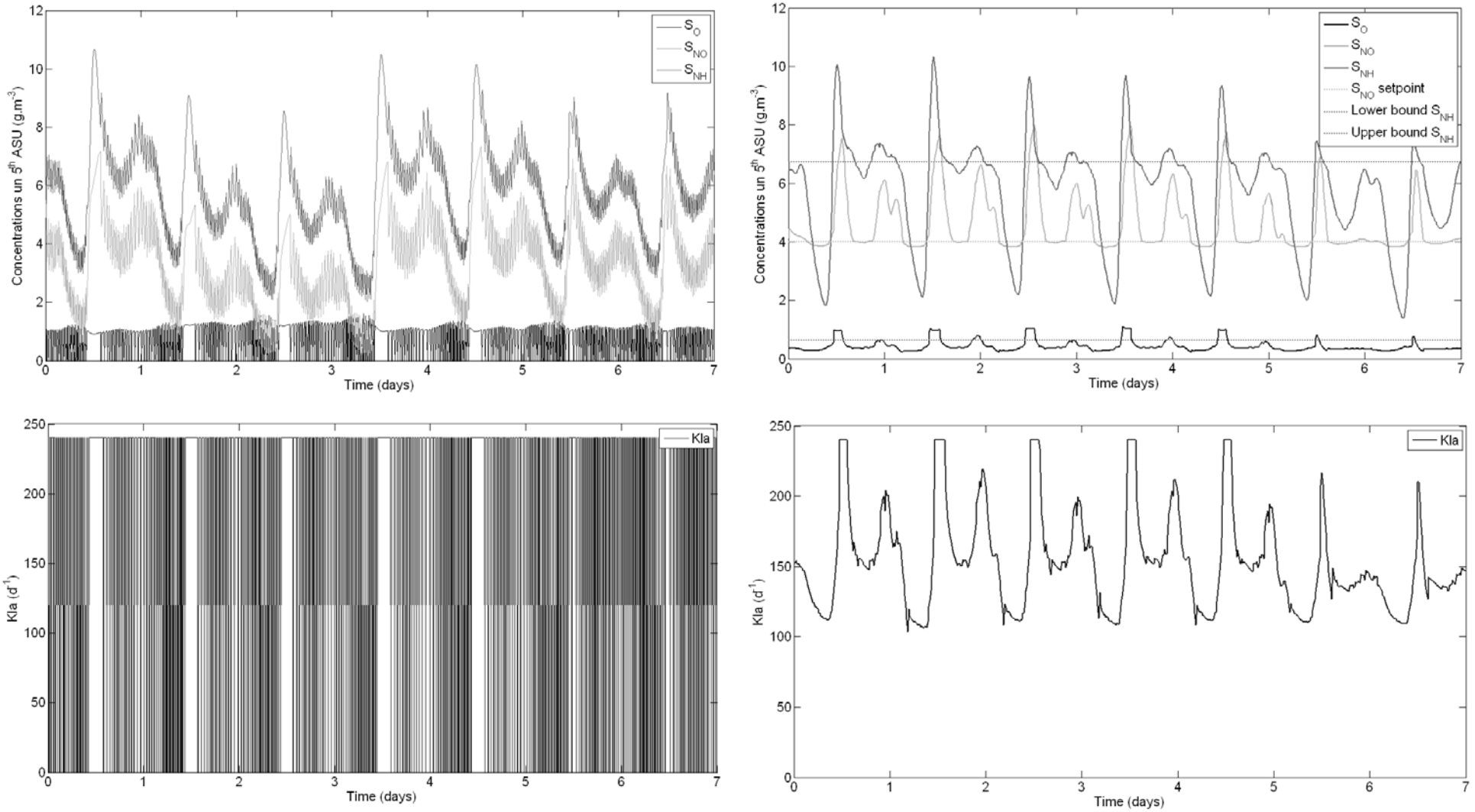

Figure 2.18: Typical curves of AMSTAR (left) and SNDN (right) control laws using BSM1

Looking at the AMSTAR figure, it may first be observed that the dissolved oxygen concentration does not reach the target of 2 g(-COD).m^{-3} during aeration phases. This is in fact due to the upper bound of the aeration coefficient which is limited to 240 d^{-1} in BSM1. A second remark is that high variations of concentrations of ammonia and nitrate are still present in the ASU. This is however also the case with SNDN and is in fact linked with the loading of BSM1 which is above EPA recommendations (Nopens *et al.*, 2008). The fast oscillations of the concentrations are induced by the sequencing of AMSTAR aeration. A third observation is the adaptation of AMSTAR phase lengths to incoming pollution. During the pollution peaks, the aeration phases are much longer and they slowly decrease at night as well as during the weekends. This is required to avoid excessive ammonia concentrations, even at the cost of an increase in nitrate concentrations as well.

In the right part of the figure for SNDN, the set point of nitrate and the bounds of ammonia are depicted on the upper graphs. A first remark is that the set point of nitrate is not followed during most of the time. This is in fact due to the upper bound of ammonia as well as the upper bound of the aeration coefficient which avoid proper operation of the nitrate controller during the whole day. The loading of BSM1 is also probably an important limitation in the application of SNDN. However, the main goals of this control scheme are reached: the nitrate concentration is kept at an acceptable value while a maximum of ammonia is removed. Oxygen concentrations are kept at low values allowing the simultaneous occurrence of both reactions. The oxygen concentration resulting in the same reaction kinetics for the oxidation of organics (with oxygen) and denitrification (without oxygen) in ASM1 is located at 0.4 g(-COD).m^{-3}. In the simulated case, the oxygen concentration evolves between 0.28 and 0.69 g(-COD).m^{-3} during 80 % of the time. Simultaneous nitrification and denitrification thus actually occur most of the time.

In order to further analyze the results of this comparison, the mass of pollutant reduced, the mass of oxygen transferred in the water and the mass of oxygen produced by the aeration system are computed based on mass-balances and depicted in Figure 2.19. For further details on the computation of these values, explanations are given in Appendix B.

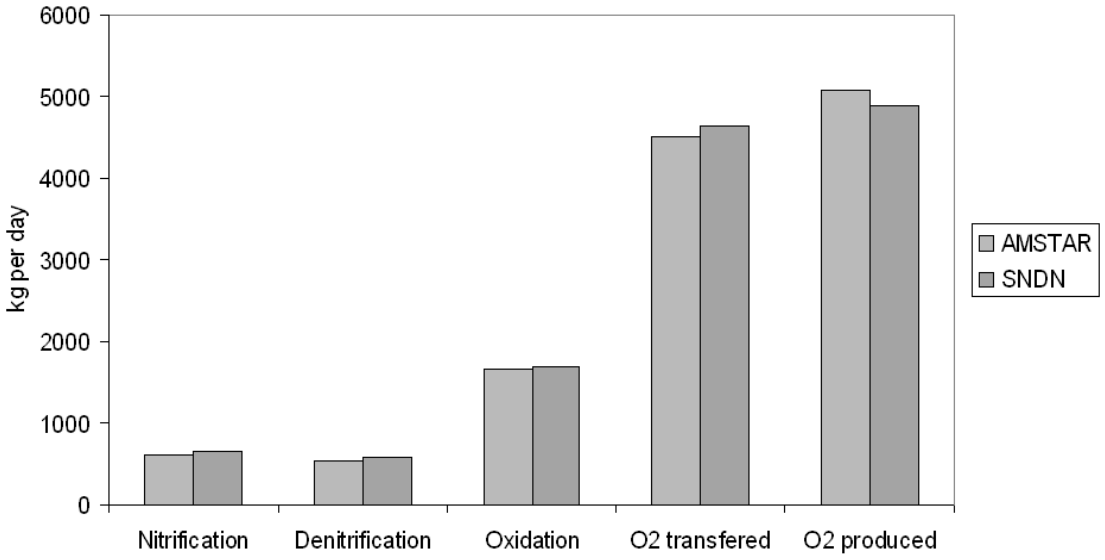

Figure 2.19: Total mass transferred with the two candidate settings of the AMSTAR and SNDN control laws (units are respectively kilograms of N per day for nitrification and denitrification, kilograms of COD per day for oxidation and negative kilograms of COD per day for oxygen).

We can observe on this figure that slightly more pollution has been removed with SNDN than with AMSTAR, even if the same effluent total nitrogen mean concentrations are observed in both cases. This is the consequence of the different units of these two criteria. One is expressed as load removed, hence taking into account the flow rate, while the other is expressed as a mean concentration. With SNDN, more oxygen has been transferred to its dissolved form in water. However, total oxygen produced by the aeration system is less with SNDN than with AMSTAR. These two observations are in fact directly linked with the model structure of ASMs where the mass of oxygen transfer is the multiplication of the tank volume, the oxygen aeration coefficient and the difference between the saturated oxygen concentration and the current one. Therefore, smaller oxygen concentrations in the tanks allow better transfers. This is the main reason why SNDN achieves lower energy consumption than AMSTAR with better pollution removal.

These presentations of the two control laws have highlighted their main characteristics. The numbers provided in this section however have to be moderated. This is only a single point of comparison, where the criterion considered to compare the two control laws is total nitrogen. It is then quite normal that AMSTAR does not perform as well as SNDN, since the latter is more targeted for an equilibrium between ammonia and nitrate (i.e. low total nitrogen values)

while AMSTAR is more intended to achieve very good elimination of the overall pollution and is hence less efficient in term of energy consumption.

This example is, in fact, a practical illustration of the risk of using single point optimization that may lead to the conclusion that continuous aeration is better than sequenced aeration. This is true for this specific operating point but there is no proof that it can be generalized to all kinds of settings of the control laws. This is the main reason why multiobjective optimization is necessary to provide clear insights into the real tradeoffs between respective performance of the two control laws and to solve the question of best control law choice.

2.5 Conclusion

The disturbances induced by high variations of the influent concentrations and flow rate have been introduced in this section. Typical processes used in WWTPs have been presented as well as the corresponding models used in the secondary treatment. Main control laws developed for appropriate aeration of the activated sludge unit and the key process needing precise control for proper pollution removal have been listed. Finally, the two BSMs have been detailed with the corresponding objectives that can be considered in process optimization. SNDN and AMSTAR control laws on BSM1 have been illustrated. These two control schemes will be considered for optimization in Chapter 4.

Chapter 3 - Multiobjective optimization with genetic algorithms

Chapitre 3 – L'optimisation multi-objectifs à l'aide d'algorithmes génétiques

Having examined WWTP models and control laws in the previous chapter, optimization techniques will now be detailed. As stated in Chapter 1, a multiobjective approach is required to handle the opposite objectives of our problems. Genetic Algorithms (GAs) are good for solving such problems and multiobjective adaptations are available in the literature.

In this chapter, the theoretical background on GA functioning is given in the first section. This is intended to clarify GA capability. However, many implementation of GAs are available, each with many subtleties. These details are not given in this chapter for the sake of clarity and conciseness. For more information on this topic, readers are redirected to the huge amount of books on GAs. A good start can be Reeves (2003).

The second section of this chapter focuses on multiobjective optimization.. The importance of this characteristic and existing GAs capable of solving such problems are explained. The algorithm chosen for this study (NSGA-II) is then detailed.

3.1 Genetic algorithms

Genetic algorithms (GAs) are a class of optimization algorithms. They are stochastic algorithms classified as metaheuristics (i.e. techniques capable of solving general optimization problems). Their origin is attributed to John Holland, who published a book exposing the scientific roots of this technique in 1975 (Holland, 1975). It was also that year that Ken De Jong, a graduate student of Holland, completed his dissertation thesis providing a clear insight into the optimization capabilities of GAs (De Jong, 1975). However, at that time, only very few real applications were developed, mainly due to the high computing power required. The real interest of the scientific and practitioner communities took off few years later, in 1989, when David Goldberg (Goldberg, 1989), another graduate student of Holland, published a very influential book that was a real catalyst for the application of the GA theory. Another reason is also probably linked with the increase of computing power that began to be available in many research centers at that time.

GA functioning is based on a mimic of the evolutionary theory developed by Charles Darwin (1859). The operations and concepts used in these algorithms (and even their name) are therefore mainly based on the vocabulary of genetics. For instance, a set of solutions is usually referred to as a population and the main operations in the algorithm are named

selection, crossover and mutation. These three main operations are the key structure of the algorithm: the selection consists in choosing the best solutions in the current population, the crossover is the generation of new candidate solutions based on the selected ones and the mutation alters the current solutions to avoid the risk of being trapped in local minima. The correspondence with Darwin's theory of the evolution of species is easy to follow, and this algorithm works based on Darwin's assumption that only the fittest individuals are capable of surviving in the long term. The main drawback of these algorithms is that the true optimum is very difficult to find and only near optimal solutions are generally found (but they are generally optimal).

3.1.1 Presentation of the algorithms for the search in binary spaces

Genetic algorithms were first developed for the optimization in a binary search space and this paragraph therefore focuses only on this case to present the main operations occurring during optimization. Extensions to continuous search space are presented in the next subsection.

Let us first assume a discrete search space B^n

$$B^n = \{0,1\}^n \tag{III.1}$$

and a function f

$$f : B^n \to \Re \tag{III.2}$$

The general problem of optimization consists in searching for

$$\min_{x \in B^n} f \tag{III.3}$$

In this expression, f is the objective function and x, named an individual, is a vector of decision variables composed of n binary values. This set of n binary values is named the chromosome of the individual. If the problem consists in maximizing an objective function, the modification to transform it into a minimization problem is obvious.

The flowchart of the operations of a genetic algorithm is presented in Figure 3.1. As in any optimization algorithms, an iterative procedure is used. For the initialization of the GA, a random initial population is generated. Then, the objective values are evaluated for this first

population. The three operations already described above are then performed on this population: selection of parents, crossover of these parents to create the new population and mutation of some of its bits. This new population is then evaluated and the procedure is repeated until some termination criterion is satisfied.

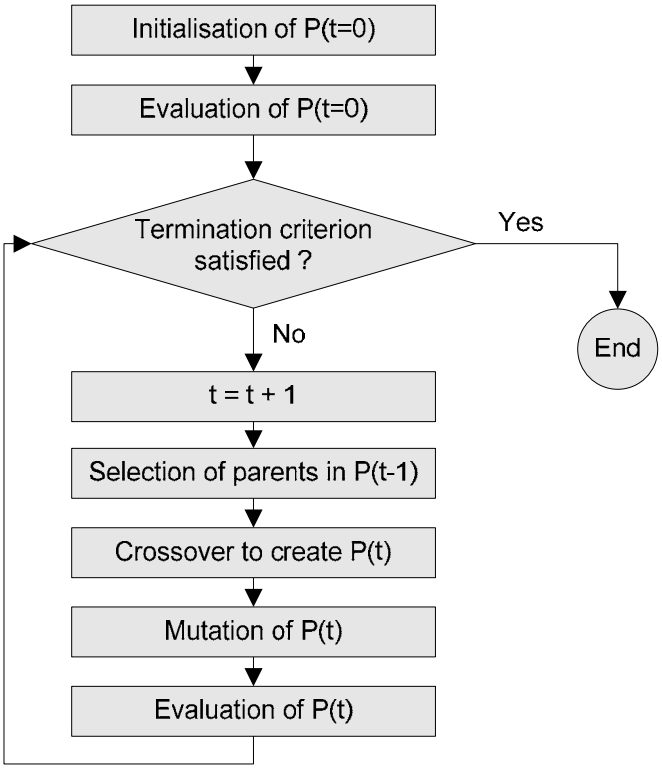

Figure 3.1: Flowchart of a genetic algorithm

Initialization

Before the initial population can be built, its size has to be chosen. This is one of the drawbacks in the use of GAs: no general rule is available for this choice and it is mainly based on the experience of the GA user. A general guideline for this choice is that the size of the population has to be related to the size of the search space (*n* in this binary case). Some theoretical propositions in this area are presented in Reeves (1993), leading to the conclusion that both values (0 and 1) of the chromosome should be found in at least one individual of the initial population and therefore a population size of $O(\log n)$ is sufficient to cover the search space.

The next problem is the choice of a method for the generation of this initial population. A traditional method consists in using random values. This method has the advantage of being very simple, but obviously it may happen quite often that the randomly chosen population does not cover the whole search space. Another solution therefore consists in using a model for the generation of this population. A simple and efficient method for this is the generalization of the Latin hypercube, which ensure good distribution of the population among the search space. This type of model is especially effective for searching in non-binary spaces.

A final important point to consider is the inclusion of a known high quality solution into the initial population. This may help in obtaining a quick convergence of the genetic algorithm, but there is a risk of premature convergence, where the algorithm will be stuck in the local optimum near this initial solution. This may happen by a loss of the diversity of the population, as the known solution is really good compared to other potential solutions randomly generated and the GA will therefore go very fast in this direction without necessarily having other good candidate solutions (Reeves 2003). It is therefore a risky choice, with benefits and drawbacks.

Selection

This step consists in choosing the individuals (which will be called parents) among the current population that will be used to generate the children for the next population. The selection of the parents should be related to the objective function values in order to promote better solutions. For this step, the roulette-wheel selection and the tournament selection are the two most popular mechanisms.

An example of these two mechanisms is illustrated in Figure 3.2. In this example, only four parents are considered for clarity's sake even if bigger population sizes are usually considered in real optimizations. The fitness associated with each individual (measure of the value of the objective value) is randomly chosen for this example. In real problems, fitness is usually measured as a distance from the current objective value to the worst objective value ever found (or to a known worst reference). The only condition on this measure is that it must always be positive and bigger values should be assigned to best solutions. Relative fitness is

the division of the fitness of an individual by the sum of all individuals fitnesses in the current population.

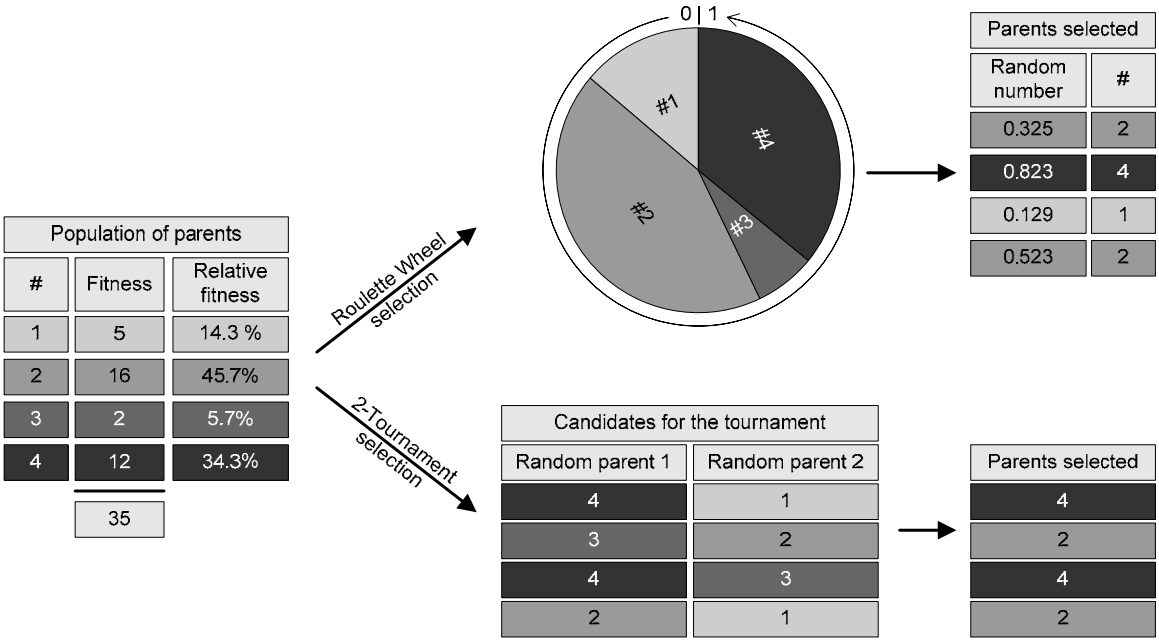

Figure 3.2: Roulette-wheel and tournament selections

The roulette wheel selection is the first mechanism developed. In this method, the space between 0 and 1 is divided in intervals whose size is related to the fitness of the current individuals (the size of the interval increases with the fitness). A random number is then generated between 0 and 1 and the selected parent is the individual whose interval contains the random number. This is repeated unless the required number of parents is reached. The main drawback of this selection method is the required use of a valued fitness.

Another mechanism is the tournament selection. In this method, *k* individuals (usually 2) are randomly chosen in the population and compared. The best one is selected as a parent. This selection is repeated unless the required number of parents is reached. The main advantage of this method compared to the roulette wheel selection is that it only needs an operation of comparison between potential solutions. Formal measure of fitness is not required. This selection can therefore work even when the value of the objective function is not formally defined but only comparison between solutions is possible. This is a technique that is typically used in multiobjective GAs like the one used in this thesis. From the example given below, one may note that only two parents are selected, and they are therefore selected twice

each. This is in fact not a problem as these parents are then used in pairs to generate children with random crossover, so children generated with these four parents will probably all be different.

Crossover

A crossover operation consists in generating new candidate solutions out of selected parents. Just as in natural evolution, the parents are randomly grouped in pairs that are used to produce two children for the new generation. A single random crossover point (1X crossover) is the traditional operation, but there is now a consensus that two-point crossover (2X) is preferable (Eshelman *et al.*, 1989). 1X crossover consists in randomly choosing a position in the chromosomes of the two parents and exchanging the values of the two chromosomes after this position. 2X crossover consists in randomly choosing two positions and exchanging the values of the two chromosomes in between these two positions. Figure 3.3 illustrates these two operations. Other more advanced mechanisms of crossover can be: reduced surrogates (Booker, 1987), uniform crossover or half-uniform crossover (HUX) (Eshelman, 1991).

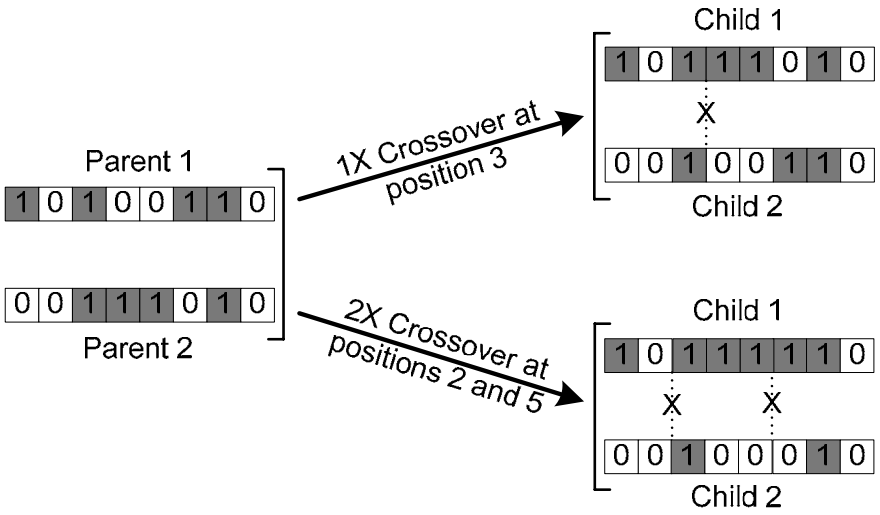

Figure 3.3: Operation of 1X crossover (top) and 2X crossover (bottom)

Mutation

Mutation consists in randomly altering one decision variable value. This operation is usually based on a random probability of occurrence. The main objective of this operation is to safeguard diversity in the population so as to avoid becoming stuck in a sub-optimal region and also to explore the whole search space. Mutation rate (probability of occurrence of this

operation) is a compromise between exploration of the search space and the number of function evaluations before reaching the optimum. Figure 3.4 gives an example of this operation.

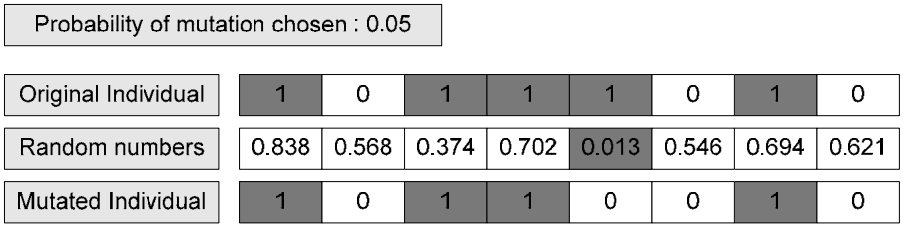

Figure 3.4: Operation of mutation

Termination

The choice of the criteria for the termination of the GA is probably the most difficult one. As GA is a overall and stochastic optimization procedure, it is quite difficult to know when the algorithm has reached its optimum. The solution usually consists in using a combination of a maximum number of generations without any change in the best solution found, a maximum total number of function evaluations and sometimes a criterion on the minimum diversity of population (when the diversity is too low, the algorithm can stop as we may assume that no more enhancement will be possible).

3.1.2 Genetic algorithms adaptations for the search in continuous spaces

The above introduction to GAs was based on the assumption that they were optimizing binary variables. However, in real life problems, optimization variables are usually in integer or real search spaces, and the direct transformation of these numbers in their binary coding is not the most adequate choice (crossover and mutation typically perform very poorly when such a simple transformation is used).

For integer variables, Gray coding can be used and has been proven to provide higher quality results than binary coding (Mathias and Whitley, 1994; Whitley, 1999). The key principle of this special binary coding of integers is that only one bit of the binary code has to be changed to move from one integer value to the next one (ire. plus one or minus one). An example is provided in Figure 3.5. The interest of this coding for GAs is that the mutation and crossover operations cause the integer values to change in their neighborhood, which is not the case with conventional binary coding.

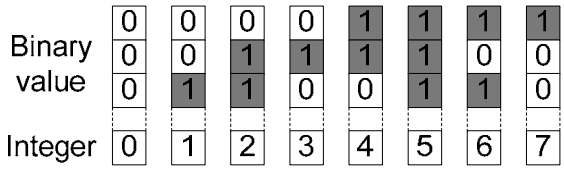

Figure 3.5: Gray coding with 3 bits for integers from 0 to 7

For real variables, the main problem is the size of the search space. When the precision required on the decision variable is high, the dimension of the problem when using a binary alphabet (either traditional binary code or Gray binary code) can be very high. It has been the main motivation for the developments of real-coded genetic algorithms. Instead of using bits as genes in the individuals, real values are used. Then, the crossover and mutation operations are redefined to provide good performance of the optimization procedure and similar functioning compared to the binary case. Many variants are available in the scientific literature (see for instance Herrera *et al.*, 1998; Deb, 2001).

3.2 Multiobjective optimization

The previous section presented the main characteristics of GAs. In this thesis, multiple objectives need to be optimized. Traditional GAs therefore need to be adapted to solve such problems, either with a transformation of the problem formulation or with specific GAs capable of handling all objectives at the same time. Both approaches are detailed in the next subsection. The approach based on adapted GAs is chosen for this study. Existing adaptations of GAs for multiobjective optimization are hence presented in the second subsection. Finally, NSGA-II, the multiobjective genetic algorithm chosen for this thesis, is explained in the third subsection.

3.2.1 Introduction of multiobjective optimization

The problem of using genetic algorithms as presented in the previous section is that many real-life optimization problems have more than one objective. Moreover, these objectives very often conflict and have to be optimized “at the same time”. Traditional optimization approaches can handle only one single objective. A way to aggregate all objectives therefore needs to be found. Usually, a weighting scheme is used, as in equation III.4, where obj_i are the individual objectives values, α_i are the weighting coefficients and obj is the aggregated objective value:

$$obj = \alpha_1 \cdot obj_1 + \alpha_2 \cdot obj_2 + \alpha_3 \cdot obj_3 + ... \quad \text{(III.4)}$$

This is convenient as it is the only way to use traditional algorithms. However, the choice of the weighting coefficients α_i is subjective and the optimal solution found is sensitive to these weights.

Another approach consists in adapting existing optimization algorithms to search for what is called the Pareto front, which is the set of best compromises considering all objectives.

The mathematical definition of this Pareto front is based on the concept of dominance, which allows comparison of two solutions x_1 and x_2, with $f_i(x)$ the value of objective number i for solution x:

A solution x_1 is said to *dominate* another solution x_2, denoted $x_1 \prec x_2$, if and only if

$\forall i \in \{1,2,\ldots,n\}, f_i(x_1) \leq f_i(x_2)$ and $\exists j \in \{1,2,\ldots,n\} / f_j(x_1) < f_j(x_2)$.

Reformulated, this definition says that solution x_1 dominates solution x_2 when all objective values of x_1 are smaller or equal to the objective values of x_2 and there is at least one objective value of x_1 which is strictly smaller than the corresponding objective value of x_2 (i.e. the first condition allows some objectives values of the two solutions to be equal while the second condition ensures that x_1 is better than x_2 for at least one objective). An example of four solutions in a minimization problem with two objectives is given in Figure 3.6. In this example, x_1 dominates x_3 and x_4, x_2 dominates x_4, and x_3 dominates x_4.

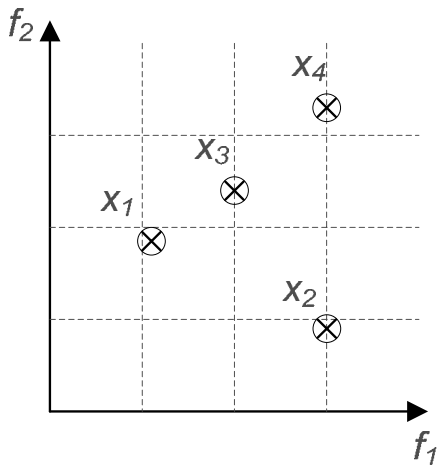

Figure 3.6: Example of four solutions in a minimization problem with two objectives

The Pareto front is then the aggregation of all Pareto points, based on the following mathematical definition:

If C is a set of potential solutions, a point $x^* \in C$ is said to be *Pareto optimal* or an *efficient* point or to belong to the *Pareto front* if and only if there does not exist $x \in C$ satisfying $x \prec x^*$. The vector x^* is then called *non-dominated* or *non-inferior*.

Reformulated, this definition says that a solution belongs to the Pareto front if it is not dominated by any other solution in the set of potential solutions. In the example of Figure 3.6, x_1 and x_2 forms the Pareto front. Another example with more solutions is illustrated in Figure 3.7. All squares correspond to potential solutions; only the dark ones are Pareto optimal.

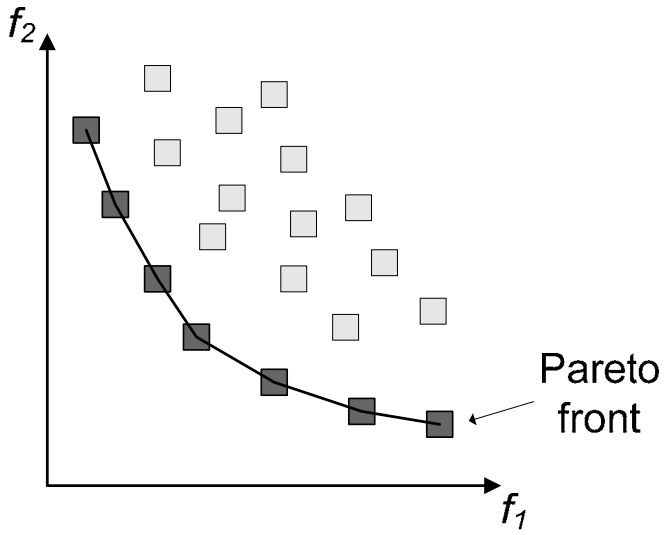

Figure 3.7: Example of a set of potential solutions (light and dark grey) and the associated Pareto front (dark grey)

The overall solution can then be chosen inside this set of best solutions, having a clear insight into the trade-offs involved. In an industrial context, multiobjective optimization is therefore very attractive. In industries, there are typically at least two people involved in optimization: a decision-maker and an analyst (concept first introduced by Savic, 2002). The decision-maker is the person responsible for making choices and selecting the best alternative. If he/she is aware of optimization methods, he/she will ask another person (the analyst) to carry out optimization of the problem. This will allow the decision-maker to make the right decision with more information where required. It is then the analyst who carries out the optimization, taking into account the formulation of the decision-maker's problem as well as its opposed

objectives. The analyst then provides the optimization results to the decision-maker, who now has more information to take the decision.

If a mono-objective optimization is performed by the analyst with a weighting scheme, the decision-maker will have only one solution proposed as a result by the analyst. This may be counter-intuitive for the decision-maker since only one trade-off is provided between the multiple opposed objectives of the problem. Usually, the analyst will have to explain that a weighting scheme was used and the decision-maker will usually ask about the sensitivity of the solution obtained to changes of the weights used. The analyst will not be able to answer this question except by making other optimizations with other weights. This may lead in the end to very long computing time and no real conception of the Pareto front and its visualization.

The analyst therefore has another solution: run a multiobjective optimization. In such a case, he/she may simply discuss the choice of the objectives with the decision-maker (whom can be more easily involved in this choice than in the choice of a weighting scheme). At the end of the optimization, the analyst will be able to provide the Pareto front of best alternatives to the decision-maker. The decision-maker will be informed of the various trade-offs in his/her problem, and will be able to base his/her decision on personal preferences among Pareto-optimal solutions. In this second scheme, almost no decision is made by the analyst. This is a more normal way of working from an industrial point of view.

3.2.2 Introduction of multiobjective genetic algorithms

Adaptations of the traditional genetic algorithms presented in the previous section are hence required to solve the problem of searching for the Pareto front of a given multiobjective problem. Many solutions are proposed in the literature and such GAs are named Multiobjective Genetic Algorithms (MOGAs). The three most common ones that are widely accepted and work on a wide range of problems are: Strength Pareto Evolutionary Algorithm 2 (Zitzler et al., 2002), Pareto Envelope-based Selection Algorithm II (Corne et al., 2001) and Non-dominated Sorting Genetic Algorithm II (Deb et al, 2000).

Other very simple algorithms are also available such as PAES (Knowles and Cornes 2000) or NPGA (Horn et al., 1994). These two algorithms are mainly used due to the criticism that the

three algorithms listed in the previous paragraph are time-consuming in their operation, as each solution has to be compared to all other ones to sort them. In fact, this criticism is of absolutely no concern in the application made in the present study . The computation time required for the internal logic of the GA is very small compared to the time required for the evaluation of a single solution, which implies a simulation run of the WWTP model. An order of 100 is usually observed between the two operations (internal logic of the GA and evaluation of the solutions). In our case, the number of function evaluations required to reach the optimum and the robustness of the MOGA are more significant than the time required for its internal computations.

For all common problems, the average performance level of the three algorithms of the first paragraph (SPEA2, PESA-II and NSGA-II) is quite similar (Zitzler et al., 2002; Corne et al., 2001) and the choice of using one instead of the others depends on the problem considered. No criteria have been found in the literature to help us in the choice of the algorithm to use, probably mainly due to the high variety of problems that can be solved with MOGAs. Moreover, it is difficult to transform most real world problems into a problem used for MOGA benchmarking. Finally, as the difference in performance among the three algorithms is very slight, the best performing one can change from one WWTP optimization problem to another one.

The choice of the algorithm to use in the present study is therefore mostly based only on *a priori*. Moreover, the goal of the present study is not to develop a particular optimization algorithm with very good performance but to clearly define and develop an optimization methodology. This methodology must be easily and widely applicable in the future since the present study is conducted in an industrial context. In our case, the choice is made to use the algorithm NSGA-II. This choice is based on the number of successful and real applications found using this algorithm (Reed and Minsker, 2004; Majumdar et al., 2005; Bekele and Nicklow, 2007). Other MOGAs would probably also have worked on our problem or even outperformed NSGA-II but this is not the topic of the present study .

3.2.3 The Non-Dominated Sorting Genetic Algorithm

The Non-Dominating Sorting Genetic Algorithm is one of the first adaptations of GAs for the search of the Pareto front proposed by Srinivas and Deb (1995). The second version (NSGA-

II) was proposed by Deb et al. (2000) to answer three main criticism on NSGA-I: the high computational complexity of $O(MN^3)$ for the GA internal computations (M is the number of objectives and N is the population size), the lack of elitism (the concept named 'elitism' consists in storing in memory the best Pareto front found so far together with the current population during the run of the genetic algorithm – traditional GAs just store the current population) and the use of a technique for diversity preservation in NSGA-I that is sensitive and based on a parameter that has to be user-defined. All these criticisms are solved with the NSGA-II.

Before describing the internals of NSGA-II, two new concepts used in this algorithm have to be introduced: the non-dominated sorting and the crowding distance computation.

The non-dominated sorting consists in assigning a 'rank' to each solution of the population, based on the concept of dominance previously introduced. The procedure to compute this rank is the following (see Figure 3.8). First, the solutions belonging to the Pareto front of the current population are assigned a rank 1 (F1). They are then temporary removed from the population and the solutions belonging to the new Pareto front are assigned are rank 2 (F2). They are then temporary removed. The procedure is repeated until all solutions have a rank. The solutions with the lowest rank are, of course, the best ones and will be favored during selection operations.

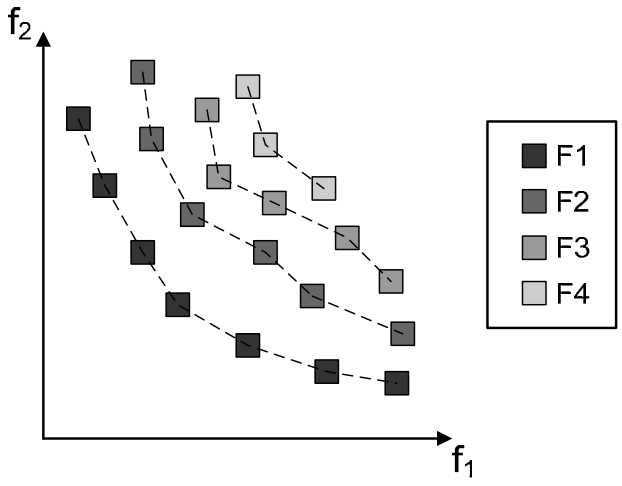

Figure 3.8: Example of the non-dominated sorting of a set of solutions

It is then necessary to differentiate among solutions with the same rank. This is done with the second concept of crowding distance. The main goal of a MOGA is to preserve the diversity

of the solutions in order to have effective coverage of the Pareto front. For a given solution, the crowding distance is the measure of the sum of the distances between the two nearest solutions along each objective. Solutions with big crowding distances are important to preserve while solutions with small crowding distances can be discarded as they have other solutions close to them. Therefore, solutions at the boundaries of the Pareto front (worst and best solutions for each objectives) are assigned an infinite crowding distance as they are very interesting ones, considering preservation of diversity.

Execution of NSGA-II is then based on these two concepts, as illustrated in Figure 3.9.

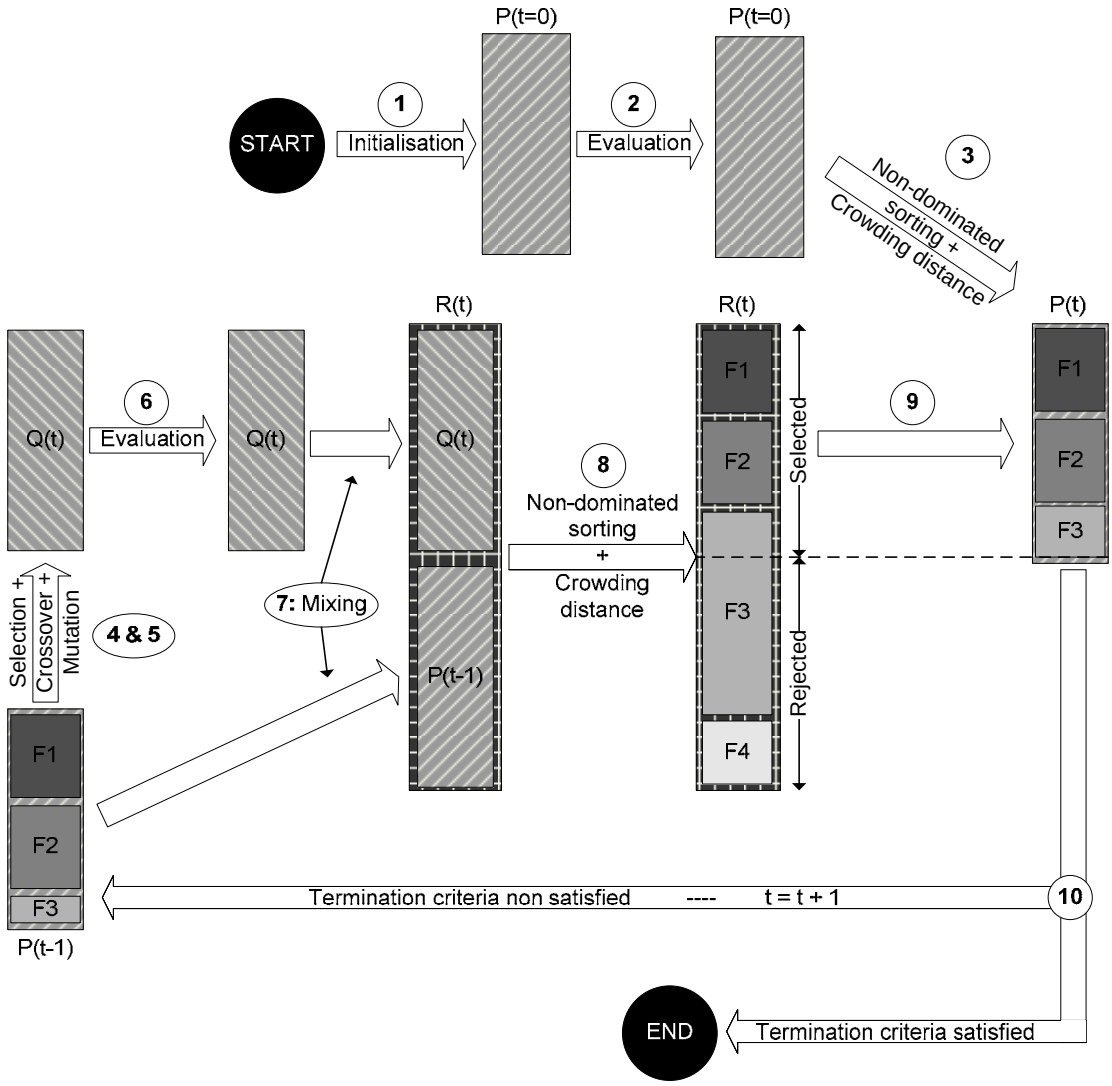

Figure 3.9: Flowchart of NSGA-II operations

After the classic initialization of the first population (1) and its evaluation (2), the solutions are sorted according to their non-domination rank and crowding distance (3). This set forms the first population P(t=0). An iterative procedure is then performed:

- tournament selection on the current population P(t) based on the previous sorting of the population (4),
- classic crossover and mutation to produce children Q(t) (5),
- evaluation of objective values for each solution in Q(t) (6),
- mixing of current population P(t) and children Q(t) in combined population R(t) (7),
- non-domination sorting of R(t) and computation of crowding distance (8),
- selection of the new population P(t+1) (9) :
    - first fronts whose summed number of solutions is lower than the population size are automatically selected,
    - the best solutions of the next front are selected according to their crowding distance,
    - other solutions of this next front and of following fronts are rejected;
- evaluation of the termination criteria, if not satisfied, a new generation is performed (10).

Using constraints during the optimization is obviously a good choice in order to keep only relevant solutions in terms of objective values but also in terms of parameter values or simulation functioning. A proposition for the handling of constraints is included in the original publication of NSGA-II (Deb et al., 2000). This mechanism handles constraints with modification of the definition of domination between two individuals, with the introduction of the constrained-domination concept:

A solution i is said to constrained-dominate a solution j, if any of the following conditions is true:
(i) Solution i is feasible and solution j is not.
(ii) Solution i and j are both infeasible but solution i has a smaller overall constraint violation
(iii) Solution i and j are both feasible but solution i dominates solution j

A solution is said to be feasible when it does not violate any constraint. The overall constraint violation is defined as the sum of all violated constraints (each measured as the distance to the feasible value of each constraint).

This new definition as the new concept of dominance in the non-dominated sorting of NSGA-II and constraints are then handled.

This has been a brief presentation of NSGA-II characteristics and how they function. To conclude, NSGA-II is very competitive compared to others like SPEA-2 or PESA-II and its main quality is the very good diversity of the solutions proposed (Deb et al., 2000). Another important point is the robustness of its default parameters which are assumed to provide good performance for most problems. The only parameter that has to be adjusted on all problems is the population size. Finally, a scheme for constraints handling is proposed in the original publication. This algorithm is therefore perfectly suited for dynamic optimization of combined models of WWTPs and their control laws.

3.2.4 Performance evaluation for multiobjective genetic algorithms

To compare various genetic algorithms or various parameters of the same algorithm for a given problem, performance measurements must be defined. These measurements will be used in this thesis to choose the best internal parameters of NSGA-II for a given problem (see subsection 4.7.1). This performance must be assessed at each generation of the GA run to compare the speed of convergence, for instance. When a single objective is used, a single metric is required and this is usually defined as the current best objective value found. When using multiple objectives, two types of performance are expected from the genetic algorithm: quick convergence to the true Pareto front (as in the single objective case) and good coverage of the Pareto front (or diversity). Mathematical definitions of these two concepts are proposed in Deb and Jain (2002).

For both metrics, a reference set considered as the true Pareto front of the problem studied is required. In a real-world problem however, this front is unknown. The advised solution consists in building this reference set with all non-dominated solutions found in the investigation of the various GA parameters and the various repetitions.

Convergence metric

The first metric is a convergence metric, used to evaluate the convergence of the running non-dominated solutions $F(t)$ of the current population $P(t)$ towards the reference set P^* considered as the true Pareto front.

At each generation for the population *P(t)*, the convergence metric is computed in the following manner:

Step 1: Identification of the non-dominated set *F(t)* of *P(t)*.

Step 2: For each point *i* in *F(t)*, calculation of the smallest normalized Euclidean distance to *P** as follows:

$$d_i = \min_{j=1}^{|P^*|} \sqrt{\sum_{k=1}^{M} \left(\frac{f_k(i) - f_k(j)}{f_k^{\max} - f_k^{\min}} \right)^2} \qquad \text{(III.5)}$$

where, f_k^{max} and f_k^{min} are the maximum and minimum values of the *k*-th objective in *P**.

Step 3: Calculation of the convergence metric by averaging the normalized distance for all points in *F(t)*:

$$C(P^{(t)}) = \frac{\sum_{i=1}^{|F^{(t)}|} d_i}{|F^{(t)}|} \qquad \text{(III.6)}$$

The convergence metric itself can then be normalized by dividing it with the worse convergence metric in the current optimization or in all optimizations used in the comparison.

To sum up, this convergence metric is an average distance of the current front to the true Pareto front. This metric has to be minimized.

Diversity metric

A second metric is used to evaluate the diversity of the current population. Diversity is a measurement of the distribution of the solutions on the Pareto front. The metric defined here is a small deviation (see Step 1 and footnote) from the original one proposed in Deb and Jain (2002).

To compute this metric, solutions first have to be projected on a suitable hyper-plane of the solution space. Instead of assessing the equal distribution of the solutions on the Pareto front, the solution will have to be equally distributed on this hyper-plane. For this, the hyper-plane is divided into a grid, the limits of which include the projection of all solutions among all generations. The number of divisions can be different in each dimension.

The diversity metric is then an average of the distribution of the solution among these hyper-boxes, related to the best possible distribution, considered as the one of the reference Pareto front.

Thus, the parameters required from the user are:

- the hyper-plane to use for projections (which can, for instance, be represented by its cosine direction),

- the number of divisions for each dimension of the hyper-plane.

The procedure for the computation of the diversity metric *D(P(t))* at the generation *t* is then the following:

Step 1: Determination of the set *F(t)* of non-dominated solutions from *P(t)*.[1]

Step 2: For each hyper-box indexed by (i,j,…), calculation of these two arrays:

$$H(i,\ j,...) = \begin{cases} 1,\ \textit{if the hyperbox has a representative point in } P^* \\ 0,\ \textit{otherwise} \end{cases} \quad \text{(III.7)}$$

$$h(i,\ j,...) = \begin{cases} 1,\ \textit{if } H(i,j,...) = 1 \textit{ and the hyperbox has a representative point in } F^{(t)} \\ 0,\ \textit{otherwise} \end{cases} \quad \text{(III.8)}$$

Step 3: Computation of the arrays *m(h(i,j,…))* and *m(H(i,j,…))* according to the neighbors in each dimensions of the hyper-plane. No formal definition of this neighborhood function *m* can be made but a proposed definition is given in Table 3.1 (the same definition of *m* is used for *h* and *H*). The goal of this function is to indicate the occupation of each hyper-box.

[1] Step 1 is a deviation from the Deb and Jain's proposition of 2002, which proposed set *F(t)* as the solution of *P(t)* that are non-dominated by P^*. This original definition is not really adequate, since during first generations, all solutions of *P(t)* are dominated by P^* so the diversity metric cannot be computed. Moreover, the goal of this metric is to evaluate the diversity of the best current solutions, so considering all non-dominated solutions of the current front seems more adequate.

Table 3.1: Definition of the neighborhood function m()

h(... j-1 ...)	*h(... j ...)*	*h(... j+1 ...)*	*m(h(... j ...)*
0	0	0	0
0	0	1	0.5
1	0	0	0.5
0	1	1	0.67
1	1	0	0.67
0	1	0	0.75
1	0	1	0.75
1	1	1	1

As this function requires the neighboring values of *h* and *H*, it is not defined on the boundaries of the hyper-plane. For this, imaginary neighbors are used on the boundaries and their *h* and *H* values are assumed to be 1. An example is shown in Figure 3.10. In this example, two objectives are optimized (f_1 and f_2). The current front *P(t)* is represented by squares (open and shaded), the true Pareto front *P* is represented by circles. Shaded squares form the current set of non-dominated solutions *F(t)*. The hyperspace chosen for the projection is the axis of the first objective and ten divisions are used. The table under the figure represents the computation of *m* for *H* and *h*, as well as for $\underline{0}$, a vector with zero values for each position that will be used in the next step.

Step 4: In the same manner as above, computation of $m(\underline{0})$ where $\underline{0}$ is an array filled with zeros.

Step 5: Calculation of the diversity by averaging *m(h(…))* with respect to *m(H(…))*, with a correction factor due to the boundary effects (*m(0)*):

$$D(P^{(t)}) = \frac{\sum_{H(i,j,\ldots)\neq 0}^{i,j,\ldots} m(h(i,j,\ldots)) - \sum_{H(i,j,\ldots)\neq 0}^{i,j,\ldots} m(\underline{0})}{\sum_{H(i,j,\ldots)\neq 0}^{i,j,\ldots} m(H(i,j,\ldots)) - \sum_{H(i,j,\ldots)\neq 0}^{i,j,\ldots} m(\underline{0})} \tag{III.9}$$

In the example of Figure 3.10, *H(i)* is always equal to one so the diversity metric is equal to:

$$D(P_{example}) = \frac{6.85-1}{10-1} = 0.65 \tag{III.10}$$

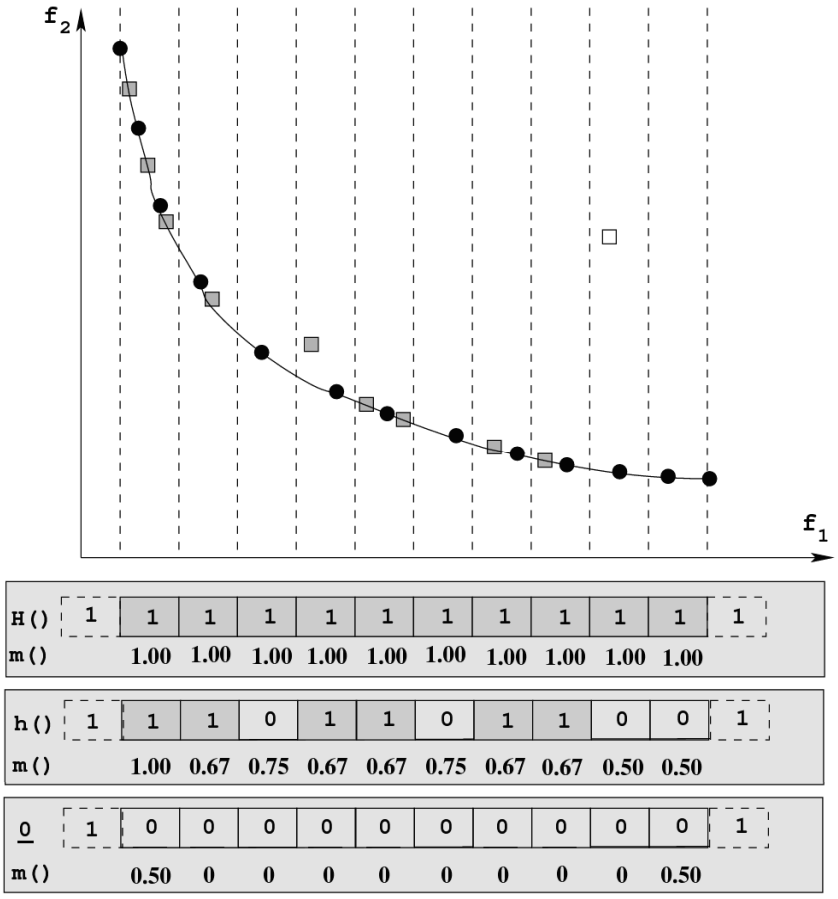

Figure 3.10: Example of diversity computation (adapted from Deb et al., 2002) (the squares represent the current solutions and the circles the true Pareto front P*)

In summary, this diversity metric is an average of the current Pareto front distribution normalized with the true Pareto front distribution. This measurement must be maximized.

Conclusion

With these two metrics of convergence and diversity, it is now possible to clearly compare the runs of the genetic algorithm during repetitions of an optimization. It should be noted that the performance comparison should be based on the current number of WWTP model evaluations performed. This is the right criterion since different runs of a GAs must be compared in terms of computation time. In our application, this computation time is related to the number of evaluations and not to the number of generations performed.

3.3 Conclusion

A general presentation of genetic algorithms has been made in this section. The benefit of using a multiobjective approach has been clearly identified for our problem. NSGA-II, a genetic algorithm capable of solving such multiobjective problems has been selected for this study. Finally, a scheme for constraints-handling and metrics for the evaluation of its performance have been proposed. In the following section, the methodology developed during the present study will be explained, based on the combination of NSGA-II with BSM1 simulations. The final application of the methodology in this literature case study will also be detailed.

Chapter 4 - Optimization methodology development on a literature case study

Chapitre 4 – Développement de la méthodologie d'optimisation sur un cas d'étude de la littérature

Chapter 2 and 3 presented the theoretical roots required for the development of a new methodology for multiobjective optimization of WWTP control laws. This chapter illustrates this new methodology based on BSM1 case study.

As our goal is to check the long-term impact of the control law settings in a plant-wide context, BSM2 would have been the perfect case study. Unfortunately, a single evaluation of the whole BSM2 model over a two year simulation typically takes one hour on a modern computer (an Intel Xeon 3.0 GHz was mostly used in this study). Since many evaluations of the whole model are done by the genetic algorithm (between 1000 and 10000), the use of the whole BSM2 was unfortunately not possible. This would have led to an overall computation time between 1.5 and 15 months, which is not feasible during a PhD program, especially for the development of a new methodology.

The development of the methodology is therefore based on the simpler BSM1 test case. In this configuration, the control law for simultaneous nitrification and denitrification (SNDN) and the aeration module of STAR (AMSTAR) presented in Chapter 2 will be optimized. Implementation of BSM1 is done with Matlab/Simulink C S-functions. Validation of the simulation results in accordance with BSM1 procedure is presented in Appendix A.

In the present chapter, a general outline of the methodology is detailed in the first section. Details are then given for each aspect of this approach. The results of this application to a literature case study are presented in the last section of this chapter while a full industrial context is discussed in the following chapter from the results obtained in the real case study of Cambrai WWTP.

4.1 Presentation of the methodology

The optimization methodology proposed is based on a combination of a multi-objective genetic algorithm (NSGA-II) and dynamic simulations of WWTP models (including control laws). This combination is named dynamic optimization and its procedure in our methodology is illustrated in Figure 4.1. The optimization procedure presented in this figure is as follows.

First, the genetic algorithm proposes new solutions to test, based on best solutions found so far (or initial solutions for the first generation). These solutions contain different values of the decision variables that are manipulated by the genetic algorithm. In our application, these decision variables are set points, bounds or internal parameters of WWTP control laws.

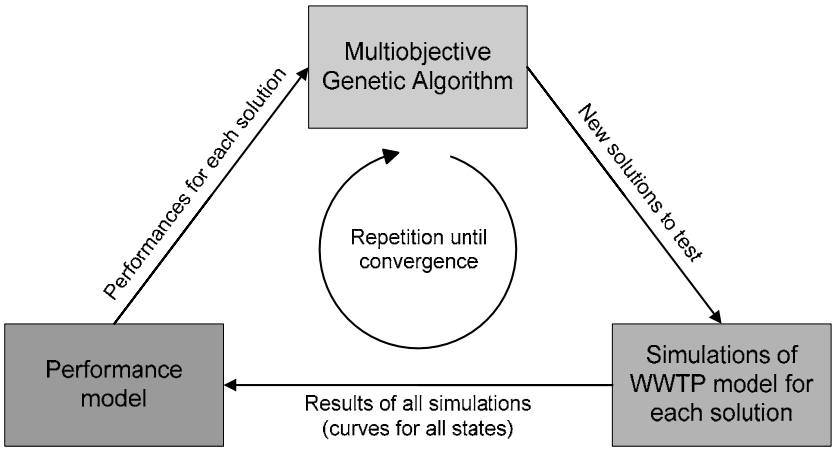

Figure 4.1: Flowchart for the dynamic optimization of WWTP

Then, an evaluation of the WWTP operation is made for each solution through dynamic simulations of the WWTP model. This model is modified according to decision variable values of each solution. These simulations are based on a new simulation procedure proposed in this thesis and presented in section 4.2. This procedure ensures that the simulations have converged to a stabilized operating point before evaluating its performance. This stabilization is required to ensure a reliable evaluation of the obtained performance. This stabilization is based on one-week repetitions of dynamic dry weather. A refinement reducing the average simulation time required to reach this stabilized operation is also presented in next section. After this stabilization, a last week of simulation is conducted for evaluation of the solution performance. Experiments showed that the use of the rain weather influent dataset of BSM1 ensures more robust evaluation of BSM1 performance. This last point will be detailed in section 4.3.

Following these simulations, objectives and constraint values can be computed with the performance model. This model is simply the aggregation of the objective and constraint computations. The importance of the objective choice is stressed in section 4.5. Constraints are also significant to ensure that performance automatically evaluated by the GA corresponds to an expected operation in practice. Further details are given in section 4.6.

The whole procedure is then repeated for each generation of the genetic algorithm until the optimization has converged. An evaluation of the robustness of best found solutions has then to be performed. This is necessary since only BSM1 datasets representing dry weather and rain weather were used during the optimization. The robustness of the long-term performance with many different disturbances therefore needs to be assessed. This is done with the simulation of each final optimization solution with BSM1_LT long-term influent dataset (Rosen *et al.*, 2004). This dataset represents typical disturbances occurring on a real WWTP and allows long-term (609-day) simulations. Further details are given in section 4.4, as well as the procedure developed to analyze these 609 days of simulation.

This methodology will now be detailed in the following sections.

4.2 Enhancement of the simulation procedure

In order to ensure reliable evaluations of each solution proposed by the GA, simulations used for the performance evaluations must represent a stabilized operation of the WWTP. This is required to avoid finding minima that only correspond to low performance accuracy.

The procedure proposed in BSM1 consists in using first a steady state simulation of 100 days, followed by a two-week dynamic simulation with dry weather and then, a two-week dynamic simulation with the weather dataset to be tested. Only the last week of simulation is used for performance evaluation.

This procedure is highly computing-intensive. Stiff solvers can sometimes be used for faster evaluation of the steady-state simulation. This is possible for continuous control laws but it does not work with hybrid controls involving sharp fronts such as sequenced aeration. On average, if T_s is the time required for the simulation of one week of dynamic data, a continuous system will, on average, require $5\,T_s$ for the whole BSM1 procedure while a hybrid system will require $18\,T_s$ (on average, both systems require $4\,T_s$ for the 4 weeks of dynamic simulations but the continuous system requires only $1\,T_s$ for the steady state simulation instead of $14\,T_s$ for the hybrid system).

Another weakness of BSM1 procedure is the lack of guarantee that the simulations have "converged" to a stabilized operation before the last week used for performance evaluation. This will, of course, be the case in most simulations but this assessment is based only on

observations of the simulator operation in typical cases. As the optimization algorithm is automatically run during these evaluations, it will not be possible to check the convergence of BSM1 procedure and this may induce incorrect performance evaluation. A solution hence had to be developed.

This problem of "convergence" of the simulations is mainly induced by the slow evolution of microbial populations within the system. In BSM1 procedure, the first three weeks of dynamical simulations are, in fact, one-week repetitions representing dry weather. The system hence evolves in a cyclic mode the length of which is one week (in addition to the quasi-cyclic mode induced by the daily repetitions, although the weekend may somewhat disturb this smaller repetition). Techniques to find the pseudo steady-states (i.e. the stabilization of the states under cyclic conditions) are available (e.g. see Platte et al., 2005). These techniques compute the pseudo steady-state based on the cyclic property of the system. It is, however, not clear if these tools can provide relevant results in limited computing time and this solution was therefore not selected. Moreover, these techniques use specific solvers which would have required the translation of the WWTP model into a new software platform. This would not have been in accordance with one industrial objective of the present study which is the use of traditional simulation software already used by engineers to model WWTPs. It was therefore decided to develop a new solution based on the properties of our application.

First, this solution is based on one-week repetitions representing dry weather. In order to check if these simulations have converged, we had to look at the differences of all state values between the beginning and the end of the week of simulation. If these differences are null, it means that the next week of simulation will be same since its initial values will be identical to those of the current week of simulations. The system has hence converged. In fact, due to the tolerances in the simulator solver computations, it is significant not to check that the difference is null but just inside a given tolerance. Finally, after convergence, a specific last week of simulation is carried out for performance evaluation. In our case, the rain weather influent dataset of BSM1 is here used in order to give reliable performance evaluations and hence realistic optimization results (details are given in next section). These one-week simulation repetitions and the convergence criteria are the solution proposed to ensure that simulations have converged but this implies slower evaluations.

A refinement of this solution is hence proposed to make the evaluations run faster. It is based on the fact that many repetitions of the same model with only few parameter changes are performed in the optimization procedure. As a consequence, the initialization of a simulation can be based on simulations previously performed. The best way to choose which one to use is then based on the values of the parameters that are changed by the genetic algorithm since solutions with few differences in parameter values are assumed to represent almost the same operation. The solution therefore consists in storing the final states found at the end of the "convergence" simulations in a "knowledge base" together with the values of the parameters issued from the genetic algorithm for this specific evaluation. For the following evaluations, the model states of the first simulation week are initialized with the final ones of the previous evaluation having the lower difference in parameter values. To compute this difference, equation IV.1 is used:

$$d_{1,2} = \sqrt{\sum_{i=1}^{N} \left(\frac{x_{i,1} - x_{i,2}}{x_{i,\max} - x_{i,\min}} \right)^2} \qquad \text{(IV.1)}$$

where $d_{1,2}$ is the distance between parameters of solutions 1 and 2, $x_{i,1}$ and $x_{i,2}$ are the values of parameter i for solutions 1 and 2, $x_{i,max}$ and $x_{i,min}$ are the maximum and minimum values of this parameter in the optimization (i.e. the bounds of the space to be explored by the GA, defined before running the optimization) and N is the number of parameters of the optimization. With this equation, the distances between each parameter of the two simulations compared are first normalized taking into account the minimum and maximum bounds of this parameter in the optimization. A Euclidian norm of these normalized distances is then computed as the total distance between the two solutions parameters.

This simulation procedure proposed thus ensures the model convergence to a stabilized state before the evaluation and this convergence is faster on average than BSM1 procedure because of its use of previous simulations as initial guesses of the process states (even if these initial guesses may be far from real pseudo-steady states). Finally, a specific influent dataset is used for a last week of simulation for the performance evaluation.

This procedure is summarized in Figure 4.2 and consists in:

1. Initializing the states with the final ones obtained from the nearest simulation in the parameter space (equation IV.1). If no previous simulation was performed, the initialization is done with the steady state values obtained in open loop,

2. Repeating one week of simulation with the dry weather input dataset (DWID) until the difference between the values of the states at the beginning of the week and the values at the end of the week are within a given tolerance. For each iteration, the initial state values are set to the final values of the previous iteration,

3. Storing the final state values together with the parameter values in the knowledge base for initialization of future simulations,

4. Simulating the final week with the rain weather input dataset (RWID). The results of this last week of simulation are then sent to the performance model.

The repetition of the convergence simulations is based on DWID since a normal operation of the process should be used to ensure that a normal stabilization is reached.

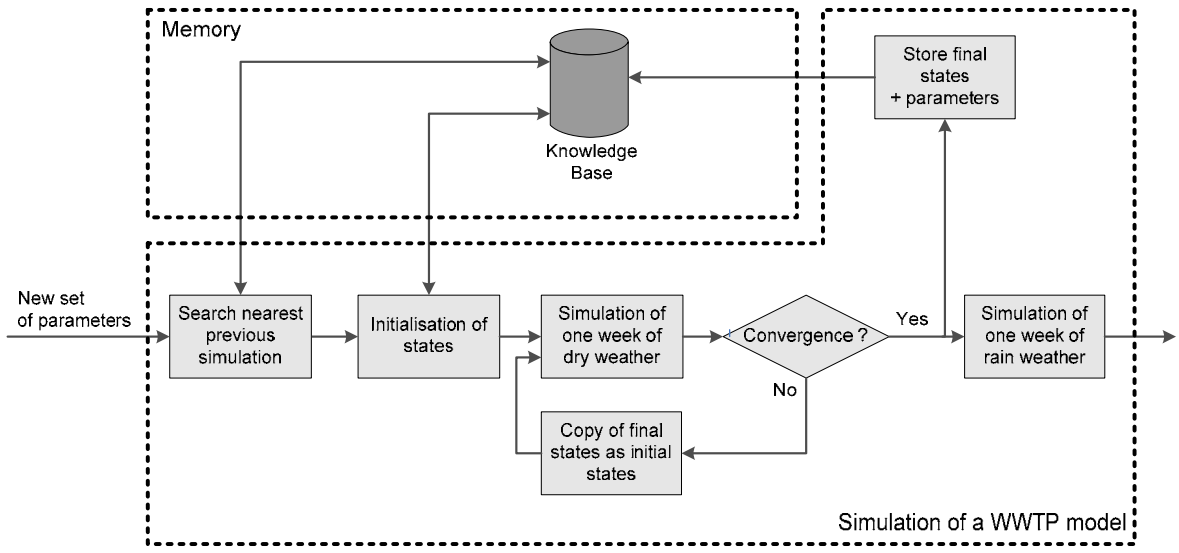

Figure 4.2: Simulation procedure for consistent and quick evaluation of parameter sets

Figure 4.3 shows the total simulation time curve (repetitions and final evaluation, normalized with the simulation time of one week of dynamic dataset, T_s) during an optimization for SNDN case. The time that would have been used with BSM1 procedure is also depicted as a

reference (which is constant, and takes 5 T_s). It can be seen on the figure that during the first evaluations, many weeks are repeated (between 3 and 9). Then, after around 200 evaluations, only one or two weeks are necessary to achieve convergence. This proves that the proposed procedure achieves truly better performance in terms of simulation time using SNDN case than BSM1 procedure, thanks to the repetitive aspect of GAs evaluations, and the insurance of convergence to stabilized operation.

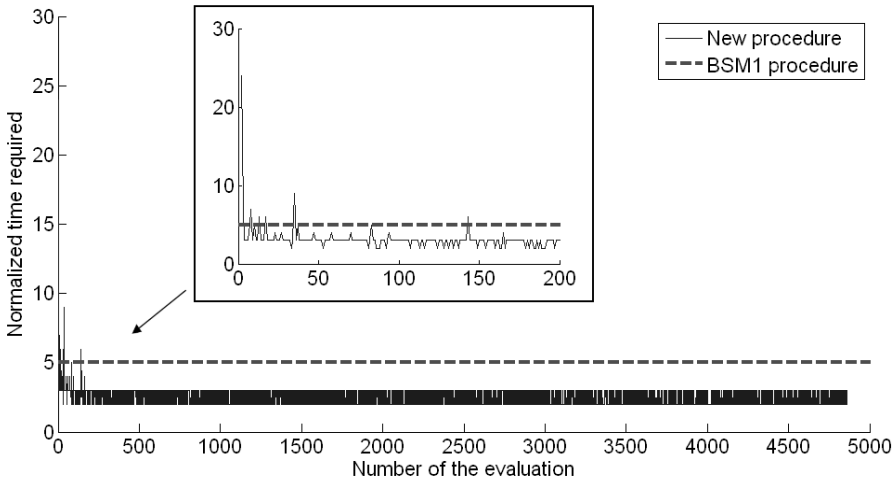

Figure 4.3: Normalized time spend for each evaluation of SNDN optimization on BSM1

The same figure for AMSTAR optimization is not depicted since the normalized time corresponding to these evaluations is much more variable from one evaluation to the next and does not decrease as the number of evaluations performed increases.

In order to still compare the average results with BSM1 procedure, cumulative frequencies are computed and illustrated in Figure 4.4 for AMSTAR (left) and SNDN (right). Concerning SNDN, the same observation is made as before. Most of the evaluations (99.3 %) require 3 T_s or less. With regard to AMSTAR, there is much more variability in each evaluation simulation time. However, 80.3% of the evaluations take less time to perform than they would have required with BSM1 procedure.

The main problem illustrated here is that 15.1 % of AMSTAR evaluations took 31 T_s. This constant value for this portion of the graph is linked with an upper limit of 30 repetitions that is used to avoid too many repetitions. This limit is required due to the sequenced nature of AMSTAR which might imply oscillations from one simulation week to the next.

For instance, a simulation initial state may correspond to start of aeration, whereas its final states may correspond to end of aeration (there is no fixed number of cycles per week; it is controlled just by ammonia and nitrate concentrations). In such an example, the stabilization will not be achieved, since there is a big difference between the initial and final ammonia, nitrate and oxygen concentrations. This is currently the main drawback of the method, which may still not provide insurance of convergence in all cases. However, 30 weeks of simulation is already a large number that is certainly sufficient to achieve this convergence.

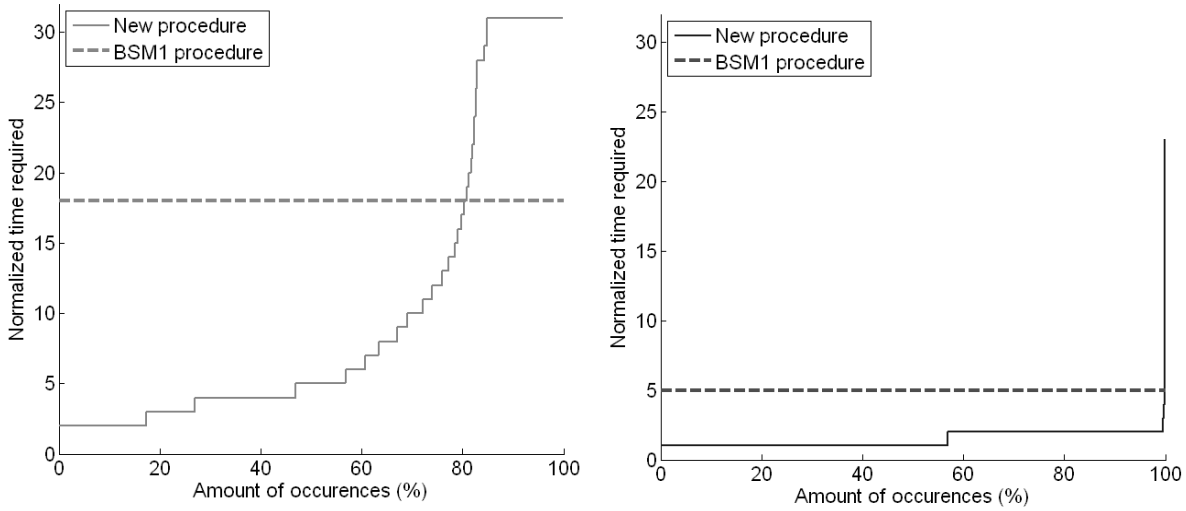

Figure 4.4: Cumulative frequencies of simulation normalized time for the optimization of AMSTAR (left) and SNDN (right)

The two figures above show that the procedure proposed for the evaluation of a GA solution also provides benefits in terms of computation time for the two examples considered, even if there is no formal proof of this assumption. However, this finding was not the main objective of this simulation procedure. Its main goal is indeed to guarantee simulation convergence before a performance evaluation. This reduction in average evaluation time is a greater benefit of the procedure developed.

4.3 Choice of the evaluation dataset

In order to provide a reliable evaluation of solution performance, the input dataset used for the final simulation week of the procedure presented in the previous section should be carefully selected. It is to represent a wide range of disturbances typically induced by the influent on the WWTP performance. In BSM1, three influent datasets are proposed, representing,

respectively, dry, rain and storm weather. The storm weather input dataset is not considered for the performance evaluation since its impact on aeration control is too limited. Then, either the dry weather input dataset (DWID) or the rain weather input dataset (RWID) can be of interest.

Figure 4.5 illustrates the optimization results obtained when using DWID (left) or RWID (right). The SNDN was optimized on BSM1 in this case and the objectives of the optimization were effluent quality and energy consumption (sum of aeration energy and pumping energy, as defined in BSM1; see subsection 2.4.3 for further details). Each point on the figure corresponds to the performance of a final solution obtained at the end of optimization. Short-term performance corresponds to the evaluation of these best solutions obtained during the optimization. Long-term performance corresponds to mean values of the daily performance obtained at the end of the optimization for each final solution with an additional simulation based on the 609-day influent dataset of BSM1_LT (only the median is considered here for simplicity but more information on the analysis of these long-term results is given in subsection 4.4 where 5^{th} and 95^{th} percentiles are also considered). The lines indicate the correspondence for each solution between its short-term and long-term performance.

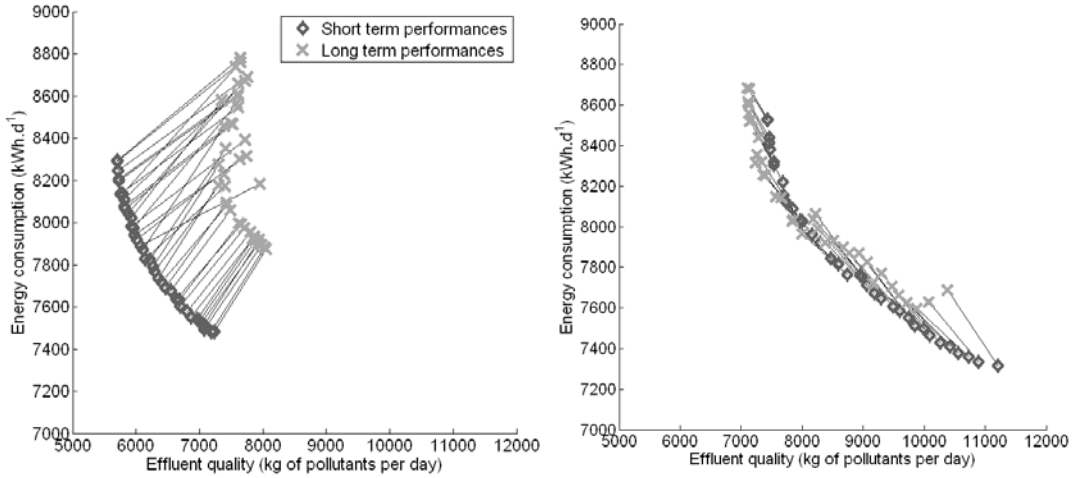

Figure 4.5. Comparison of short-term and long-term performance obtained with DWID (left) and RWID (left).

On the left part of the figure, a strong deviation is observed between short-term and long-term performance. This is quite normal because the daily WWTP loading and flow rate have stronger variations in the BSM1_LT input dataset than in BSM1 DWID. On average, it is usual that the WWTP has worse performance when the loading and/or flow rate are increased.

This first problem supports the use of these long-term evaluations which will be discussed later in the section 4.4.

Another problem that can be observed on the left part of this figure is that some solutions that are optimal (i.e. non-dominated) on the short-term are not optimal anymore in the long-term. In fact, only half of the solutions proposed after the optimization is optimal in the long-term. Moreover, the solutions that are not optimal in the long-term are among the ones with the better effluent quality (lower than 6000 kg(pollutant units).d^{-1}). This is quite damaging since solutions in this range of effluent quality may have been missed because their objective values were slightly worse than those of selected solutions. Moreover, these missed solutions may have been more robust in the long-term than the selected ones.

On the right part of the figure, the deviation between short and long-term performance is much smaller and almost constant for all solutions. This results from using RWID for the performance evaluation instead of DWID. This low deviation is; however, only the consequence of the objectives chosen and WWTP model considered. This low deviation between RWID and long-term performance must not be generalized to all optimization problems. In this figure, a few solutions are not optimal anymore in the long-term but they do not have a significant impact since they can simply be dismissed from the set of optimal solutions, since they are regularly spaced on the Pareto front. Moreover, their deviations from the long-term Pareto front are limited. Only the two solutions with the lower energy have larger deviations and should be removed from the final set of solutions. They may however have been interesting since their removal will limit the extension of the Pareto front.

Overall, this figure indicates that the use of RWID for the performance evaluation provides more robust performance. This is certainly related to the fact that more disturbances are included in RWID than in DWID .

In order to further analyze this result, the second set of solutions resulting from the optimization with RWID can be simulated again and evaluated with DWID. Dry weather performance can then be compared with the performance of the first solution set. This should indicate whether or not some promising solutions are overlooked when RWID is used during optimization. This comparison is made in Figure 4.6. Original BSM1 performance (in open and closed loop) is also indicated for reference. This figure indicates that performance is very

tight considering solutions with very good effluent quality. When effluent quality increases, the difference in performance of the two set of solutions increases too, the solution obtained with DWID being more optimal. Finally, there is a range of low energy consumption which is reached by the first set of solutions but possible with the second one. This clearly indicates that there are not so many differences between the two solutions considering this criterion of short-term performance during dry weather.

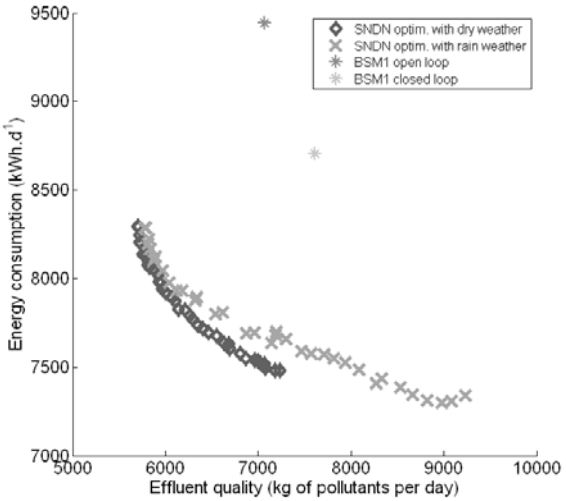

Figure 4.6. Comparison of short-term performance during dry weather for the solutions obtained with a performance evaluation based on DWID (dark grey) and RWID (light grey) compared to original BSM1 performance (stars).

A final comparison of the two evaluation datasets can then be based on their long-term performance, individually depicted in Figure 4.5 based on mean performance. In Figure 4.7, daily performance of these two alternatives during the 609 days of simulation are considered. For each solution, the 5^{th}, 50^{th} and 95^{th} percentiles are depicted (the medians are represented with straight lines while the extremes are in dotted lines). More information on this comparison is detailed in section 4.4, below. What is clearly visible in this figure is that the two medians are very similar, except that the median of the optimization based on evaluations with RWID is capable of reaching lower energy that the other one. As stated before, around a half of the optimization solutions based on DWID are not optimal in the long-term, unlike most optimization solutions with RWID. This is especially visible with the 95^{th} percentiles, even if other percentiles also indicate this non-optimality. The only limitation of the use of RWID is that some solutions obtained with DWID are reliable and perform better than solutions found with RWID.

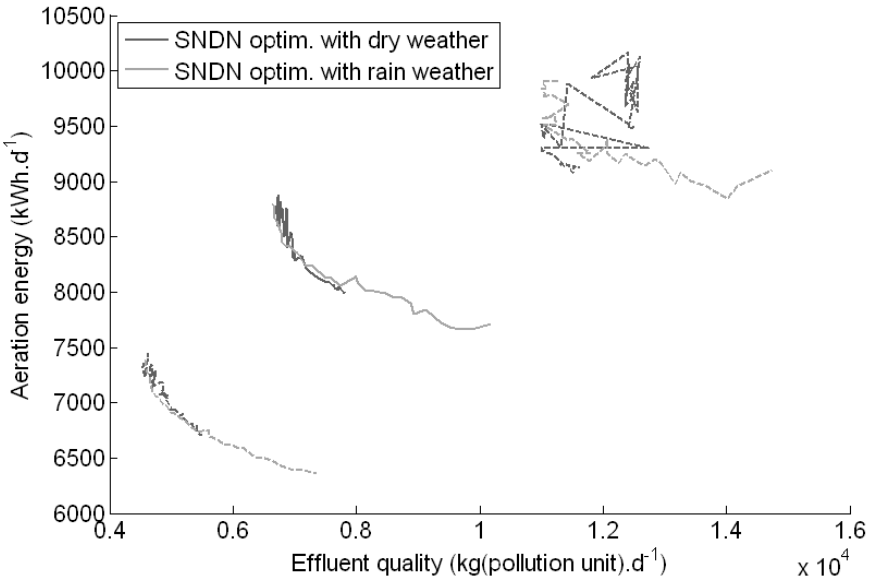

Figure 4.7. Comparison of daily long-term performance of optimized SNDN based on DWID (dark grey) and RWID (light grey). The 5th, 50th and 95th percentiles are shown.

To conclude, this section has focused on the choice of the input dataset for performance evaluation. The choice of RWID previously presented was justified in this section, based on the observation of the short-term and long-term performance obtained with both alternatives. With these evaluations, it is clear that the results obtained with RWID tend to be better than those obtained with DWID. However DWID still should be used in the convergence simulation procedure as it is more representative of a "normal" WWTP operation.

4.4 Evaluation of the robustness of the optimization in the long-term

As the performance evaluation is based only on one week of data, the number of events represented is small. The performance of the optimized control laws have therefore been assessed only for a limited amount of disturbances. However, in practice, a large number of disturbances occur every year and the WWTP must be capable of handling these events optimally.

Unfortunately, using simulations with a long-time simulation horizon is not realistic due to the very long computing time required. One way to evaluate the robustness of solutions could have been to use a synthetic influent signal with a large spectrum of frequencies but its impact on the non-linear processes is not clearly known. Moreover, the attempt of generating this synthetic signal resulted in a long dataset due to the very large influent spectrum of frequencies. The selected solution hence consists in an *a posteriori* check of solution

robustness, as proof of the optimization results application. This is reliable in terms of computing time since only a limited amount of simulations corresponding to the final population should be made. If the results show that some solutions are not reliable in the long-term, they may be removed from the set of optimal solutions. Another possibility is to analyze these specific solutions in order to make conclusions about necessary modifications of the control law or some additional constraints that have to be included in the optimization problem or even some modifications to the optimization methodology that can be made (as with the choice of RWID instead of DWID presented in the previous section).

This evaluation of solution robustness is very significant to avoid providing incorrect results at the end of the optimization. Indeed, if a solution with low robustness is implemented on a real WWTP, it may lead to very poor performance. In the end, this may lead to the conclusion that the optimization methodology is not working adequately since the solution chosen for practical implementation was not robust. The previous section showed a perfect example, where many solutions were not robust in the long-term and therefore were not to be considered at the end of the optimization.

In order to evaluate this long-term robustness, each final optimization solution is simulated again with a long influent dataset containing many different disturbances. This dataset represents 609 days and illustrates typical variations of the influent of a WWTP. It is part of BSM1_LT, an extension of BSM1 for the assessment of its long-term performance (Rosen et al., 2004 ; Gernaey et al., 2006b). The difficulty when using this dataset is related to the interpretation of these 609 days of simulation, especially when the goal is to infer conclusions about control law robustness. The goal of this step is, in fact, to evaluate the possible deviation between short and long-term performance as well as the variability of the long-term performance.

To reach this goal, the solution proposed here is based on the following procedure. First, for each solution, the daily performance of the WWTP (based on the chosen objectives) is computed, hence providing 609 sets of performance for each solution. Then, for each solution the 5^{th}, 50^{th} and 95^{th} percentiles of daily performance is computed for each objective. This allows us to have only three points for each solution. These three points can then be compared to those of another solution, with information on the median performance and their variability.

The choice to only consider the median (50^{th} percentile) and not the mean allows us to have more insight about the overall expected performance, while the mean might have been disturbed by extreme values of very poor days. The choice of the 5^{th} and 95^{th} percentiles is then based on the fact that it is necessary to assess which are the worst performance of the WWTP. Minima and maxima are not considered as they are too extreme. The 10^{th} and 90^{th} percentiles lack quite a lot of information as the performance of the WWTP needs to be fulfilled most of the time and are therefore not considered here either.

In order to give more information, each set of percentiles (5^{th}, 50^{th} and 95^{th}) is plotted with a line linking the individual solutions in a precise order. This sorting of the solutions is based on short-term performance, with only a monotonically increasing or decreasing objective considered. This sorting and linking of long-term solution performance allow us to see if there is just a constant deviation of all solutions during the long-term evaluation or if there is more uncertainty in final performance.

Figure 4.7 in the previous section is a practical example of this visualization of long-term performance. In this example, we see that the solutions obtained with DWID are quite jittered and hence not as robust as the one obtained with RWID.

The combination of this long-term influent dataset of 609 days and the new visualization technique proposed in this paragraph hence provide insight into the robustness of the solutions obtained with the optimization.

4.5 On the importance of the choice of the objectives

Now that the way to perform the evaluations of the model is clear, the objectives of the optimization have to be defined. As stated in section 2.4.3, many objectives are available for the evaluation of a WWTP performance. The appropriate choice will ultimately be up to the decision-maker. For instance, one may consider for WWTP optimization that the most significant objectives are total effluent quality and total energy consumption. This is true, generally speaking, in order to limit the environmental impact of WWTPs. However, in practice, the decision-maker will also base his decision on the local regulations or on the risk

of the various alternatives. Choosing the right objectives is very significant to provide clear insight into these alternatives that will in the end allow a clear decision[1].

As an example, the control law for SNDN is optimized twice with different objectives on BSM1 case study and the results are presented in Figure 4.8 and Figure 4.9. In the first optimization, only effluent quality and the energy consumption are considered. In the second optimization, effluent quality is replaced by the mean concentration of ammonia and the mean concentration of total nitrogen in the effluent. In Figure 4.9, the three subplots correspond to the projection of the Pareto front on the three sub-axes (obtained with permutations of the three objectives values). In both figures, each point corresponds to a final solution at the end of the optimization.

As can be seen in Figure 4.9, the use of the second set of objectives allows us to discover that there is a point of minimum total nitrogen concentration (indicated by an arrow) which is not visible in Figure 4.8 when only the first set of two objectives is considered. If a decision-maker is more interested in the second set of three objectives, this indication of a minimum total nitrogen mean concentration in the effluent might help him/her by indicating that if ammonia concentration is further reduced, this will induce a higher total mean nitrogen concentration in the effluent.

This is just a single example showing the importance of the objective choice, highlighting that they should be in accordance with the decision maker's point of view. Of course, each optimization problem will benefit from an appropriate choice of objectives (other examples are provided in the application of section 4.7 as well as in chapter 6). A general recommendation is however to also consider the use of constraints when they are more adequate than explicit objectives.

[1] The respective roles of the decision-maker (who makes the final choices) and the analyst (who performs the optimization) are detailed in subsection 3.2.1, as well as the importance of this difference for the application of the methodology in the industry

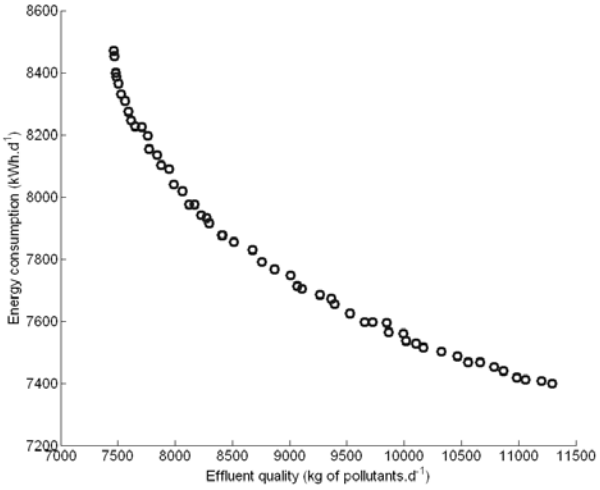

Figure 4.8: Optimization results with two objectives considered (effluent quality and energy consumption, sum of aeration energy and pumping energy).

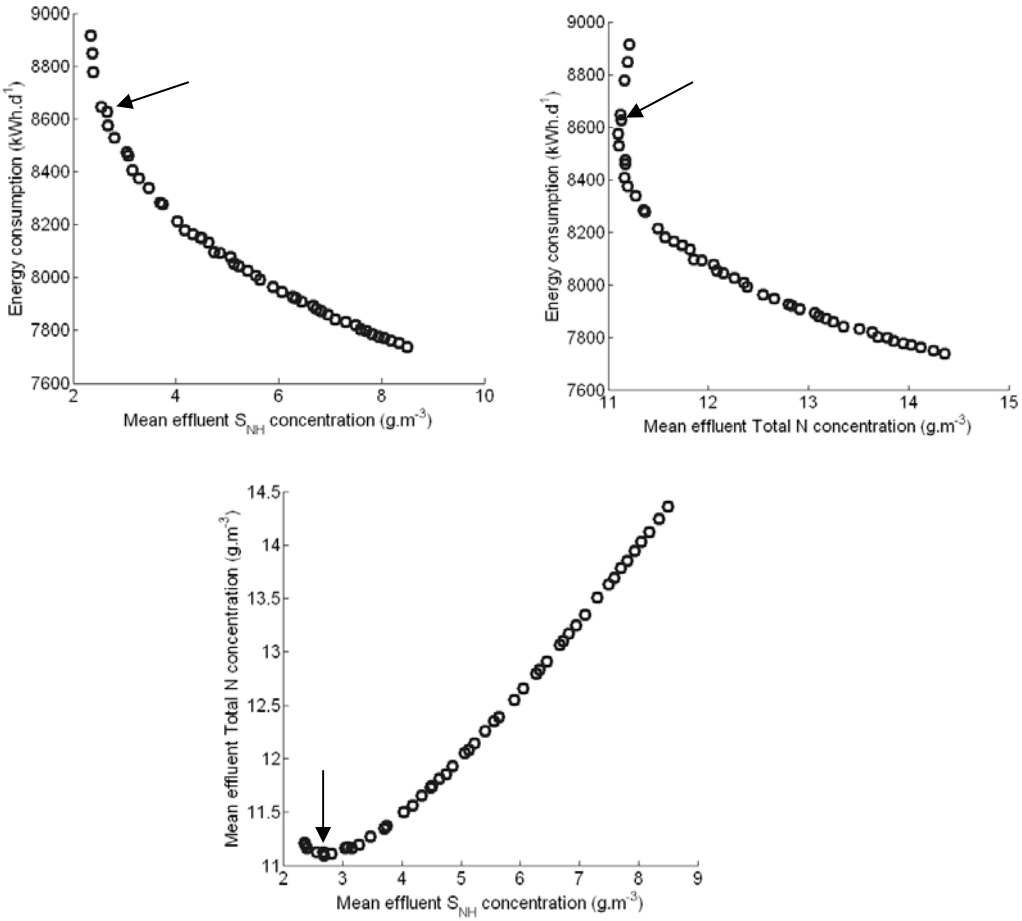

Figure 4.9: Optimization results with three objectives considered (mean effluent concentrations of total N and ammonia, and energy consumption)

4.6 On the importance of using constraints

4.6.1 Definition of the constraints to consider

Controller performance is typically better handled when expressed as constraints. This is due to the multiobjective optimization procedure. In such optimizations, the goal is indeed to find trade-offs among all objectives of the problem. In our case, it is nonsense to consider a trade-off between controller performance and effluent quality for instance. With regard to controller performance, the goal is indeed to ensure that the controller is behaving normally during optimization of effluent quality or other objectives. Therefore, a constraint on minimum and/or maximum admissible controller performance is sufficient and more adequate. Such constraints must not be omitted since they ensure that the solutions considered by the GA truly represent an operating point where the controller is acting on the system. Without these constraints, the performance obtained might only be “random” for the specific influent dataset used but performance with another dataset is likely to be worse.

Another type of constraint that can be considered concerns the performance of the WWTP itself. Such constraints are useful to avoid too many solutions being disregarded by the decision-maker at the end of the optimization because they are not feasible or inadequate. Restricting the search space can make the problem more difficult to solve for the GA but it can also increase the chance of focusing the optimization on only the most interesting solutions and hence discover new compromises.

The constraints to consider in our application can be divided in two groups: (i) constraints ensuring relevant operation of the control law and (ii) constraints only ensuring adequate performance ranges. Three main types of constraint can be used. In order of strictness, these are:

- constraints that must not be violated at any time during the simulation
- constraints that must not be violated for overall performance during the entire period
- constraints that must not be violated during more than a given maximum time.

Examples of constraints applicable to the optimization of BSM1 are listed below together with illustrations of their impact on the optimization results.

4.6.2 Example in the case study

In BSM1 case, it is first desirable to limit the concentration of ammonia in the effluent to ensure good treatment. The legal mean level expressed in BSM1 is 4 $g.m^{-3}$. This level is, however, very difficult to achieve with the influent datasets and the sizing of the processes proposed in BSM1. In order to have a good set of potential solutions, the maximum level thus chosen is 15 $g.m^{-3}$. The value is computed as a mean level during the last week of simulation in order to average the performance over the whole week (constraint type 2):

$$\frac{1}{7}\int_{t_0}^{t_{7\,days}} S_{NH,effluent}(t)\,dt \;\leq\; 15\ g.m^{-3} \qquad \text{(IV.2)}$$

Apart from the ammonia concentration, it is also significant to limit the concentration of total nitrogen in the effluent. This leads to a limitation of the nitrate amount. The maximum mean concentration chosen is 30 $g.m^{-3}$. This value is the legal limit expressed in BSM1.

$$\frac{1}{7}\int_{t_0}^{t_{7\,days}} NGL_{effluent}(t)\,dt \;\leq\; 30\ g.m^{-3} \qquad \text{(IV.3)}$$

For all other legal limits (i.e. *TSS*, *COD* and BOD_5), no constraints are added, as the corresponding concentrations are far below the legal limit during most of the week. Such constraints are therefore not required. They must only be considered when there is uncertainty about the impact of the control law modifications on the corresponding concentration. For instance, if the wastage flow rate was also optimized, a constraint on the *TSS* mean concentration in the effluent would have been needed. Considering aeration control optimization, the modifications made have no impact on the *TSS* concentration in the effluent and very limited impact on the *COD* and BOD_5 concentrations since constraints on ammonia and total nitrogen are already considered (the pollutants corresponding to *COD* and BOD_5 degrade much faster than ammonia in activated sludge systems).

These constraints on admissible effluent quality bring out only interesting solutions. Optimization with only these two constraints is performed based on SNDN and BSM1, with the three objectives already considered in the previous subsection (i.e. mean concentrations of ammonia and total nitrogen in the effluent and energy consumption). The results in terms of objective values are presented in Figure 4.10 with only the final solutions projected on the three sub-axes of the objective space. In this figure, if only the upper left projection on the axes of ammonia mean concentration in the effluent and energy consumption is considered,

the results seem to be fine. But when the total nitrogen mean concentration in the effluent is considered, it is difficult to interpret the results, since the GA has found solutions with similar ammonia concentrations and energy consumption but very different total nitrogen concentrations (the points which seem randomly distributed in the upper right corners of both subplots). The GA did not manage to reduce the total nitrogen concentration of these points.

One hypothesis could be that uncertainty of the numerical solver used in the simulations has caused these worse performance. This is however unlikely to be the case since no solutions which dominate all the others have been found but only competing ones (in terms of dominance). When a problem of solver arises, the simulation results and the performance is quite random and solutions largely dominating other ones are found during the optimization.

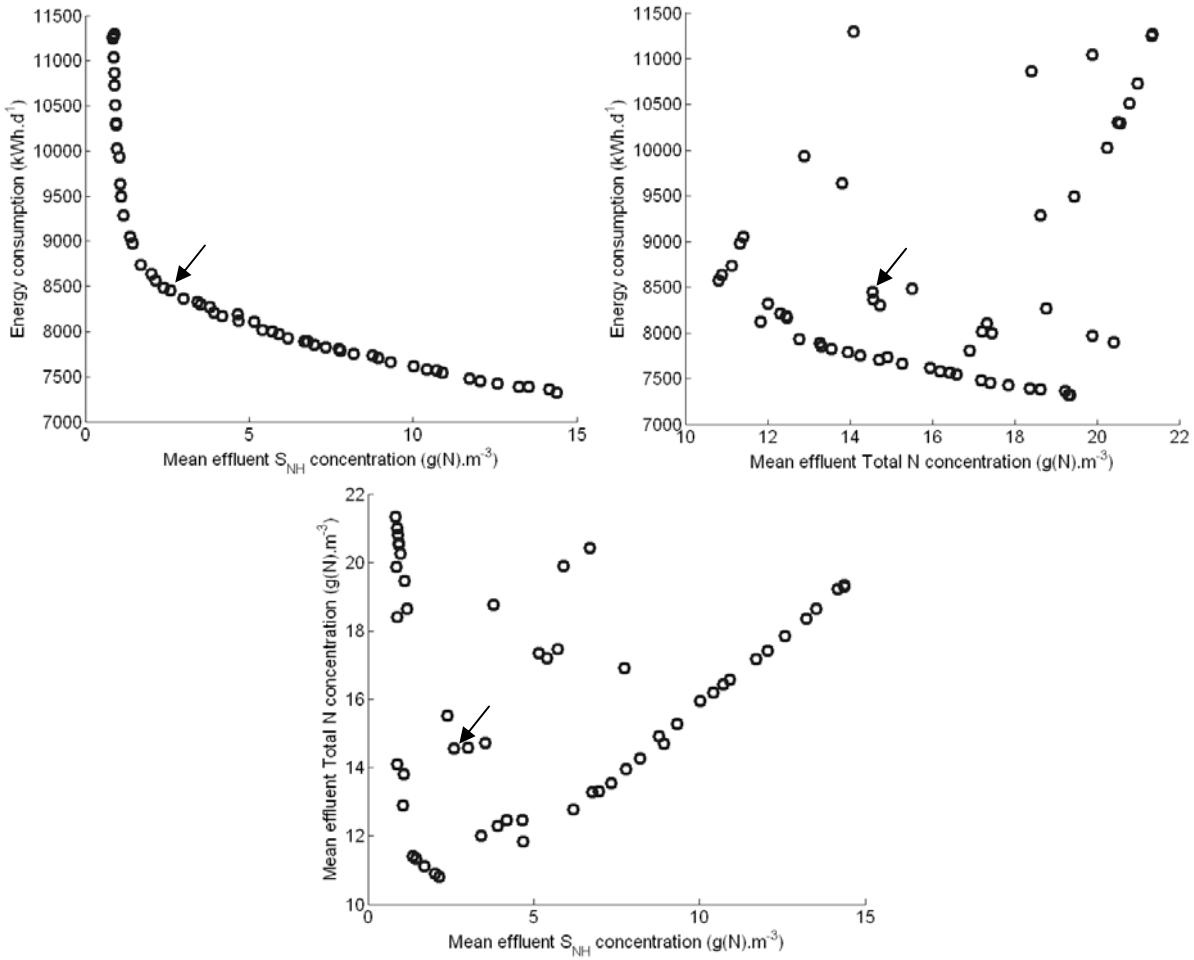

Figure 4.10: Optimization of SNDN with two constraints (on effluent mean concentrations)

In fact, these unexpected results are the consequence from incorrect SNDN control law operations that induced the introduction of incorrect solutions in the population. An example

corresponding to the settings of the point indicated by an arrow in Figure 4.10 is illustrated in Figure 4.11. Concentrations of ammonia and nitrate in the 5th ASU of BSM1 are depicted in the figure, as well as the fixed nitrate set point and the dynamic output of the first controller, which is the dynamic set point of ammonia. For the SNDN settings corresponding to this point, the nitrate set point is never reached. This is linked with the levels of nitrate and ammonia targeted that cannot be reached with the limitation on aeration and process sizes used in BSM1. As a consequence, nitrate evolves almost freely. The performance of such a solution is not interesting at all since it is specific to the influent dataset used for the evaluation and long-term performance will surely degrade significantly. It should therefore be discarded.

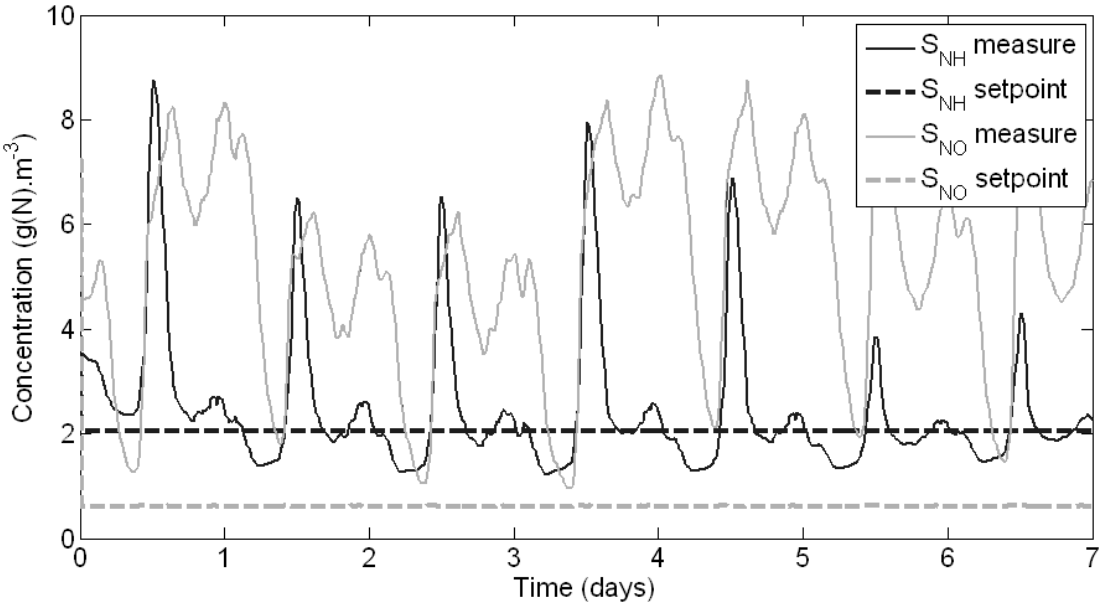

Figure 4.11: Example of problem with SNDN optimization with two constraints

In order to enhance the performance of the GA in this optimization example, it is necessary to add constraints on the performance of the control law. As previously stated, these constraints are necessary to ensure that the control law is really controlling the system and that it is not just the system which is freely evolving because the control laws have reached their allowed limits. In the SNDN case, two constraints are required, one per sub-controller.

For the nitrate controller, it happens quite often that the set point is not reached due to the bounds of ammonia and this is a normal operation of the control law (when the level of ammonia reaches one of the controller limits, the nitrate set point is not followed anymore).

The nitrate set point should therefore only be followed at night, when the load is reduced. As a consequence, the constraint consists of ensuring that the absolute error is below 1 g(N).m^{-3} for more than 3 days (out of 7). This constraint uses a tolerance on the error to take into account the delay between change of actuator and change of measurement (due to oxygen transfer and biomass activity):

$$\begin{aligned} &\int_{t_0}^{t_{7\,days}} f_1(t)\,dt \;\geq\; 3\;days \\ &where\;\; f_1(t) = \begin{cases} 1\;if\;\; \left|Setpoint\_S_{NO} - S_{NO}(t)\right| \leq 1\;g.m^{-3} \\ 0\;otherweise \end{cases} \end{aligned} \tag{IV.4}$$

For the ammonia controller, the set point should be followed during most of the simulation, the exception being the high load period where the maximum admissible air flow rate may be reached. The constraint is hence that the absolute error of this second controller is below 0.5 g(N).m^{-3} for more than 6 days (out of 7). A tolerance on the error is also used:

$$\begin{aligned} &\int_{t_0}^{t_{7\,days}} f_2(t)\,dt \;\geq 6\;days \\ &where\;\; f_2(t) = \begin{cases} 1\;if\;\; \left|Setpoint\_S_{NH}(t) - S_{NH}(t)\right| \leq 0.5\;g.m^{-3} \\ 0\;otherweise \end{cases} \end{aligned} \tag{IV.5}$$

New optimization of SNDN was finally performed with these two additional constraints and the comparison of the results with the previous ones is illustrated in Figure 4.12. The results of the new optimization are very smooth compared to previous ones. This is, of course, not the goal of all optimization problems but is expected for this problem due to the choice of the objectives and the operation of the control law in combination with the activated sludge processes.

Results of the second optimization are competitive with results found in the first one. This proves that the two additional constraints do no penalize optimization.

However, a first remark is that low levels of ammonia are achieved with the first set of constraints but not by the second one. Another remark is that the first optimization produced results with a high level of ammonia (and total N) that are not achieved by the second optimization. In both cases, this is linked to the fact that the controller does not really act on the system in such concentrations ranges. For solutions with a low level of ammonia, this is almost obvious since such solutions are very similar to the ones presented in Figure 4.11 and the results are almost identical. To illustrate what happens with a high level of ammonia, an illustration of a solution (referenced “1” on Figure 4.12) is provided in Figure 4.13. In this

figure, it can be observed that the nitrate set point is not followed most of the time. This will obviously cause problems of robustness in the long-term.

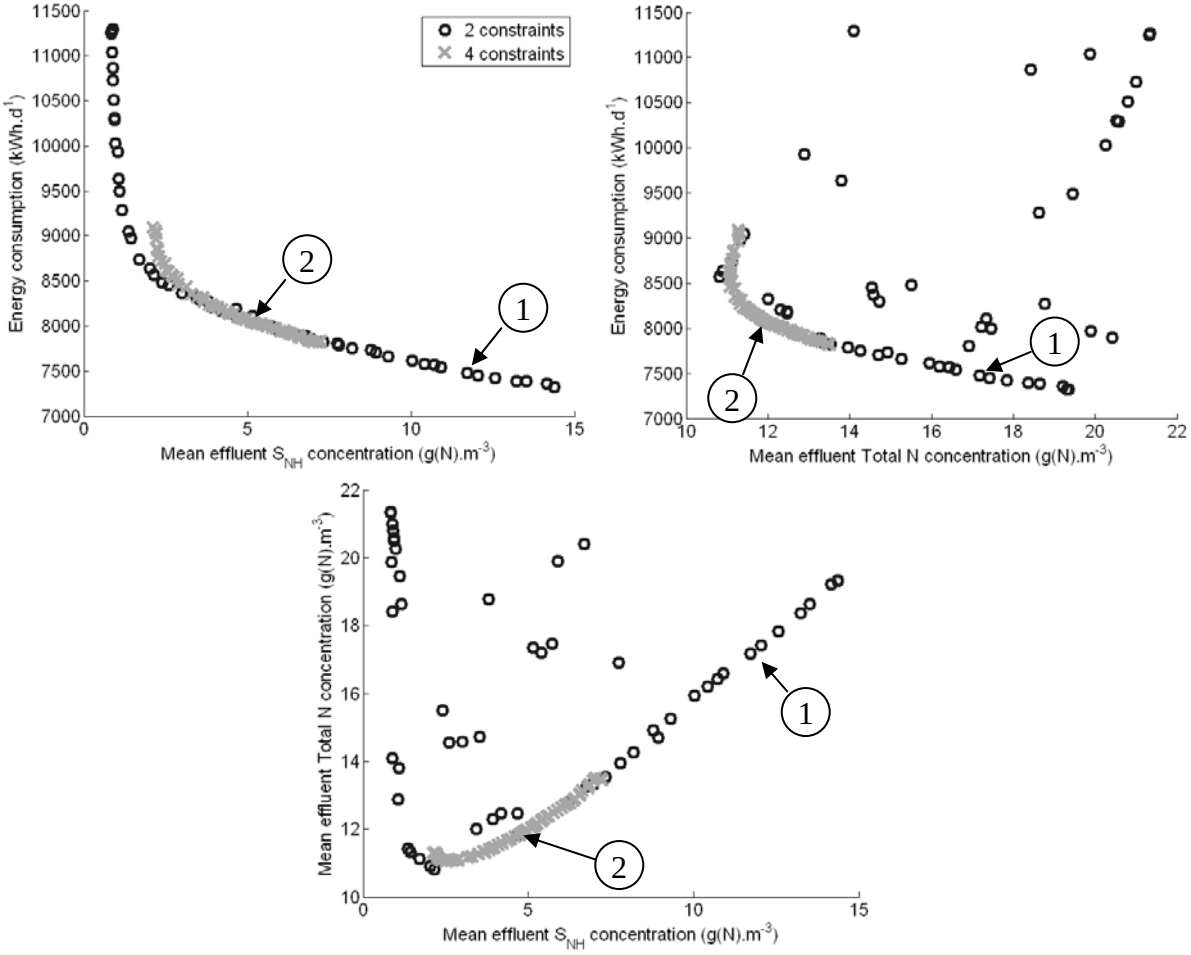

Figure 4.12: Comparison of SNDN optimizations with two constraints (on effluent mean concentrations) and four constraints (on effluent mean concentrations and controller performance)

As a reference, the curves of a solution of the second optimization with four constraints (referenced "2" on Figure 4.12) are depicted in Figure 4.14. It shows that both set points of ammonia and nitrate are adequately followed, except during the peak of ammonia (which is limited by the upper bound) and during the low load at night (which is limited by the lower bound and allows a reduction in the nitrate concentration during this period). This second point clearly represents a normal operation of the SNDN control law.

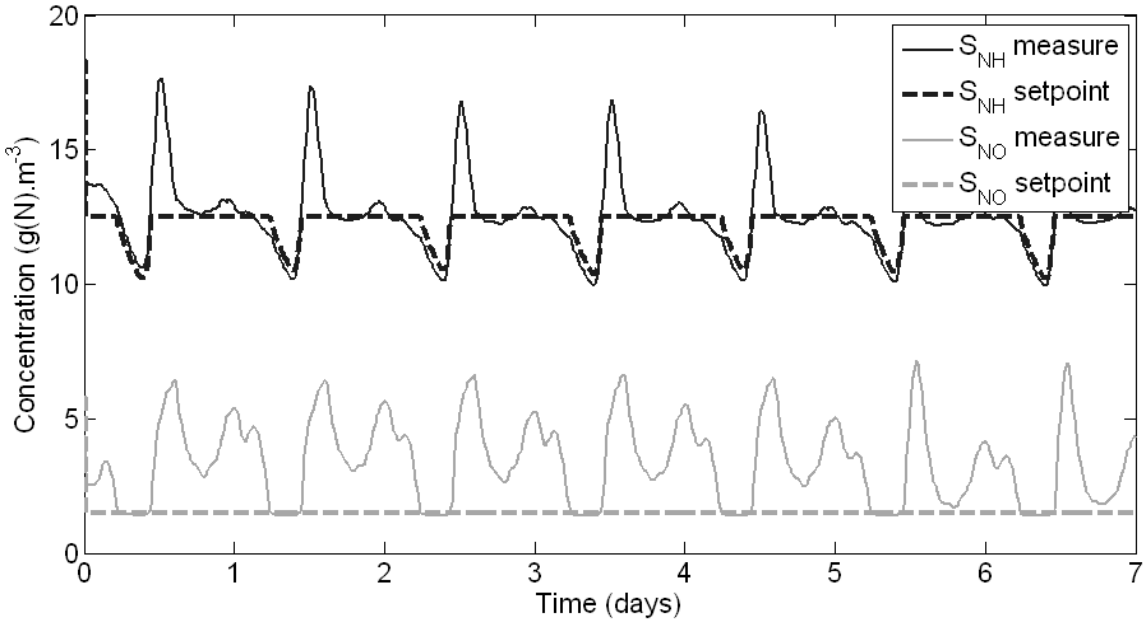

Figure 4.13: Second example of problem with SNDN optimization with two constraints

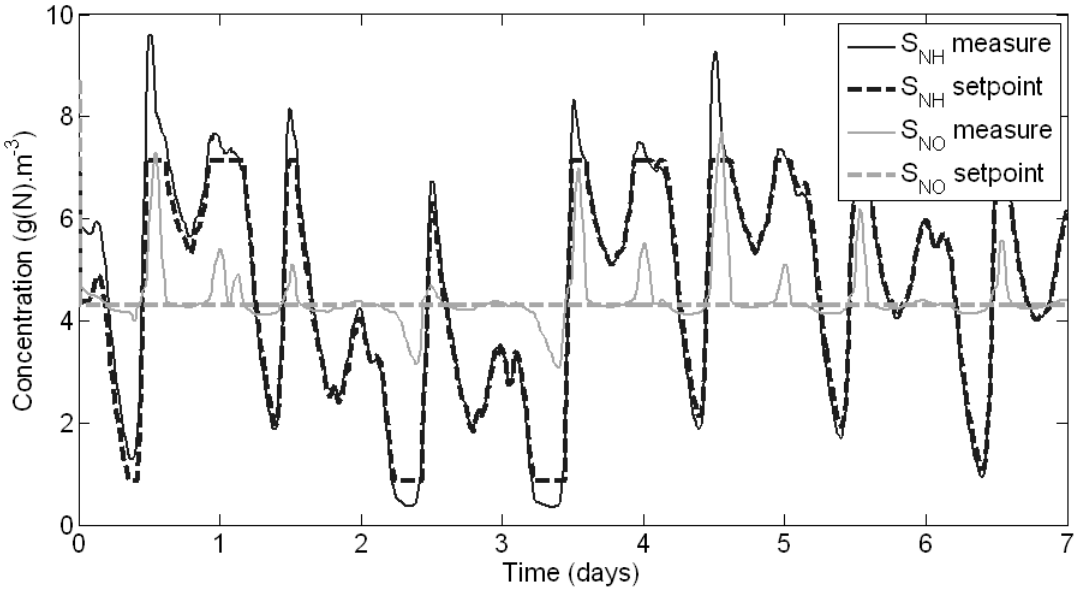

Figure 4.14: Example of good solution obtained with SNDN optimization with four constraints

This was a practical illustration of the importance of the choice of adequate constraints on the SNDN case.

For AMSTAR, the constraint(s) to add for reliable operations of the controller must ensure that the current operating point stays between the two curves f_N and f_{DN} most of the time

(small errors are admissible due to the system response time when the aeration is switched on or off). In our implementation, two constraints are hence considered, one per switching curve:

$$\int_{t_0}^{t_{7\,days}} f_3(t)\,dt \;\le\; 1\,days \qquad \text{where } f_3(t) = \begin{cases} 1 \text{ if current functionning point is above } f_{DN} \\ 0 \text{ otherweise} \end{cases} \tag{IV.6}$$

$$\int_{t_0}^{t_{7\,days}} f_4(t)\,dt \;\le\; 1\,days \qquad \text{where } f_4(t) = \begin{cases} 1 \text{ if current functionning point is below } f_{N} \\ 0 \text{ otherweise} \end{cases} \tag{IV.7}$$

To conclude, this subsection has defined the constraints that are necessary for relevant optimization of AMSTAR and SNDN for BSM1. The benefit of their use was illustrated in the case of SNDN. It is important to note that the controller performance presented here as constraints should not be considered as objective for the optimization since there are no tradeoffs between this performance and the real objectives such as effluent quality or the energy consumption.

4.7 Application of the optimization methodology to the BSM1

After the detailed presentation of the methodology developed in previous sections, this section considers its complete application to the comparison of AMSTAR and SNDN control laws on BSM1. As already explained in section 2.3, the implementation of these control laws is based on measurements of concentrations in the last tank of BSM1 plant. The controller output is a single oxygen transfer coefficient (preferred to air flow rate for literature simulation case studies) which is applied to the three aerated tanks (see Figure 4.15 below). The oxygen transfer coefficient is limited to between 0 and 240 d^{-1} in BSM1.

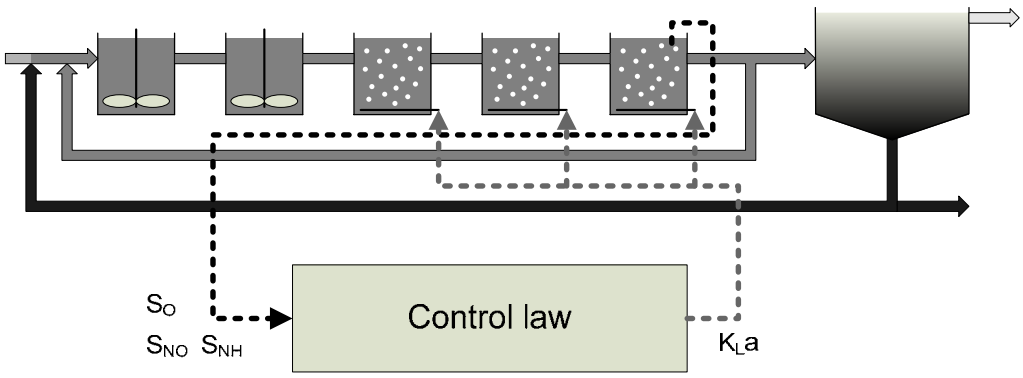

Figure 4.15: Implementation of the control laws to BSM1 layout

The parameters chosen to optimize AMSTAR are the coordinates of the point at which the two criteria functions crosses and the slope of the two curves, considered as straight lines in this study (see section 2.3.5 for more information on AMSTAR). The limits of each parameter to optimize are summarized in Table 4.1. The optimal intersection of the two curves may have negative coordinates due to the minimum durations of each phase. These are the main parameters that control WWTP behavior and therefore its performance. The set point of the oxygen controller is set to 2 $g.m^{-3}$. This parameter is not optimized since the same value must be used in all cases. The chosen value is based on practical experience in order to avoid problems of bulking. Moreover, lower concentrations will, in fact, lead to operations very close to SNDN, which is not the goal of this control law. Choice of the slope of f_N and inverse slope of f_{DN} bounded between 0 and 1 avoids the generation of non-feasible solutions considering these two parameters. It therefore avoids wasting time searching in non feasible spaces.

Table 4.1: List of parameters and their limits for AMSTAR optimization

Parameter	Unit	Lower limit	Upper limit
$[S_{NO}]$ of crossing point	$g(N).m^{-3}$	-4	4
$[S_{NH}]$ of crossing point	$g(N).m^{-3}$	-4	4
Slope of f_N	-	0	1
Inverse of slope of f_{DN}	-	0	1

The parameters used for SNDN optimization as well as their limits are summarized in Table 4.2. It should be noted that instead of directly using the two limits of ammonia, only the lower limit is used and the upper limit is represented as a difference with the lower limit. This is done to ensure that all solutions proposed by the genetic algorithm are feasible (when the upper bound is smaller than the lower bound, the solution proposed by the GA is not-feasible as it does not have any physical meaning).

Table 4.2: List of parameters and their limits for SNDN optimization

Parameter	Unit	Lower limit	Upper limit
Set-point of S_{NO}	$g(N).m^{-3}$	0	10
Lower limit of S_{NH} set-point	$g(N).m^{-3}$	0	10
Difference between upper and lower limit of S_{NH} set-point	$g(N).m^{-3}$	0	10

One may notice that the limits of most parameters are very large, compared to what may typically be expected in practice. This is done in order to avoid restricting the optimization to

solutions of which we are already aware and to give the optimization procedure a chance to propose new solutions.

In this application, three objectives are optimized: the mean concentrations of ammonia and total nitrogen in the effluent and the energy consumption. These three objectives are highly dependent on the chosen parameters and they typically correspond to the decision-maker's point of view in France. In order to ensure that only reliable and interesting solutions are selected, the constraints previously defined in subsection 4.6.2 are also considered.

4.7.1 Tuning of the GA parameters

Before optimizing previous control laws, the GA internal parameters should be adequately chosen. It is clear that the performance of a GA on a specific problem can be enhanced with a fine tuning of its parameters. Considering NSGA-II, only two main parameters need be adjusted to each problem: the population size and the number of generations (when no more sophisticated termination criteria is used). The other parameters are very robust to provide solid performance on most problems and will be set to their default value for all optimizations. These default parameters are the probability of crossover (0.9) and mutation (0.1) and the indexes for the distribution of real values during crossover (10) and mutation (20).

The adjustment of the population size and number of generations is more critical since incorrect values might prevent the convergence of the algorithm to the Pareto front. To find the best values, expert knowledge is the best solution. Another solution is to use the parameters of another problem with similar characteristics. If this is not available, many parameters have to be tested with multiple repetitions of the same parameters to ensure that incorrect performance is not just the consequence of "bad luck" in the stochastic evolution of the GA.

In our case, we did not have any similar problem at hand. Therefore, in order to assess the best values of these two parameters, SNDN control law is optimized and various GA parameters are tested. Five population sizes are assessed: 12, 20, 48, 100 and 200 individuals. For each population size, four repetitions of the optimization are done and the total number of WWTP performance evaluations is 6000 (so for a population size of 12, 500 generations are computed, for a population size of 20, 300 generations are computed, and so on).

In order to compare the GA performance in these various runs, the performance metrics defined in subsection 3.2.4 are used, namely the convergence metric (measure of the distance between the current Pareto front and the optimal one) and the diversity metric (measure of the average distance between all points of the current Pareto front). The optimal Pareto front is assessed based on all repetitions that are made for this test. All non-dominated solutions found are selected in this front.

Raw results of the convergence metric computation for the four repetitions are provided in Figure 4.17 and those of the diversity metric are presented in Figure 4.18. Each subplot in these figures is related to a different population size. For each test, four repetitions are made and hence four curves are provided, one per repetition. The convergence metric is plotted on a logarithmic scale for better visualization and analysis of this criterion.

For better comparison of the various population sizes, Figure 4.16 presents the mean of the convergence and diversity metrics for each population size and GA performance criteria. This computation hides the variability of the repetitions but it allows better comparisons with a superposition of the mean curves.

Considering only the convergence metric (which should be minimized), it is obvious that sizes of 12 and 20 individuals are inadequate since the convergence is too random and the final values reached are not very good. It is also clear that 200 individuals is an unsuitable size since the convergence is very slow and even if the first generations are competitive, the last ones are not any more. On the other hand, sizes of 48 and 100 individuals are quite competitive. The former is capable of very fast convergence but the convergence can then degrade somewhat and depends on the repetition (stochastic effect, but this is not as significant as it seems to be on the figure since a logarithmic scale is used). Optimizations performed with the later population size show slower convergence but the variations between the repetitions are lower, especially after the 40^{th} generation.

If the diversity metric (which should be maximized) is considered, the best solutions are once more obtained with sizes of 48 and 100 individuals. The diversity of solutions obtained with the first one is better than with 100 individuals except at the end. However, no clear conclusion on the variability of the repetitions can be made in this case.

This example shows that it is not always possible to make definitive conclusions about how genetic algorithms function, mainly due to their stochastic nature. Nevertheless, in this specific example, it is possible to conclude that a population of 48 to 100 individuals should be chosen. 100 generations seem to be sufficient if the population size is near 48. For population sizes closer to 100 individuals, a choice of 60 generations seems to be more adequate.

For our application, a population size of 60 individuals is selected and 100 generations are performed. The investigations presented here can of course not be repeated for all optimization problems due to the long computing time required but, since GAs are very robust, the chosen parameters can be used for the two control laws applied to BSM1 and Cambrai case studies.

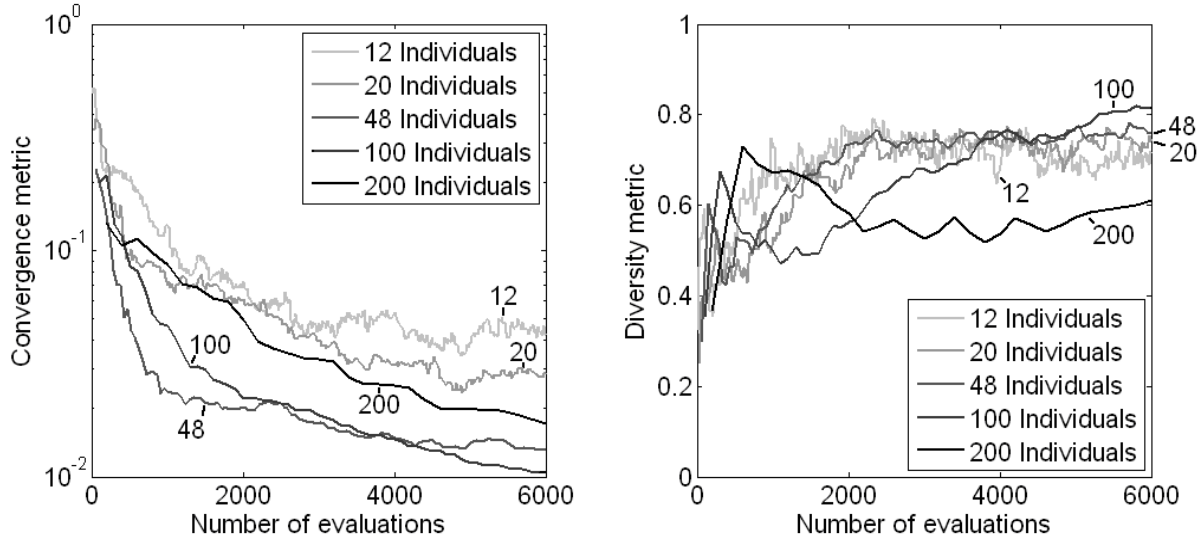

Figure 4.16. Mean convergence and diversity metrics

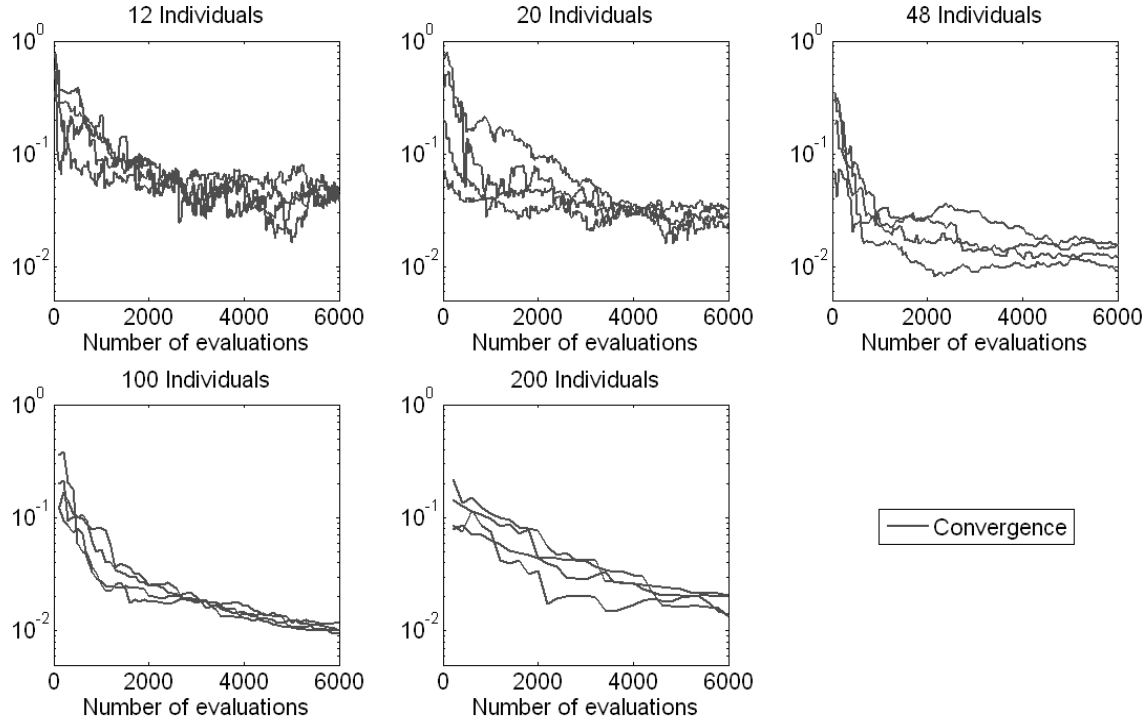

Figure 4.17. Convergence metrics for different potentials sizes (12, 20, 48, 100 and 200) with four repetitions in the case of SNDN optimization on BSM1

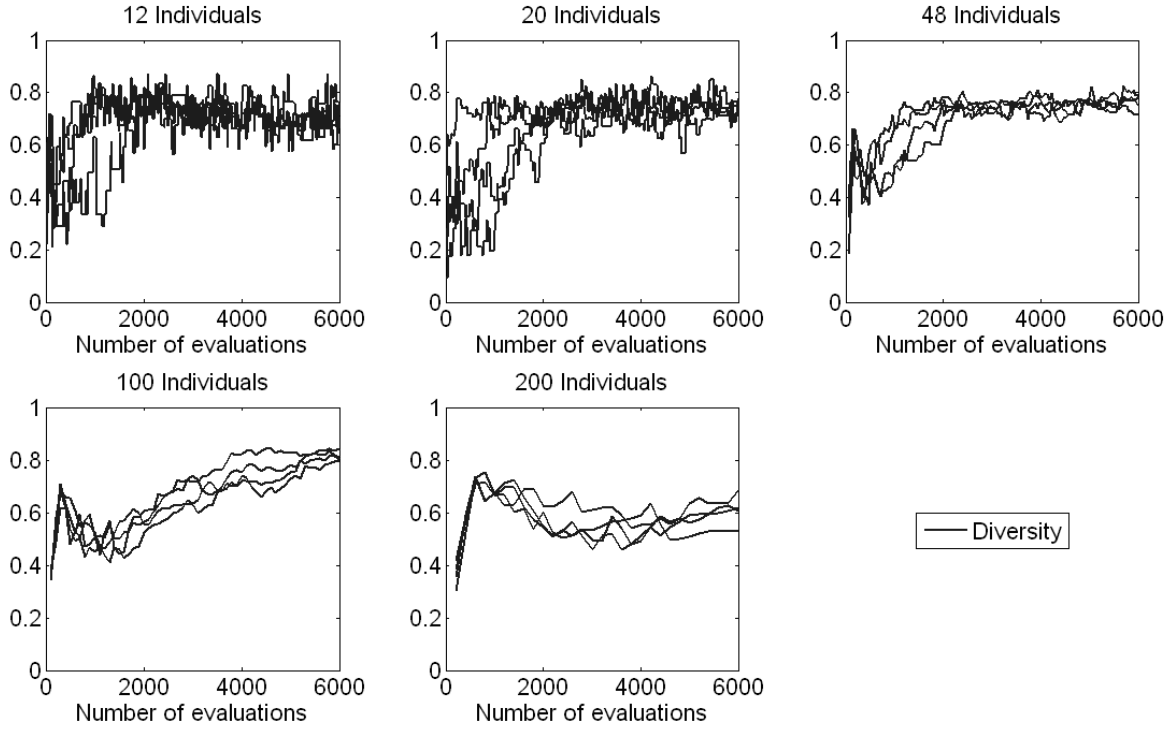

Figure 4.18. Diversity metrics for different potentials sizes (12, 20, 48, 100 and 200) with four repetitions in the case of SNDN optimization on BSM1

4.7.2 Short-term performance at the end of the optimization

With the chosen population size (60) and number of generations (100), both control laws AMSTAR and SNDN are optimized on the BSM1. Three objectives are considered (mean concentrations of ammonia and total nitrogen together with energy consumption). Evaluations of the performance are based on the rain weather input dataset of BSM1, after stabilization of the process obtained with DWID (see section 4.2). For the comparison of the two control laws, the performance of the final solutions obtained is projected on each pair of objective subspaces and depicted in Figure 4.19 and Figure 4.20. The optimized solutions of AMSTAR are represented by a star while those of SNDN are represented by a circle.

In Figure 4.19, the Pareto front is represented in three dimensions with bold points while its projections on the three two dimensional sub-axes are represented in normal points.

For reference purposes, the performance of BSM1 and modified BSM1 are also plotted in Figure 4.20. The modified BSM1 corresponds to simulations where the wastage flow rate is reduced to 220 $m^3.d^{-1}$ for both cases (i.e. open and closed loop) and the sludge recycle flow rate is controlled to 100 % of the influent flow rate for the closed loop case. Since these modifications are used in SNDN and AMSTAR evaluations (see subsection 2.4.4), they need to be reflected in the controls proposed in BSM1 for better comparison.

The first remark before analyzing these figures is that three objectives might imply the Pareto front is a surface. In fact, in this case the Pareto front is just a path in three dimensions, which is therefore very easy to observe in each of its projections. With regard to AMSTAR, the points are more jittered, especially considering solutions with low levels of ammonia and total nitrogen in the effluent (or high levels of energy consumption). This is due either to lower accuracy of the computations or weakness of controllability in this region.

It is then possible to compare the control laws based on this figure. However, as explained before, this comparison should not be based on short-term performance but on median long-term performance, which is more representative of overall performance of the controller schemes on BSM1.

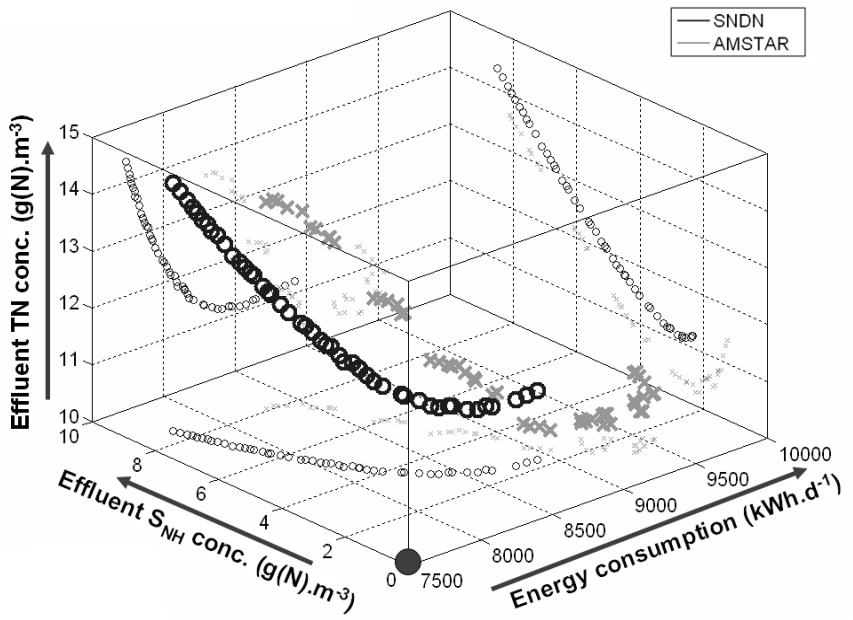

Figure 4.19: 3D short-term performance obtained with SNDN and AMSTAR for the BSM1

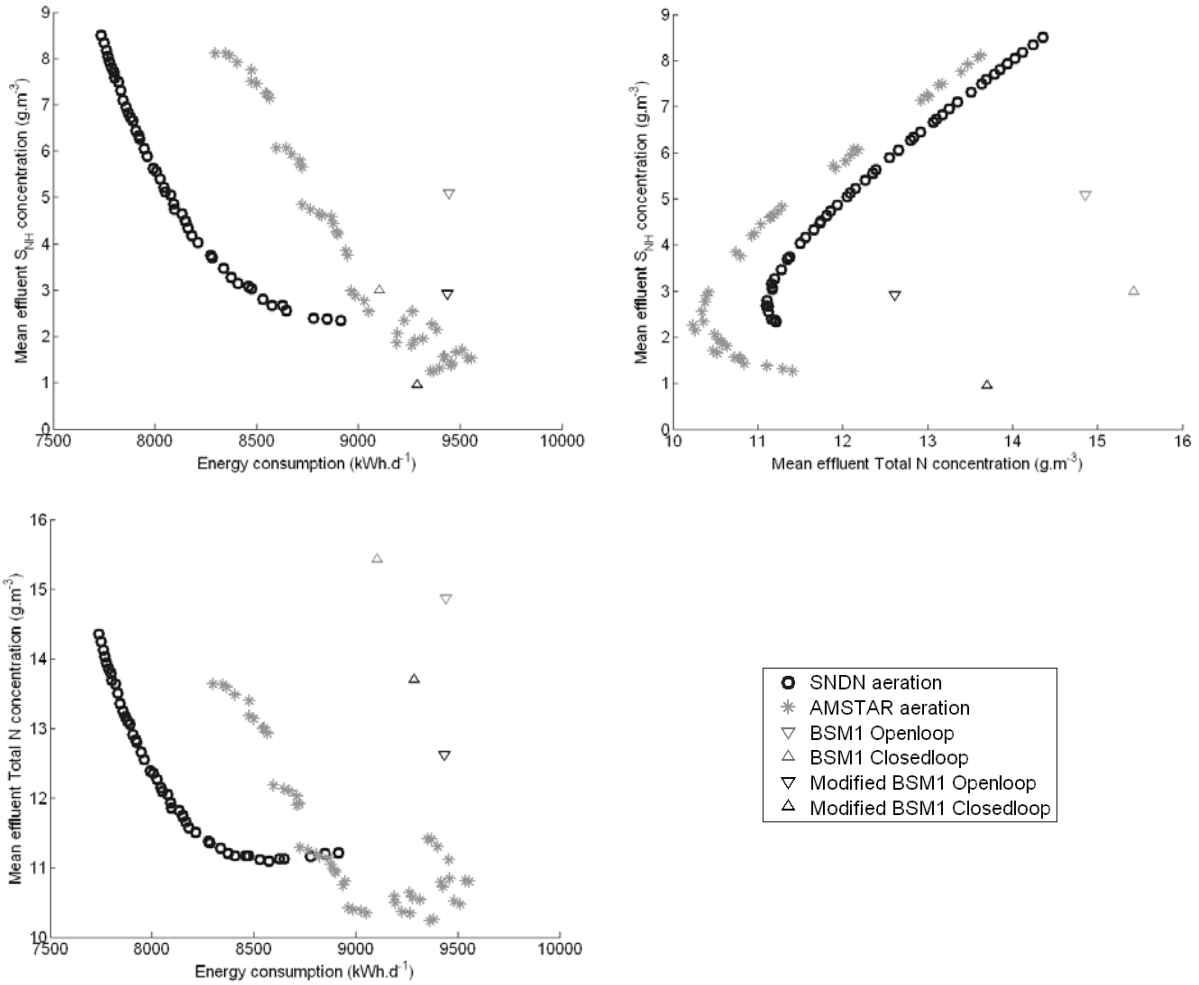

Figure 4.20: 2D projections of short-term performance obtained with SNDN and AMSTAR for the BSM1

4.7.3 Long-term evaluations of the robustness, median performance and comparison of the two control laws

As previously stated, the following step is to perform the long-term simulations, in order to evaluate the robustness of the solutions found and to perform a comparison of various control laws. For these simulations, the influent dataset of BSM1_LT is used as previously explained in section 4.4. Daily performance is then computed for each solution and their 5th, 50th and 95th percentiles are used for visualization and interpretation of the results.

The median long-term performance obtained for AMSTAR and SNDN as well as the modified versions of BSM1 are plotted in Figure 4.21 (in black), together with the short-term performance (in grey) obtained during the optimization and already presented in Figure 4.20.

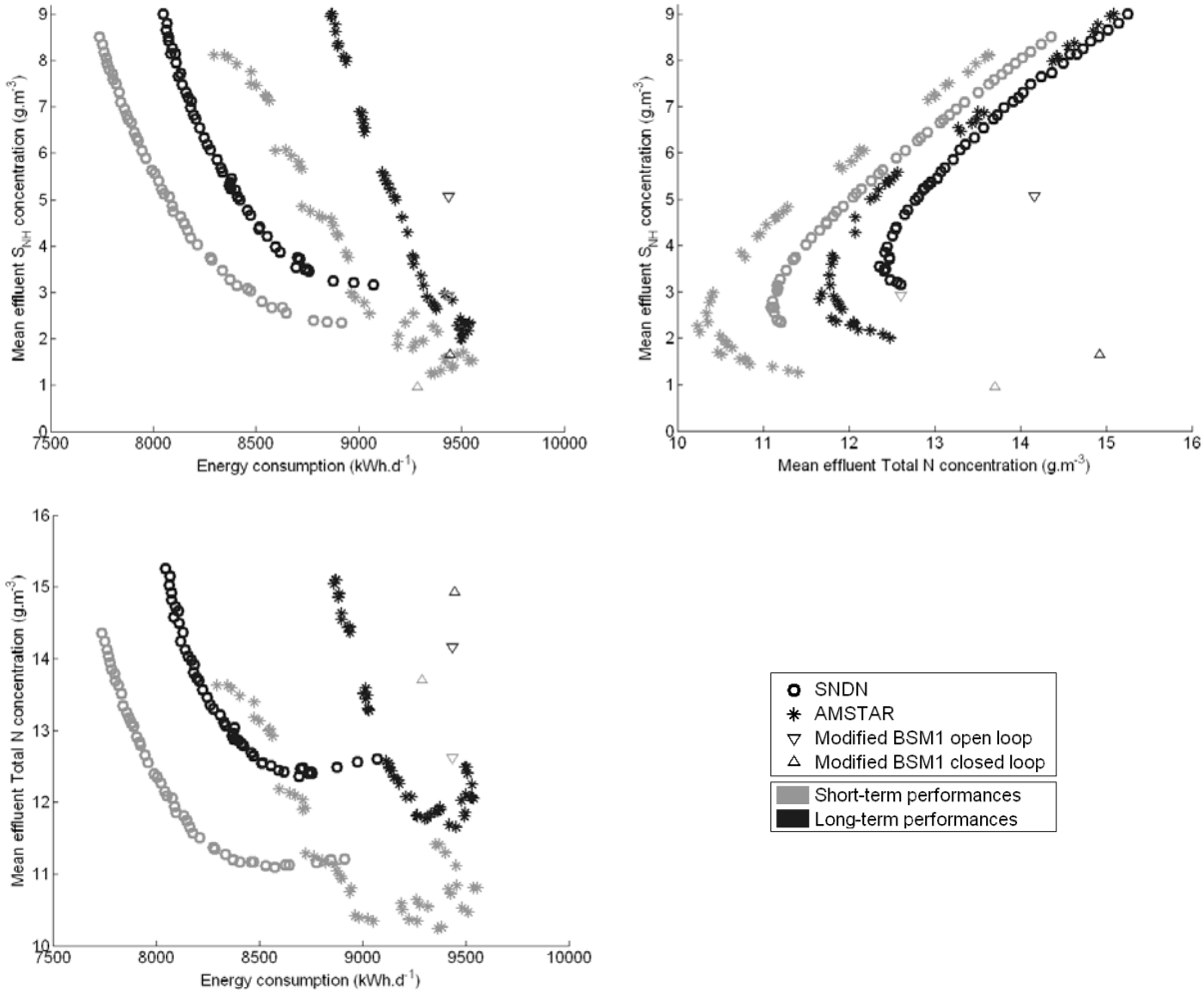

Figure 4.21: Comparison of short-term and long-term performance obtained with BSM1

The first observation is that many performance deviations are observed between short-term and long-term median performance for both control laws: between +2% and +6% for energy

consumption, +4% to +35% for mean ammonia concentration in the effluent and +4% to +12% for mean total N concentration in the effluent. These deviations are significant, especially for the mean ammonia concentration, which is, by the way, one of the most significant objectives of BSM1 as it is the legal limit which is the most difficult to meet. This observation clearly shows that only the mean performance based on the long-term simulation and including various events can be presented to the decision-maker and the comparisons of control laws or the choice of an optimal point must not be based on short-term evaluations performed during the optimization.

The second observation on this figure is that short-term and long-term performance curves are shaped almost identically. This is a very good indicator of the robustness of the solutions found during optimization. This will, however, be further analyzed with the 5^{th} and 95^{th} percentiles presented later.

Considering only the median long-term performance (in black), many outcomes can be derived from this figure in terms of comparison of the control laws. The first remark is that both optimized control laws perform better than open loop settings proposed in the BSM1. This is quite normal since the open loop control cannot measure disturbances and therefore does not challenge closed loop controls (or this would mean that the controller is very poorly tuned or parameterized).

On the contrary, the modified closed loop version of BSM1 is capable of challenging both AMSTAR and SNDN control laws since it achieves the lowest mean concentration of ammonia in the effluent. However, this operating point is achieved at the cost of high energy consumption and high concentration of total nitrogen in the effluent. It is, in fact, quite a logical and extreme result: if much oxygen is injected in the system, almost all ammonia will be depleted but this will inhibit the denitrification in the aerated tanks and consume considerable energy. As the amount of oxygen is still controlled in BSM1 control scheme, it is still not as extreme as in BSM1 open loop case but a lot of oxygen is provided (a concentration of 2 g(-COD).m^3 in the last tank means lower oxygen transfer efficiency and the resulting portion of oxygen injected is simply released into the atmosphere).

For better comparison of the three alternatives (i.e. modified BSM1 closed loop, AMSTAR and SNDN), Figure 4.22 presents the 5^{th}, 50^{th} and 95^{th} percentiles of the long-term evaluation.

The 50th percentiles are plotted with filled lines while the 5th and 95th percentiles are plotted with dotted lines. The first observation on this figure is that the simple closed loop control proposed in BSM1 is only interesting if a very low concentration of ammonia in the effluent is targeted, as already explained, based only on the median performance. It will hence not be considered anymore and only AMSTAR and SNDN will be compared.

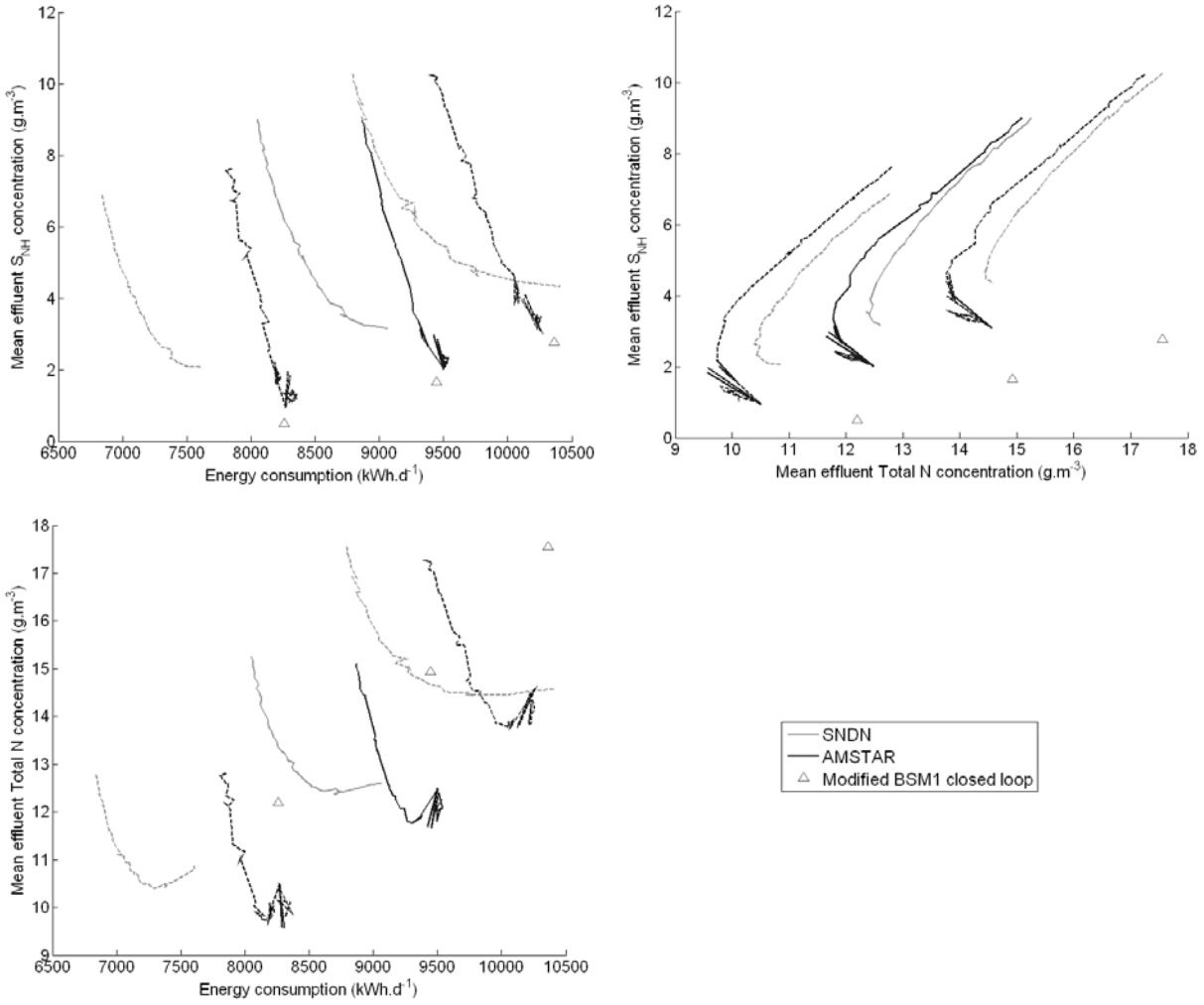

Figure 4.22: 5th, 50th and 95th percentiles of daily long-term performance obtained with BSM1 controlled with a modified version of the closed loop control proposed in BSM1, with AMSTAR and SNDN

Many analyses of these two control laws can be made with the figure. The first observation is that SNDN is better than AMSTAR for a large proportion of the solutions. For the same mean concentration of ammonia in the effluent as AMSTAR, SNDN consumes on average 8 % less energy (this reduction is observed in the upper left portion of the figure, for a median concentration of ammonia in the effluent between 5 and 9 $g.m^{-3}$). This difference is bigger for lower concentrations of ammonia in the influent (5th percentiles) or lower for bigger concentrations of ammonia in the influent (95th percentiles). On average, AMSTAR releases

between 0.5 g(N).m^{-3} and 1 g(N).m^{-3} less total nitrogen than SNDN in this range of controller settings.

For the same median concentration of total nitrogen in the effluent as AMSTAR, SNDN consumes on average 8.5 % less energy. This reduction is observed in the lower left part of the figure, for median concentrations of total nitrogen between 12.5 and 15 g(N).m^{-3}.

Hence, for median concentrations of total nitrogen in the effluent between 12.5 and 15 g(N).m^{-3} or median concentrations of ammonia in the effluent between 5 and 9 g(N).m^{-3}, SNDN is superior to AMSTAR when applied to the BSM1 test case. The figure also indicates that for lower median concentrations, the difference between both control laws steadily decreases but is still at the advantage of SNDN for median concentrations of ammonia above 3.2 g(N).m^{-3} and median concentration of total nitrogen above 12.3 g(N).m^{-3}. These two points are also the lower operating conditions achieved by SNDN. AMSTAR is then the only solution to reach lower concentrations in the effluent.

The last observation is that the evaluations of AMSTAR for low level of ammonia and/or total nitrogen are much jittered on the figure but the same kind of behavior is observed on all percentiles. The problem comes here from the optimization, for which a better parameterization of the AMSTAR control law or a more robust evaluation of short-term controller performance is required.

The observations made on this case study are very instructive. First, the difference in energy consumption is at the advantage of SNDN for most operating conditions. This difference is not very big but it should be noted that this evaluation is based on models without any impact of the operating conditions on the model parameters. Bigger differences can hence be expected with real processes. Finally, AMSTAR achieves very low concentrations of pollutants in the effluent that are not possible with SNDN. This is in fact due to the design of AMSTAR and to the high loading of BSM1 (compared to ATV and EPA recommendations). AMSTAR sequences the aeration and thus provides the best conditions and biological kinetics for both reactions of nitrification and denitrification. This is, however, reached at the cost of higher energy consumption. On the other hand, SNDN is more targeted for WWTPs with normal loadings or low loadings for which it is possible to provide exactly the amount of

oxygen required and to balance both reactions of nitrification and denitrification, even if their kinetics will be slower compared to AMSTAR.

It is hence obvious that this comparison of the control laws and the conclusions obtained are only applicable to the case of the BSM1. The generalization of these conclusions to all WWTPs is indeed a risky choice since each WWTP specific sizing and influent characteristics can induce different compromises between both control laws.

4.7.4 Results of the optimization in terms of controller settings

After comparing the performance obtained with the two control laws AMSTAR and SNDN, it is also possible to analyze the decision variable values for each solution of the two problems. This may lead to conclusions about some links between the parameters, and may help to detect non-relevant solutions. It is necessary for the real implementation of the solution that will be selected by the decision maker.

The parameter values corresponding to the different points of the Pareto front are illustrated in Figure 4.23 for SNDN. Parameters are plotted according to the energy consumption of each solution. This parameter has the advantage of being monotone in the solutions found and it allows for a better visualization of the results. This is particularly important when the spacing of the solutions on the Pareto front is not regular (this it is not really significant here but it will be for AMSTAR).

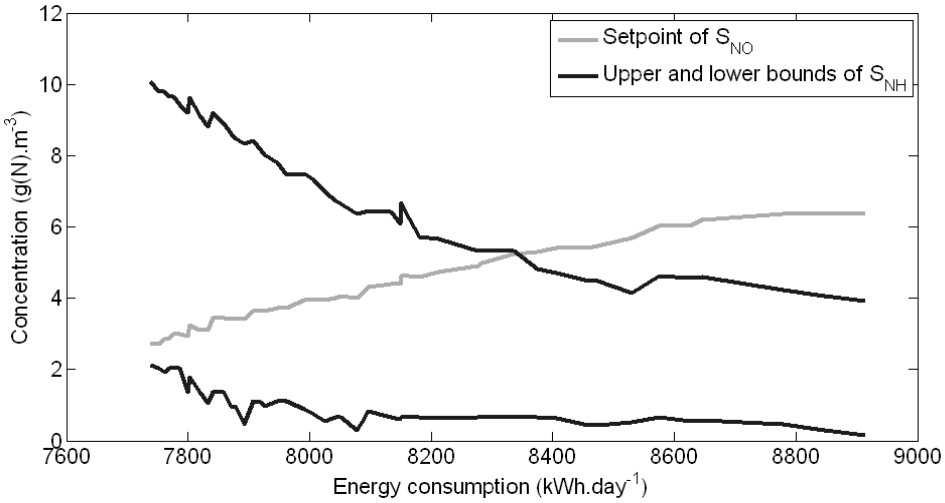

Figure 4.23: Optimal settings found for the continuous aeration for the BSM1

In this figure, energy consumption increases when the bounds of the ammonia controller decrease. This is normal since the decrease of ammonia concentration in the effluent implies the addition of more oxygen, hence consuming more energy. This reduction in the concentration of ammonia implies the production of more nitrates. As the denitrification capacity is limited, the set point of the nitrate controller needs to be increased for proper operation of the control law. One may notice that the lower bound of ammonia is almost constant. As the error on the sensors typically used in real WWTPs is around 0.5 $g.m^{-3}$, the lower bound of ammonia can probably simply be set to 1 $g.m^{-3}$ without major consequences on the control law performance. This may even be the subject of new simulations of each final solution with a fixed lower bound.

The parameter values obtained for AMSTAR are depicted in Figure 4.24. Their interpretation is more difficult. First, some points corresponding to the range of energy consumption between 8700 and 8950 $kWh.d^{-1}$ probably correspond to simulations of limited accuracy. For other solutions, it is relatively difficult to derive a tendency in the evolution. Future work will be required to fully interpret these results. A new parameterization of the problem can also be studied and may lead to better results.

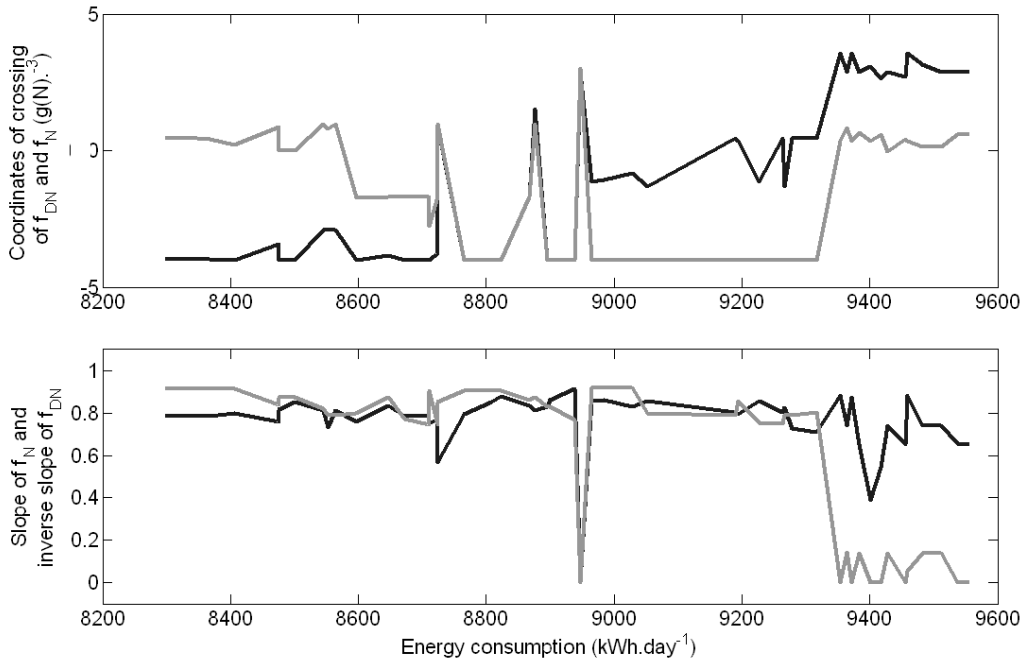

Figure 4.24: Optimal settings found for the sequenced aeration for the BSM1

This section has presented the application of the methodology on BSM1 case study for the comparison of AMSTAR and SNDN. This methodology has proven to be efficient and to provide clear insight into the performance of each control laws, as well as their corresponding settings. The full information can clearly help a decision-maker to decide if he has to choose between both control schemes. The use of multi-criteria decision-making tools can even help him in the analysis of the results. Here, only three objectives are considered so the interpretation is quite easy but when more objectives are considered and/or more complex Pareto fronts are obtained, such tools will be necessary for proper visualization and analysis of the results. Such developments are thus a good perspective of this methodology that will be further detailed in the last chapter of the thesis.

4.8 Conclusion

This chapter has presented the methodology development based on BSM1 case study. This methodology first involves a new simulation procedure ensuring convergence of the simulations before the performance evaluations. This ensures reliable objective values for the optimization algorithm. The influent dataset used for the convergence of simulations is the DWID of BSM1, while RWID of BSM1 proved to be more reliable for performance evaluations. For the evaluation of long-term performance, simulations are performed with the long-term influent dataset of BSM1_LT representing 609 days of operation. This allows the visualization of the median and 5^{th} and 95^{th} percentiles of daily performance and gives insights into the real performance of the studied control law. The importance of the choice of objectives and constraints has also been highlighted in this chapter, together with their respective influence. Finally, the methodology developed has been applied to BSM1 for the optimization of two controls laws, SNDN and AMSTAR. This has provided us with a clear comparison of both control laws, highlighting their respective domains of validity and performance.

The main weakness of this methodology is that neither short-term nor long-term variations of the temperature have been considered so far. In real WWTPs, they can however imply big changes in terms of performance since the reactions kinetics are influenced by the mixed liquor temperature. The inclusion of this aspect in the procedure might however be difficult since the long-term modifications are very slow (one cycle of variation takes one year).

Another aspect that could also be considered is the inclusion of more exact models of sensors and actuators as proposed in the BSM1_LT. This would allow a better insight into actual performance and would also avoid providing solutions that are not feasible in reality.

The main perspectives of this methodology is first its application on real case studies such as the case of Cambrai WWTP in northern France, which is presented in the next chapter. Another perspective is the application to more complex control schemes. For instance, instead of using constant set-points during the whole day and week, variable set-points might be considered according to the current incoming load for instance.

Chapter 5 - Application of the methodology to Cambrai WWTP

Chapitre 5 - Application de la méthodologie sur le cas de la station d'épuration de Cambrai

Having applied the optimization methodology proposed in this thesis to a literature case study, this chapter now applies the methodology to a real plant, the Cambrai WWTP in northern France. The goals of this section are (i) modeling of the secondary treatment influent, (ii) modeling of the secondary treatment itself and (iii) the application of the methodology developed in this thesis for the optimization of two control laws for sequenced and continuous aeration. First the case study is presented, then each section of this chapter details the results for one of these objectives, and finally, general conclusions are given.

5.1 Presentation of the case study

5.1.1 Main presentation

The city of Cambrai is located in northern France, 159 km from Paris and 62 km south from Lille. It has a population of 33,716 according to the last census in 1999. The main WWTP is located in a small town named Neuville St-Rémy and the design load of this WWTP is 63,000 population equivalent[1]. Few polluting industries are connected to the sewer system of this WWTP and the main wastewater comes from households in nearby towns and cities.

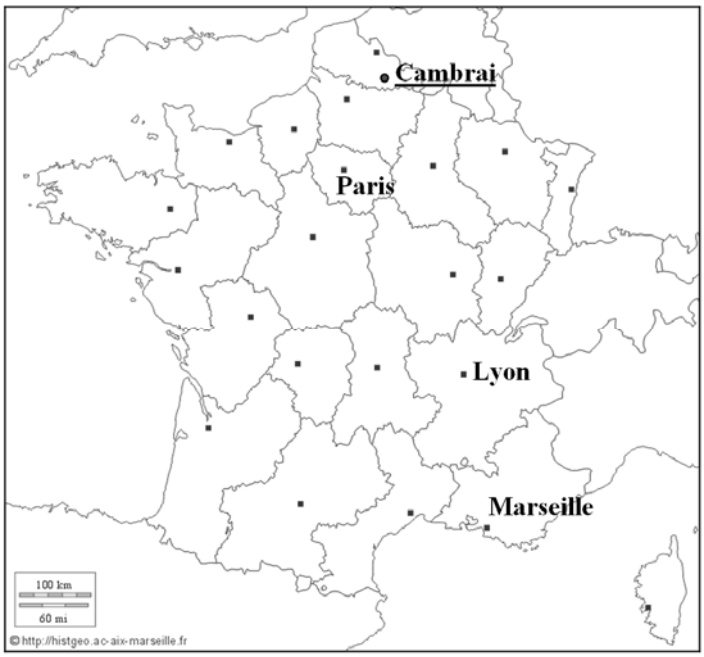

Figure 5.1: Geographical location of Cambrai

[1] The European directive of 21st May 1991 defines a population equivalent as an organic load whose BOD_5 is equal to 60 $g(O_2).d^{-1}$

The climate in this city is typically oceanic, with many days of rain (120 days per year, compared to 111 days in Paris or 57 in Marseille) but with a modest total height of rain (642 mm in Cambrai, same as in Paris, compared to 544 mm in Marseille)[2].

In the WWTP, primary treatment of the wastewater is based on mechanical screening and removal of grease and sand via fine air bubbles. Grease is then treated with special activated sludge and high concentrations of oxygen with the Biolix® process. The remaining pollution is then injected in the secondary treatment. Sands are washed and then disposed or composted.

Secondary treatment is based on activated sludge units and gravity settlers functioning in a continuous flow as illustrated in Figure 5.2. Two parallel lines of processes are used. The repartition between the two lines is assumed to be the same[3]. The raw wastewater is first separated between these two lines (1) and mixed (2) with recycle sludge (8) from the secondary clarifier (6). This mix of activated sludge and wastewater then goes into an anaerobic zone (3) enabling biological phosphate removal. Iron(III) chloride ($FeCl_3$) is also added (10) for the removal of remaining phosphate via precipitation. Activated sludge then goes into a carousel (4) where intermittent aeration is currently applied with small air bubbles generated with a suppressor and porous membranes (not represented). Porous membranes are positioned in only half of the circumference of the carousel. A deaerator (5) is then used to remove air and gaseous nitrogen bubbles that can prevent proper settling. The activated sludge then goes into the secondary clarifier (6) for the separation of solid compounds. Sludge is extracted to a small temporary storage (8) before being recycled with the inlet of the secondary treatment (2) and a small fraction is removed when sludge is in excess (9). Cleaned water (7) going out of the secondary clarifier is finally released to a river flowing nearby.

The sludge treatment is simply based on centrifugation with a first addition of polymer and a final addition of quick lime to inert the sludge and control the dryness. The sludge is then released for disposal or composting.

[2] Data from Infoclimat (www.infoclimat.fr) based on averages between 1961 and 1990
[3] None of the measurements made on the process indicated a difference of repartition between the two lines and a CFD computation was performed which did not show any difference either.

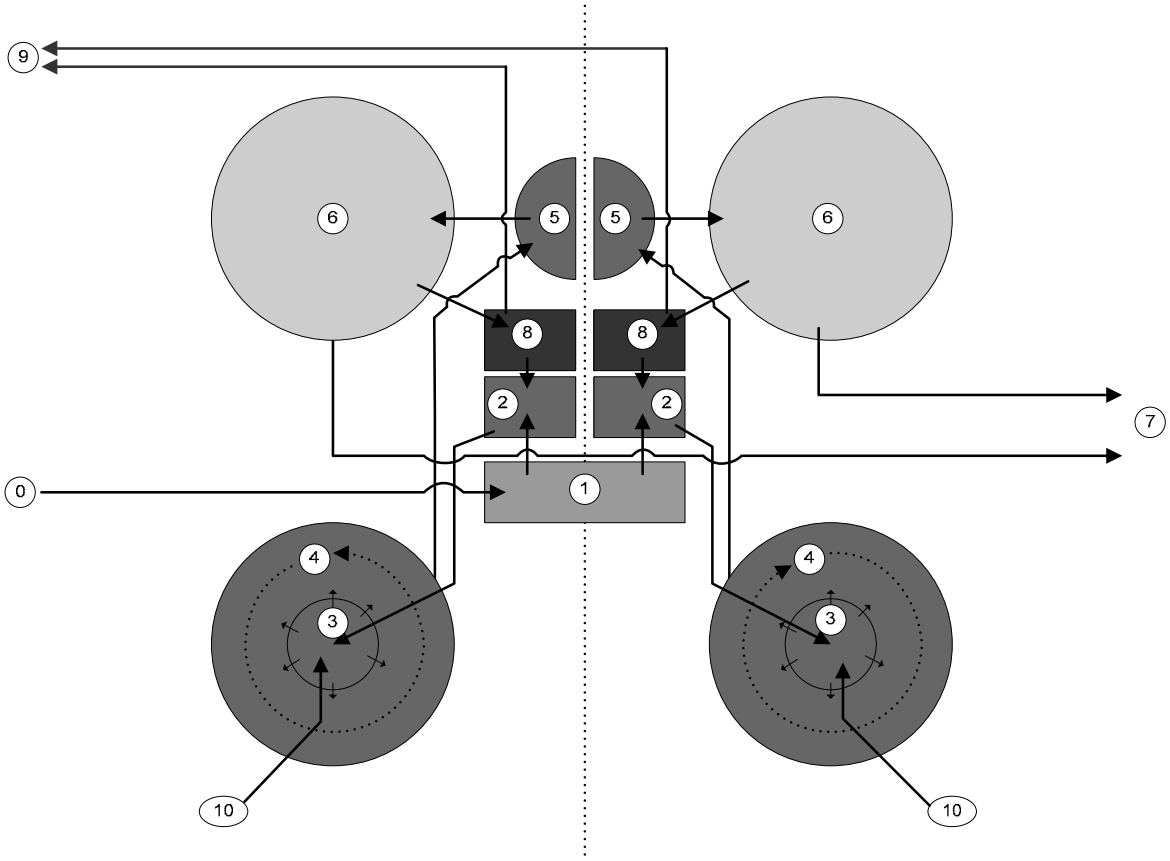

Figure 5.2: Physical layout of the WWTP and arrangements of individual processes

5.1.2 Key figures

Key physical process figures used in the WWTP modeling are summarized in Table 5.1. They represent one line of treatment (the overall volumes are hence doubled at the WWTP).

Table 5.1: Key physical parameters of one WWTP treatment line

Parameter	Value
Anaerobic basin volume (3)	800 m^3
Aerated basin volume (4)	5600 m^3
Deaerator volume (5)	80 m^3
Secondary settler volume (6)	4074 m^3
Secondary settler height (6)	3.5 m
Index of secondary settler feed layer (1 = top layer)	7 out of 10
Sludge recycling flowrate (8)	100% of influent flowrate
Average sludge residence time (9)	18 days
Addition of $FeCl_3$ (10)	11g/m^3 of influent

For calibration of the WWTP model, average influent flow rate and loads are also required and they are determined from close monitoring of the plant. This monitoring is based on 24h averages made with grab samples relative to the influent flow rate.

Analyses of TSS, COD, BOD_5, NH_4, TKN, NO_2, NO_3, TN and total phosphorus are performed during some days of the week, based on a complex scheme designed to provide information for every day of the week in a rotating manner. On average, COD analyses are performed 4 times per week, TSS twice per week and other parameters once per week. Influent flowrate and rain depth are continuously recorded in the WWTP SCADA system. Rain depth is measured based on a sensor located at the WWTP.

Datasets used for the estimation of WWTP loading come from 2006, 2007 and the first half of 2008. Total COD load is estimated at 5400 kg(COD)/day. The TSS load, which is the sum of volatile suspended solids (VSS) and inorganic suspended solids (ISS), is estimated at 3075 $kg.d^{-1}$. The TKN and NH_4 loads are estimated at respectively 516 $kg(N).d^{-1}$ and 351 $kg(N).d^{-1}$. The NH_4/TKN ratio is thus 0.68. The default value in ASMAnjou is 0.66 (ASMAnjou is the activated sludge model used in this chapter; for further information, see subsection 2.2.1).

According to a small measurement campaign performed at the WWTP for two days, the ratio between total COD and soluble COD is estimated at 0.4 and the ratio between ISS and TSS is 0.26 (typical values are 0.33). Although short, this two-day measurement campaign enabled hourly analysis of samples and daily observation of variations in these ratios, which were almost constant in the end. Longer datasets would have been preferred but soluble COD and ISS are not monitored at the WWTP.

Based on the first ratio and total COD load estimated from measurements, soluble COD load is estimated at 2160 kg(COD)/day and particulate COD load at 3240 kg(COD)/day. With the second ratio and TSS load, VSS load can be estimated at 2290 kg/day and ISS load at 785 kg/day. Values are summarized in Table 5.2.

Table 5.2: Estimation of mean incoming loads

Parameter	Load estimated
TSS	3075 kg/day
ISS	785 kg/day
VSS	2290 kg/day
Total COD	5400 kg/day
Particulate COD	3240 kg/day
Soluble COD	2160 kg/day
TKN	516 kg/day
NH_4	351 kg/day

Based on continuous measurements of influent flow rate, the mean value is estimated at 8400 $m^3.d^{-1}$ for the entire WWTP (i.e. 4200 $m^3.d^{-1}$ per treatment line).

5.1.3 Control of the aeration system

The traditional operation of the aeration system in Cambrai is based on two thresholds of oxidation-reduction potential (ORP). This is meant to be a robust measurement of the activated sludge state and was the only available aeration control measurement technology that was economical in terms of investment and operation. Over the last ten years, new ion-sensitive sensors have become available, which allow robust and reliable online measurements of ammonia and nitrate concentrations. New control laws based on these measurements are emerging, such as SNDN (see subsections 2.3.4 and 2.3.6 for more details).

Tests of simultaneous nitrification and denitrification are currently performed by other Veolia colleagues on the first line of Cambrai WWTP. The second line is still controlled with ORP. This is meant to provide proper comparison of the two systems, since the same influent is received at the same time in the two lines and the two activated bacterial populations can be adapted to each control law, thanks to the complete separation of the two treatment lines.

Before the tests of the new control law, a single air flow rate was available on the two lines. Its value was measured to 4200 $Nm^3.h^{-1}$ per line. To test the new control laws, an electric variable speed drive has been installed on the first line in order to allow air flow rate to vary between 2500 (30 Hz) and 4200 $Nm^3.h^{-1}$ (50 Hz). Air is propelled in the aeration system by means of a turbine.

This new lower air flowrate is, however, still too high to ensure continuous aeration. A degraded control law is therefore currently being tested based on thresholds of ammonia (presented in section 2.3.4) but without any control of the oxygen concentration in order to achieve almost simultaneous nitrification and denitrification with very low levels of oxygen during the aeration phases. This control law is named SABAL (Sequenced Aeration Based on Ammonia Levels) in this chapter for better clarity. The final goal of the study is the implementation of SNDN (presented in section 2.3.6) but this will require new modifications of the aeration system since the current lower bound of the air flow rate is still too high. These modifications have not yet been made due to the corresponding cost that had not been planned. Moreover, the benefits of these modifications in this case are not clearly known.

Degraded control of SABAL may be sufficient and the difference between SABAL and SNDN may not be very high. The optimization and comparison of both control laws that will be presented in this chapter can hence help us to evaluate the benefits of such modifications. Finally, the simulated benefits could then be compared with real ones in the future if these modifications are undertaken.

5.1.4 Goals of the study

The objective of this application in the case of Cambrai is first to optimize and compare the two control laws SABAL and SNDN. Moreover, as the current air production system is oversized, the goal is also to help us to define the adequate sizing of the air production system and to evaluate the benefits of this modification. This will allow us to gauge the cost of the air production system modifications against the savings on energy consumption. In the end, this can help us to decide if the modifications of the air production system are relevant.

To achieve this objective, three steps are required:

- to define the “typical” influent of the WWTP, as well as its variability,
- to calibrate the WWTP model that needs to be reliable for a wide range of functioning,
- to perform optimizations of the system for the various problems.

5.2 Calibration of an influent model

Short and long-term influent datasets representing the inlet of the secondary treatment are required for the model calibration and the control law optimization. No such datasets were available at the WWTP, mainly because no sensors are present, except concerning the influent flowrate but this is absolutely not sufficient. The phenomenological model developed by Gernaey *et al.* (2006b) is the solution chosen (see subsection 2.4.2). This model allows the generation of dynamic influent datasets based on typical daily profiles of flowrate and loads. Rainfalls are considered and a sewer model is included. A previous application of this model was already presented in Beraud *et al.* (2007) in the case of a smaller WWTP located near Toulouse, France. The same type of application is presented here.

As the modeling of the activated sludge units of the WWTP will be based on ASMAnjou (for details see sections 2.2.1 and 5.3), the influent model is modified to take into account the new fractionation corresponding to this activated sludge model. Exactly the same principles of the

original influent model are used and only the number of fractions is changed (e.g. TSS is modeled with ISS and VSS instead of particulate COD due to the inclusion of ISS in ASMAnjou). New parameters are necessarily included in this new influent model, such as VSS and ISS loads.

The influent model first is to be calibrated to demonstrate its ability to represent the incoming loading rate. To this end, a model of a storm basin must be included. Although it is not in the original influent model developed in Gernaey *et al.* (2006b), a storm basin is present in the Cambrai WWTP configuration. A schematic representation of the input and output of this model is represented in Figure 5.3.

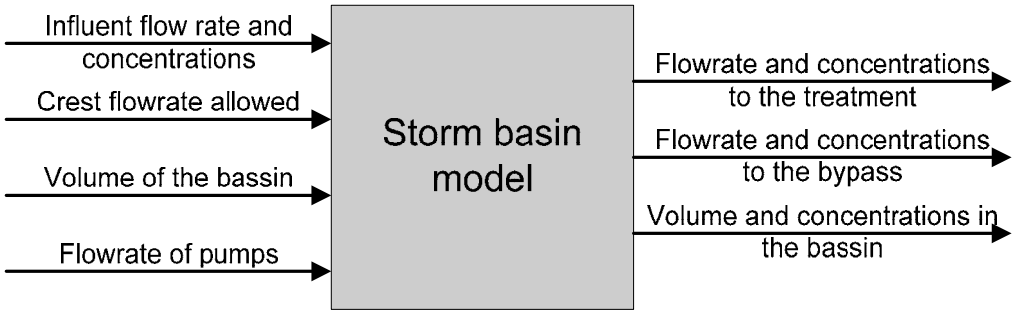

Figure 5.3: Inputs and outputs of the storm basin model

In this model, the influent flowrate is limited by the peak value allowed. When this peak value is reached, the remaining flow is directed to the storm basin. The storm basin has a limited volume. When it is full, the remaining flow that cannot go into the storm basin is directly bypassed to the outlet of the WWTP. Concentrations of pollutants in the storm basin are also computed in this model based on the concentrations in the inlet of the storm basin and the current concentrations in the basin (based on mass balances).

The last point to be defined concerns the pumping of the wastewater from the basin back to the treatment. In Cambrai, this pumping starts when the sum of the influent flow rate and the nominal flowrate of the pumps is lower than the peak flowrate allowed (due to technical aspects) and stops as soon as the basin is empty or when the influent flow rate goes back above the peak value allowed (which is rarely the case). The same behavior is included in the model.

With this additional model, the influent model can now be calibrated and validated, first, considering only the influent flow rate. For this, rain depth measured at the WWTP is used as the sole input to the model.

The measurement datasets are resulting from the months of September and October 2008 for the calibration period and from the month of May and first half of June for the validation period (during the summer period the population decreases and corresponding datasets are hence not used since they do not represent a normal functioning).

Each simulation is preceded by two days with no rain before actual rain is applied, so that the model can converge to a standard operation. The daily averages are computed based on periods of 24 hours starting at 8am.

It should be noted that during the calibration period some data acquisition problems occurred and therefore three rain events are missed. They occurred the 23rd and 24th of September (days 25 and 27 of the simulation), 3rd of October (day 35) and between the 5th and 7th of October (days 37 to 39).

In the influent model, many parameters can be tuned. The ones that were modified are summarized in Table 5.3. The number of persons equivalent results directly from the design of the Cambrai WWTP. Other parameters have been adjusted for proper calibration of the influent flowrate during dry and rain weather.

Table 5.3: Calibrated parameters of the influent model for the flowrate

Parameter	Initial value	Calibrated value
Water production per person equivalent ($l.d^{-1}$)	150	110
Number of persons equivalent		63 000
Parasite water infiltration ($m^3.d^{-1}$)	7100	2900
Conversion of rain depth in water flowrate ($m^3.mm^{-1}$)	1500	1200

The parameters of the storm basin model corresponding to the real ones of the WWTP are summarized in Table 5.4. They result directly from the design of the Cambrai WWTP.

Table 5.4: Parameters of the storm water tank

Parameter	Value
Volume of storage (m^3)	1750
Maximum flowrate to treatment ($m^3.h^{-1}$)	1350
Flowrate of the pumps ($m^3.h^{-1}$)	200

In addition to these steady-state values, the flowrate and loading daily profiles also have to be adjusted. To this end, two days of measurements of the ammonia concentration at the inlet of the secondary treatment were performed. The incoming flowrate is also continuously monitored at the WWTP. The resulting load can hence be computed. The results of these measurements are presented in Figure 5.4. No rain fell in Cambrai area during the two days of measurements or the day before. Out of these two days of experimental data, typical dry weather flowrate and load profiles have been estimated. The assumption is made that all pollutants have the same daily profile, which should provide us with sufficient accuracy for using the influent datasets that will be generated with this model.

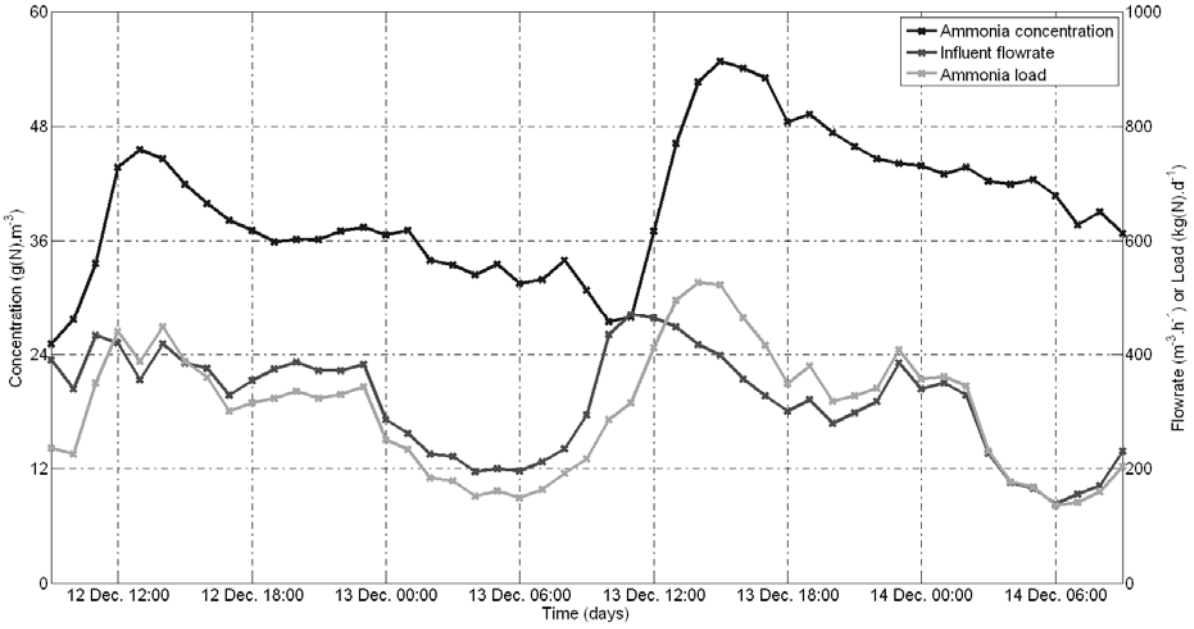

Figure 5.4: Measurements of the incoming flowrate and ammonia concentration and estimation of the ammonia load at the inlet of the secondary treatment of Cambrai WWTP.

Based on these estimated daily profiles at the inlet of the secondary treatment, i.e. the outlet of the influent model, the daily profiles at the inlet of the sewer network in the influent model are modified for a better calibration and representation of the influent. This is important since the delay between the peak flowrate and load is estimated at two hours based on the measurements performed. This choice has an impact, since the peak flowrate will flush remaining pollutants from the WWTP, while high concentrations of pollutants will come later when the flowrate is lower and thus hydraulic residence time is higher.

The result of the calibration of these daily profiles is depicted in Figure 5.5. The lines correspond to the output of the influent model (during dry weather) while the crosses correspond to the estimated profiles based on the two days of measurements. Very good calibration was achieved thanks to fine characterization of the daily profiles at the inlet of the sewer network based on 48 values per day.

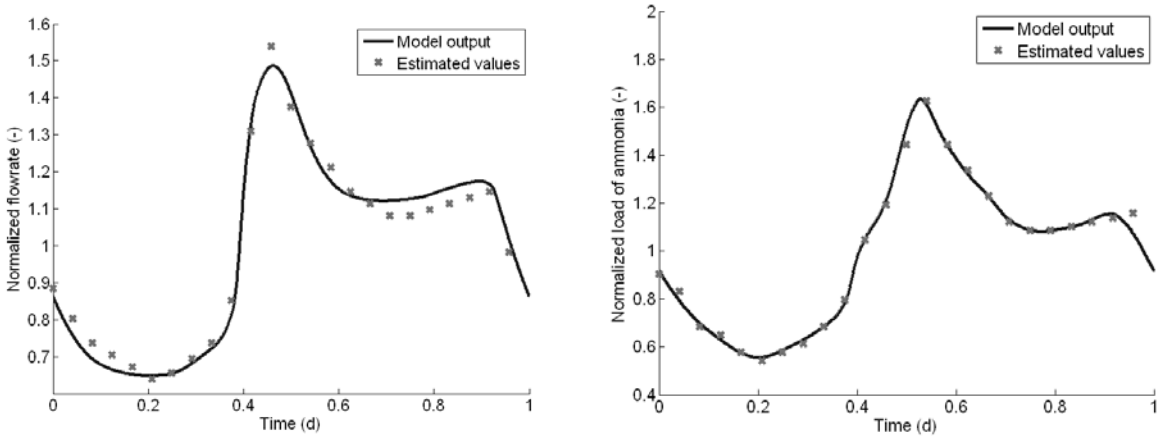

Figure 5.5: Estimation of daily profiles and output of the calibrated influent model during dry weather.

The final results of the calibration of the model (steady state values and daily profiles) are presented in Figure 5.6.

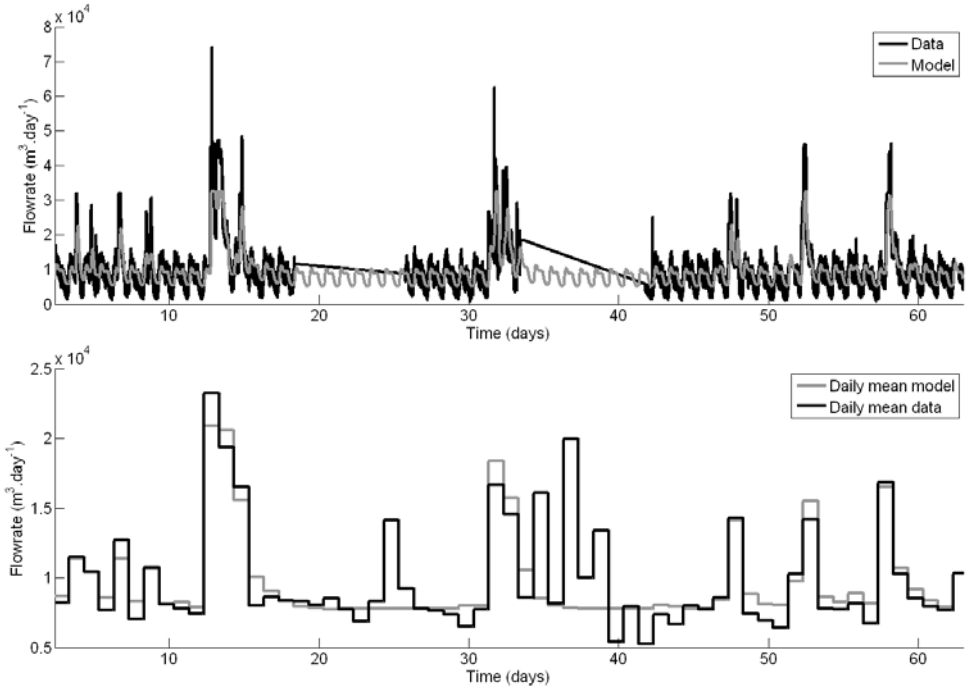

Figure 5.6: Results of the calibration of the flowrate

Except during the days where the rain data is missing, the calibration is relatively good, as can also be seen in Figure 5.7 where the relative error between the computed and measured daily averages of flowrate is depicted. If the days with missing data are not taken into account, the mean error during the calibration period is -3 %, and the standard deviation on the error is 10 %. Considering uncertainties of the measurement of rain depth due to sensor technologies as well as the fact that the measurement based on the sensor located at the WWTP can have deviations compared to actual rainfall on the entire network, the calibration can be considered as sufficient for our purpose.

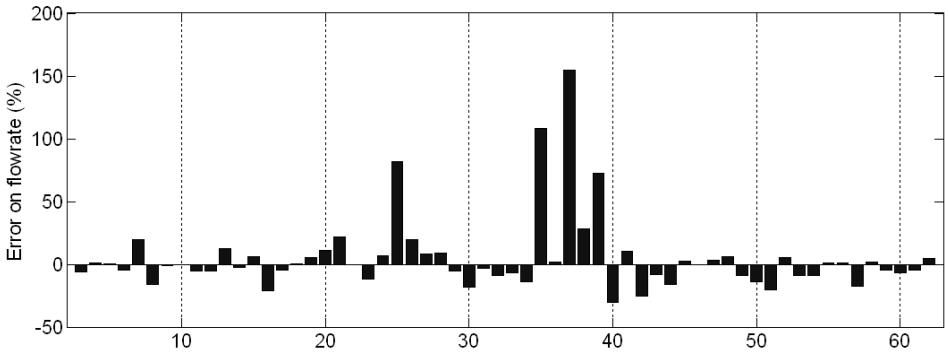

Figure 5.7: Relative error of the flowrate calibration

The results of the validation of the model calibration for the flowrate values are presented in Figure 5.8 and Figure 5.9. It should be noted that during the validation period, a bypass of the WWTP occurred due to a technical problem at the WWTP on the 6th of June (day 39 of the simulation). Without this day, the mean on the error of estimation of the incoming flowrate is 1% and the standard deviation is 15%.

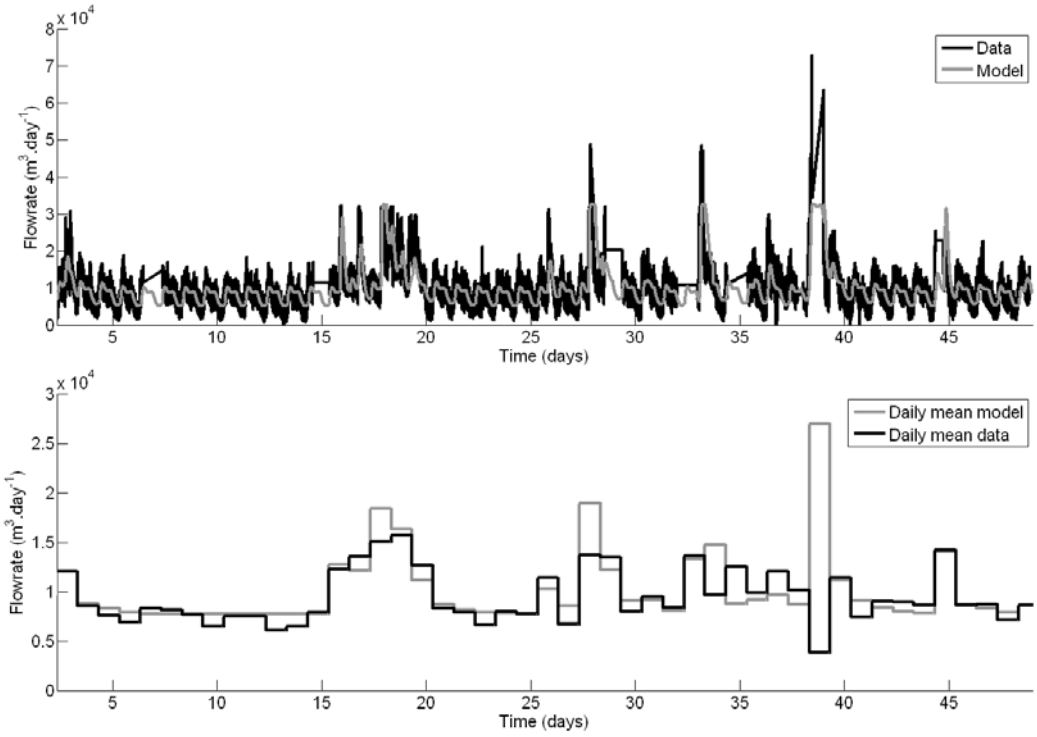

Figure 5.8: Results of the validation of the flowrate

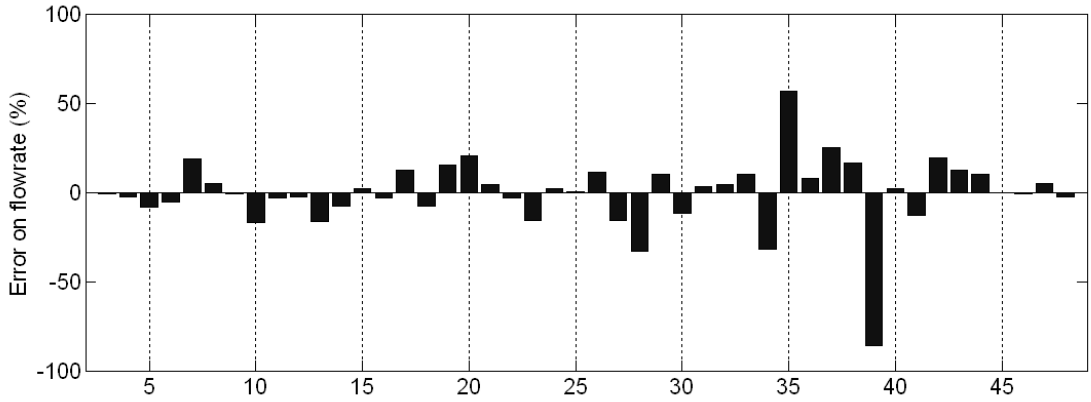

Figure 5.9: Relative error in the flowrate validation

Considering the calibration of the pollutant loads and concentrations, the only available data comes from the monitoring of the WWTP. However, this data is necessarily biased by the first errors in the calibration of the influent flowrate and some parameters had to be calibrated to minimize these errors. These parameters are summarized in Table 5.5 and Table 5.6.

In Table 5.5, the values correspond to the average incoming loads already presented in section 5.1.2 based on data from January 2006 to September 2008. Except the daily profile of variations at the inlet of the sewer previously presented, no other modification was made to

the influent model for the calibration of these loads since no additional data was available, except concerning suspended solids that will now be further calibrated.

Table 5.5: Calibrated parameters of the influent model for the pollutant loads

Parameter	Initial value	Calibrated value
Load of soluble COD (g/d/PE)	19.28	34.28
Load of particulate COD (g/d/PE)	76.72	51.43
Load of VSS (g/d/PE)[4]		36.35
Load of ISS (g/d/PE)		12.46
Load of SNH (g/d/PE)	6.89	5.57
Load of TKN (g/d/PE)	14.24	8.19

For a better representation of suspended solids loads, the parameters of the first-flush effect model have been tuned and resulting values are presented in Table 5.6. This model represents sedimentation of particulates in the network during dry weather and the flush of these particulates during storm events. The modifications of the model parameters were made of allowing more suspended solids to settle in the sewer system and to make their release happen at a lower flowrate and in a sharper manner. These modifications are made based on the dataset used for the calibration.

Table 5.6: Calibrated parameters of the influent model for the first flush effect

Parameter	Initial value	Calibrated value
Maximum quantity of SS in the sewer (kg)	1000	5000
Limit flowrate for the flush effect (m^3/d)	40000	30000
Parameter *n*	10	5
Parameter *Ff*	50	5

The results of the calibration and validation are presented in Appendix C. For all cases, the error range is high (+/- 20 % for ammonia and total N and between -50/+100% for COD and TSS). This is primarily due to the uncertainties in the flowrate (10 and 15 % of standard deviation are observed for the calibration and validation period) which impact the results on the loads. Another source of uncertainty results from the variability of the composition of the influent due to changes in activities which are not modeled with the current parameterization of the model (the model can include such changes but there is insufficient data for their correct calibration as the number of points is too low).

[4] This fraction is not included in the original model

Due to the limited number of samples available, it is not believed that better results could be achieved without additional measurement campaigns. Moreover, as our goal is to simulate and optimize the WWTP with these datasets, the modeling of specific events is not necessary. This is in fact included in the influent model with random noise and deviations which describe these events (this was of course turned off during the calibration and validation of the model). The most important characteristics that have to be represented are the average loads and the concentration profiles, which have been adequately set here.

Short-term and long-term dynamic datasets have hence been generated with this model for the optimization of the Cambrai WWTP. For short-term datasets, a dry weather dataset and a rain weather dataset have been generated in the same fashion as BSM1 datasets (i.e. the rain dataset is made of a dry weather plus a rain event occurring at days 2 and 3). For the long-term dataset, the model is simply simulated during 609 days with random rain events and random modifications of the incoming load.

5.3 Modeling of the WWTP

5.3.1 Description of the model chosen

For WWTP modeling, only a single line of the secondary treatment is considered since no sufficient measurements about the influent are available of allowing a different calibration of the two lines. The model calibration is therefore assumed to be identical for both treatment lines. The final model is presented in Figure 5.11.

The model developed by Printemps (2004), named ASMAnjou, is used for the representation of the activated sludge units. First the anaerobic zone is modeled as a single 800 m^3 unit (2). The carrousel of the aerobic zone is divided into six tanks (3) of 970 m^3 each. This value of six tanks is arbitrarily chosen based on an appropriate discretization of the effect of the aeration on only a half of the carrousel. This choice is also linked with the physical outlet of this unit which is on a weir on around a sixth of the carousel circumference. Aeration is applied to only three of the units, based on the chosen control law (4).

Due to the weir at the outlet of the carousel, a waterfall is present, which induces oxygenation of the activated sludge. Based on measurement of ammonia in the activated sludge tank and at the outlet of the secondary clarifier (see Figure 5.10), a constant reduction in ammonia of

1g.m^{-3} is estimated. A simple model for this phenomenon (5) is included in the Cambrai model, which represents the oxidation of a user-defined amount of ammonia. The corresponding growth of autotrophic biomass, production of nitrate and reduction in alkalinity is computed based on the default stoichiometry values present in the ASMAnjou models (which can also be changed by the user).

After the weir of the ASUs, the activated sludge goes into the deaerator. This unit is meant to remove all gas bubbles but it can be assumed that biomass (autotrophic and heterotrophic) is still active in this tank due to the low hydraulic residence time of the activated sludge (around 27 minutes). A standard activated sludge model is therefore used for this unit (6) with the real volume of 80 m^3.

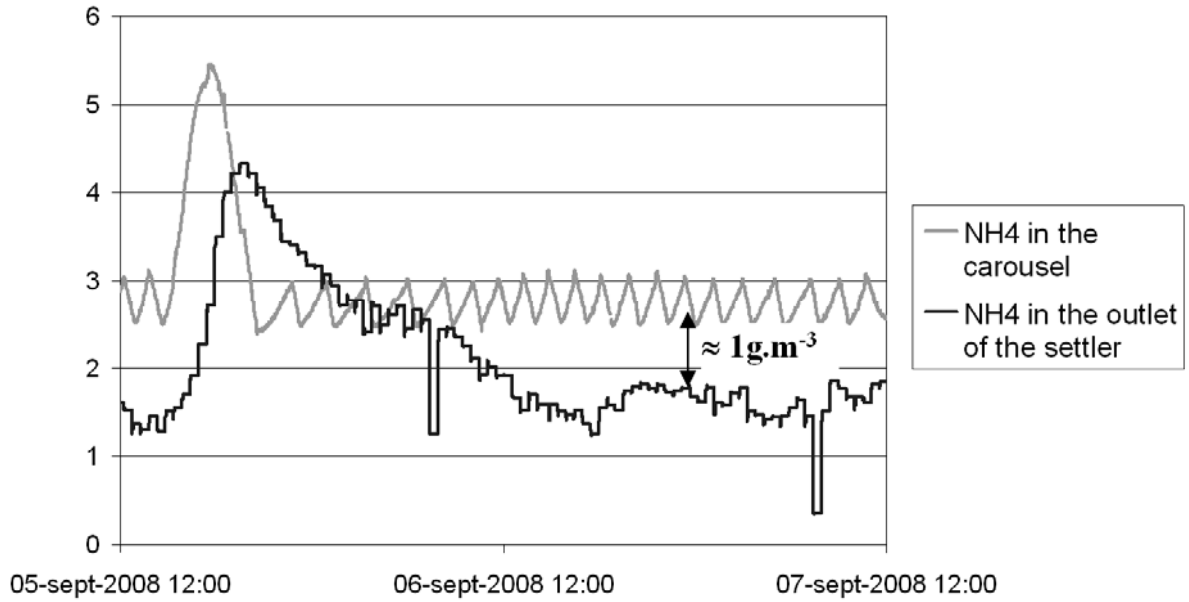

Figure 5.10: Elimination of ammonia due to the weir at the outlet of the carousel

Then comes the secondary clarifier (7). As a consequence of the choice of the ASMAnjou model, a modified version of Takács's model is used for the secondary clarifier representation, including the propagation of soluble compounds in the layers (their concentrations are not supposed constant in the entire settler but only in each of its layers).

The cleaned water goes from the overflow of the settler to the outlet of the WWTP (8).

A fraction of the sludge is extracted from the underflow of the settler to the sludge treatment (9). This extraction is based on one extraction per day during open weekdays. No extraction is

performed during the weekend. The extraction is made with a pump with a nominal flow rate of 32 $m^3.h^{-1}$. The average volume extraction during the first half of 2008 was 160 $m^3.d^{-1}$. Considering that the sludge is extracted only during open week days, the volume to extract is 227 $m^3.d^{-1}$. The pump has therefore to be switched off 7 hours per day. This is the role of the extraction control (11).

The remaining sludge is recirculated to the activated sludge units. The flowrate is controlled (10) to be equal to the influent flow rate, in order to provide adequate hydraulic velocities in the secondary clarifier and to avoid problems of sludge rising during rain and storm events. Activated sludge is mixed with the influent (14) before being fed to the activated sludge unit.

As the heterotrophic biomass is still active in the secondary clarifier, a model is included to remove all remaining oxygen and a fraction of nitrogen in the recycling flux (12). This model is based on the proposition of Gernaey *et al.* (2006a) who assessed that such a simple model is sufficient to provide accuracy without the huge computing time associated with a fully reactive secondary clarifier (in a reactive clarifier model, an activated sludge model is associated with each layer).

The final missing piece of this model is the inclusion of the treatment of phosphorus. This was not the goal of the study. However, the amount of sludge and biomass produced by this process still has to be considered for the calibration of the model. Only the mass of inorganic products in the biomass and the inorganic products of precipitation are considered to change. Based on the measurement of remaining phosphorus in the outlet of the WWTP which is always below 1 $g.m^{-3}$ (compared to 10 $g.m^{-3}$ on average at the inlet), a steady-state value is assumed to be adequate. It concerns a production of ISS which is estimated at 150 $kg.d^{-1}$ based on internal Veolia knowledge of typical phosphorus treatment[5] (both the biological treatment and the addition of Iron(III) chloride are considered in the computation of this value).

[5] This computation cannot be detailed due to the protection of Veolia's intellectual property and expert knowledge. It is based on steady-state observations of the physico-chemical treatment efficiency.

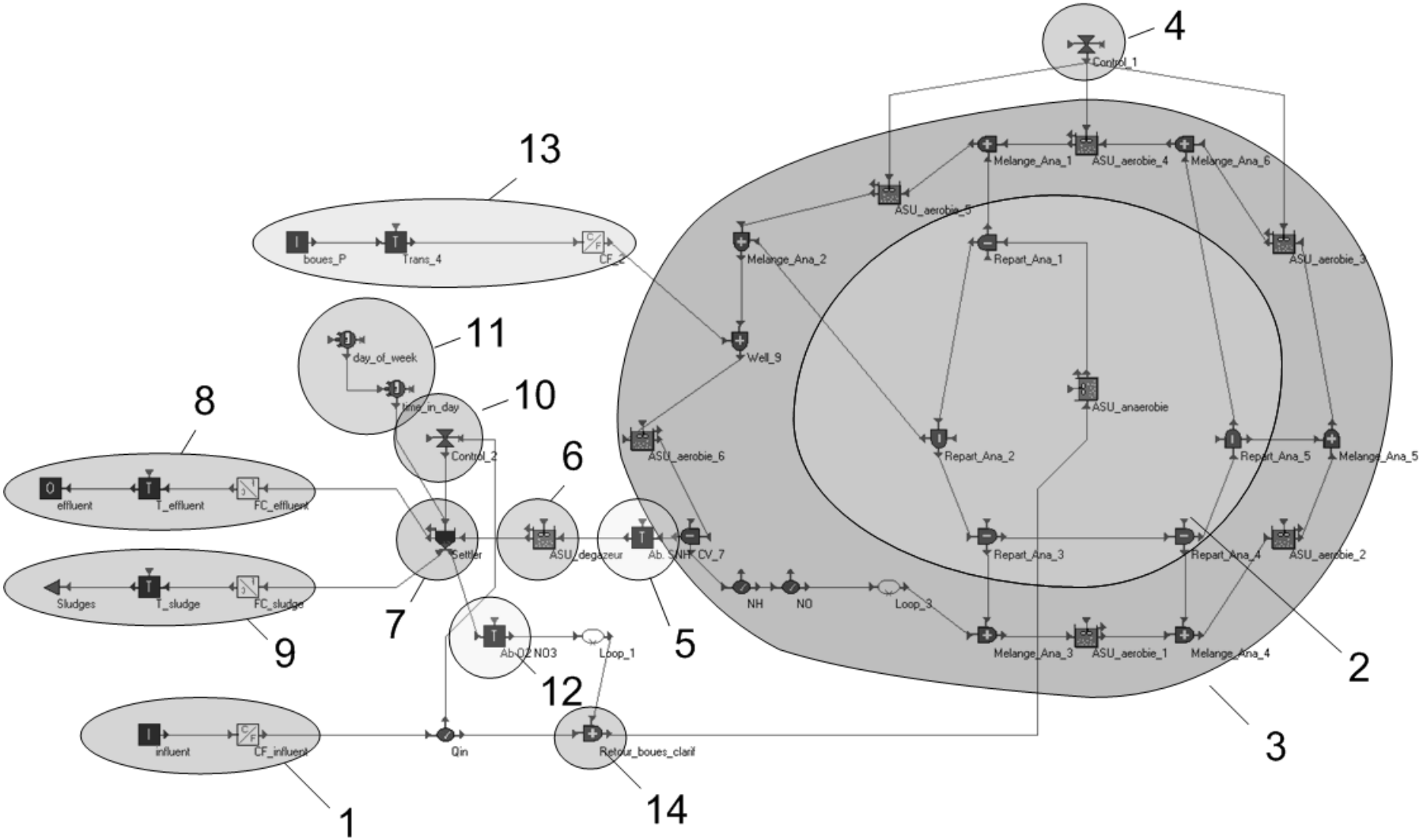

Figure 5.11: Model of one line of the Cambrai secondary treatment

5.3.2 Results of the model calibration

The calibration of the model is based on flowrate and pollutants profiles generated using the influent model previously described. Simulations of a dry weather week are repeated until simulations convergence is obtained (as explained in section 4.2). The goal of this calibration is to represent average WWTP functioning in order to be able to compare various control laws as well as the impact of their settings.

Defaults values are used for all models and only a few are modified when required during calibration. In our case, we chose to modify the yield of the heterotrophic biomass Y_H and the fractionation of the incoming pollution (*FsolMESmin*, fraction of ISS that can be transformed in soluble compounds and *Rpart*, repartition of the particulate COD between inert and biodegradable fractions). The wastage flowrate Q_w is also adjusted.

The calibration is done with the sequenced aeration that is currently tested in the WWTP based on ammonia levels for the start and stop of the aeration. The lower limit of ammonia that stops the aeration is set to 2.5 $g.m^{-3}$ and the upper limit of ammonia that starts the aeration is set to 3 $g.m^{-3}$. The air flowrate during the aeration phases is set to 2500 $Nm^3.h^{-1}$.

As little data is available, the goal is to achieve proper representation of average composition of the mixed liquor and effluent based on experimental data. These goals are the total mass of suspended solids in the activated sludge (4.4 $g.L^{-1}$), the fraction of ISS above TSS in the activated sludge (0.40) and the concentrations of ammonia and nitrogen in the effluent (respectively 2 and 1 $g(N).m^{-3}$).

The final calibrated parameters values are summarized in Table 5.7, together with their default values and typical ranges (when available, from Printemps, 2004). In this table, a fifth parameter α is also presented. This is the coefficient representing aeration system efficiency, constant ratio between air flow rate and oxygen transfer coefficient. This parameter was not calibrated based on the goal presented in this subsection but it is identified later on and is detailed in sub-section 5.3.3 below. It is represented here for reference purposes but did not impact the calibration performed in this sub-section.

In this table, Y_H is increased while *FsolMESmin* and Q_w are decreased. This allows an increase in the amount of sludge in the system. The combined modifications of these three parameters are necessary to limit their variations and be in accordance with acceptable ranges typically found in WWTP models. *Rpart* is also decreased in order to adjust the ISS ratio above TSS.

Table 5.7 : Calibrated parameters of the secondary treatment model

Parameter	Default value and typical range	Calibrated value
Y_H	0.6 (0.57 – 0.67)	0.67
FsolMESmin	0.75	0.5
Rpart	1.2	0.7
Q_w	770	700
α	0.5	0.65

Table 5.8 presents the results in terms of operational values. The difference between estimated values based on measurements and values obtained after the calibration is adequate. The results of this simple calibration are sufficient for proper representation of the WWTP functioning aimed at the comparison of control laws functioning. Refinements of this calibration can be the subject of a more detailed study but the current model seems to be sufficient for our use, especially taking into account the uncertainties encountered in the values of the influent concentrations.

Table 5.8: Calibrated operational values of the secondary treatment model

Parameter and unit	Measurement	Value after calibration
TSS in the mixed liquor ($g.m^{-3}$)	4400	4406
ISS / TSS in the mixed liquor (-)	0.40	0.39
S_{NH} in the effluent ($g(N).m^{-3}$)	2	1.82
S_{NO} in the effluent ($g(N).m^{-3}$)	1	1.38

5.3.3 Reference point for the ORP control law

One of the goals of this study is to compare the new control laws based on measurement of ammonia and nitrate with the current ORP control law. The simulation of such a control law is quite difficult as it requires a model of the ORP measurement.

Many attempts to model this signal have been made. Most works (Zipper, 1998; Brown *et al.*, 1999; Fuerhacker *et al.*, 2000; Cecil, 2008a) focus only on the detection of the curve bend corresponding to the end of the nitrification and denitrification. This is pertinent for use in sequenced batch reactors where complete cycles are desired. This is absolutely not our case.

Only a few recent publications (Meijer, 2004; Cecil and Skou, 2008b) attempt to expand knowledge of the modeling of this signal.

A small attempt is made here to model this signal based on the fact that it is related to the logarithm (base 10) of the concentration of main chemical species. In ASMAnjou, the only ones available are oxygen, ammonia and nitrate. A model based on these species based on the Nernst equation can be written as:

$$ORP = A + B \cdot \log_{10}(S_O) + C \cdot \log_{10}(S_{NH}) + D \cdot \log_{10}(S_{NO}) \quad \text{(V.1)}$$

Finding the right values for parameters A, B, C and D is very difficult. This is mainly the consequence of the imprecision of the measurements. This is particularly the case for ammonia and nitrate, for which an error of 0.5 $g.m^{-3}$ is usual. When using a logarithm of such values usually located between 0 and 3 $g.m^{-3}$, large errors can be induced.

However, a set of parameters presented in Table 5.9 has been calibrated on a period of 15 days based on measurements performed on the Cambrai line where SABAL control is tested. The results of this model on a validation period of 15 days are presented in Figure 5.12. The model is able to adequately represent ORP measurements on this period. It should however be noted that a calibration of the model on the second Cambrai line, which is controlled by ORP thresholds, was not achieved. This is assumed to be the consequence of greater variations of ammonia and nitrate concentrations which implies more uncertainties in these measurements as well as on its logarithmic value.

Table 5.9: Proposed parameters for the ORP measurement model

Parameter	Value
A	205
B	41
C	-285
D	133

The calibration of the ORP model achieved should however be used with caution since the range of ammonia explored in the dataset is very small (between 2.5 and 3 $g.m^{-3}$). The extension to wider ranges of ammonia concentrations may be at risk. This is however the best solution found so far and currently available to our knowledge. The advantage of this solution is that it allows an estimation of the ORP control performance based on a WWTP model.

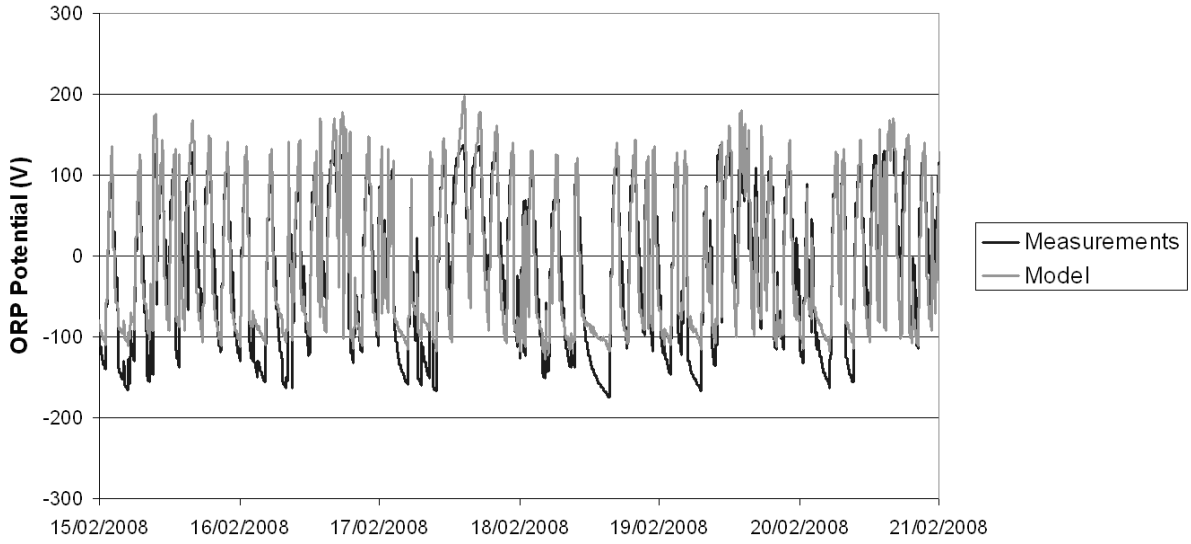

Figure 5.12: Validation of the ORP model on a 15-day dataset of Cambrai WWTP

This model was thus tested for the simulation of the ORP control law based on thresholds of -50 mV and +120 mV that are currently applied at the WWTP. The first simulation results did not provide satisfactory results since the ammonia concentration was steadily increasing before stabilizing at a concentration around 15 $g.m^{-3}$, which is absolutely not in accordance with reality.

In order to achieve a better representation of the real performance, the real air flow rate applied during the first week of September was applied to the model, with corresponding estimations of the influent concentrations and flow rates. This step showed that default aeration efficiency (α, constant ratio between air flow rate and oxygen transfer coefficient) was not well-calibrated. The default value is 0.5 while the adequate value for proper representation of the ORP functioning is estimated at be 0.65, based on real sequences of aeration. A second simulation was performed based on this new parameter with the dataset of the first week of September. The results are better since the total air volume injected per day is simulated at 25014 $Nm^3.d^{-1}$ instead of 24006 $Nm^3.d^{-1}$ in reality. The ammonia concentration in the ASU is 1.57 (instead of 0.61) and the nitrate concentration in the ASU is 0.37 (instead of 2.27). The correspondence is hence not very good for these parameters but it should be noted that there is much uncertainty about the real incoming load of the period. In order to further compare this simulation with the real functioning of the ORP control law, cumulative frequencies of the aeration phase length are presented in Figure 5.13. This figure shows that

there is quite a significant difference between the simulation and the real functioning, and more precisely that long aeration phases are not well-represented.

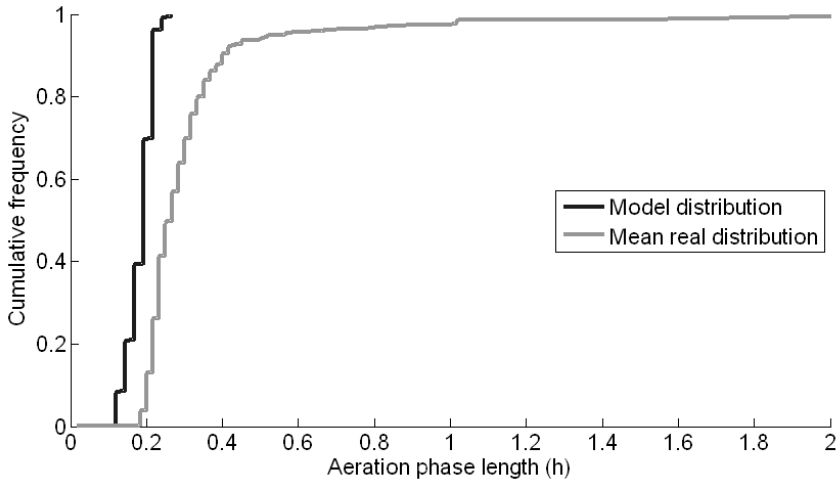

Figure 5.13: Cumulative frequencies of aeration phase length for the real measurements (light grey) and best simulation (dark grey) based on the first week of September.

The tests performed in this section thus clearly indicate that the precision of ORP control simulations based on the proposed model is limited and can still be enhanced. A better calibration or formulation of the ORP model is obviously required but this could be the subject of an entire PhD.

The developed model is however still sufficient to achieve preliminary estimations of ORP performance on the rain weather dataset, based on the simulation procedure proposed in this thesis. The performance in terms of mean volume of air injected per day and mean concentrations of ammonia and total nitrogen in the effluent are presented in Table 5.10, together with corresponding performance based on real values. The two sets of real performance is based on the months of June and September 2008. The average mixed liquor temperature was 18.6°C in June and 18.5°C in September. The simulated performance is based on the rain weather input dataset with two days of rainfall. The deviations between simulated and real performance is quite high but the difference between the two real periods is even greater. This is in accordance with the deviations previously observed in the BSM1 case with the long-term evaluations. This clearly indicates that many factors influence real performance and the comparison of control law performance on real processes is hence at high risk. The only practical solution is to make two simultaneous comparisons with two treatment lines as at Cambrai but very few WWTPs have many treatment lines with a full

separation of the two lines. This clearly indicates the benefits of the methodology proposed in this thesis, which allows an unbiased comparison.

Table 5.10: Comparison of mean performance obtained with the real and simulated ORP control laws

Parameters	Real ORP control June 08	Real ORP control Sept. 08	Simulated ORP control
Daily volume of air injected ($Nm^3.d^{-1}$)	32614	24006	29708
S_{NH} concentration in the ASU ($g(N).m^{-3}$)	0.60	0.61	2.21
S_{NO} concentration in the ASU ($g(N).m^{-3}$)	0.23	2.27	0.69
S_{NH} concentration in the effluent ($g(N).m^{-3}$)			1.18
S_{NO} concentration in the effluent ($g(N).m^{-3}$)			3.46

5.3.4 Reference point for the SABAL control law

Since the reliability of the ORP simulations is limited, a reference point of SABAL functioning is required for more detailed analysis of the optimization results. For this, two reference simulations are performed. Both are based on ammonia levels of 2.5 $g.m^{-3}$ and 3 $g.m^{-3}$ but two air flow rates are tested: 2500 $Nm^3.h^{-1}$ and 4200 $Nm^3.h^{-1}$. These values correspond to two values of the aeration system that have been tested and the practical performance can hence be compared with the simulation ones. For both simulations, the aeration efficiency α is set to 0.65 since it was demonstrated to be the best parameter in the search of an ORP reference previously presented.

The resulting performance is presented in Table 5.11 for the air flowrate of 4200 $Nm^3.h^{-1}$ and Table 5.12 for the air flowrate of 2500 $Nm^3.h^{-1}$, together with corresponding performance based on real values for reference purposes. This real performance is based on the month of June 2008 for the first flowrate and September 2008 for the second one.

Table 5.11: Comparison of mean performance obtained with the real and simulated SABAL control law with an air flowrate of 4200 $Nm^3.h^{-1}$ during aeration phases

Parameters	Real SABAL control (June. 08)	Simulated SABAL control
Daily volume of air injected ($Nm^3.d^{-1}$)	36666	29228
S_{NH} concentration in the ASU ($g(N).m^{-3}$)	2.85	2.66
S_{NO} concentration in the ASU ($g(N).m^{-3}$)	1.14	0.70
S_{NH} concentration in the effluent ($g(N).m^{-3}$)		1.68
Total N concentration in the effluent ($g(N).m^{-3}$)		5.31

Table 5.12: Comparison of mean performance obtained with the real and simulated SABAL control law with an air flowrate of 2500 $Nm^3.h^{-1}$ during aeration phases

Parameters	Real SABAL control (Sept. 08)	Simulated SABAL control
Daily volume of air injected ($Nm^3.d^{-1}$)	21313	26210
S_{NH} concentration in the ASU ($g(N).m^{-3}$)	3.05	2.72
S_{NO} concentration in the ASU ($g(N).m^{-3}$)	0.18	0.77
S_{NH} concentration in the effluent ($g(N).m^{-3}$)		1.78
Total N concentration in the effluent ($g(N).m^{-3}$)		5.42

In order to analyze these results, they are depicted in Figure 5.14, together with the ORP performance obtained in the previous paragraph. The observation on this figure is that there is a significant deviation between simulated performance and real performance. However, it can be clearly identified that some performance corresponds to the month of June and some to the month of September.

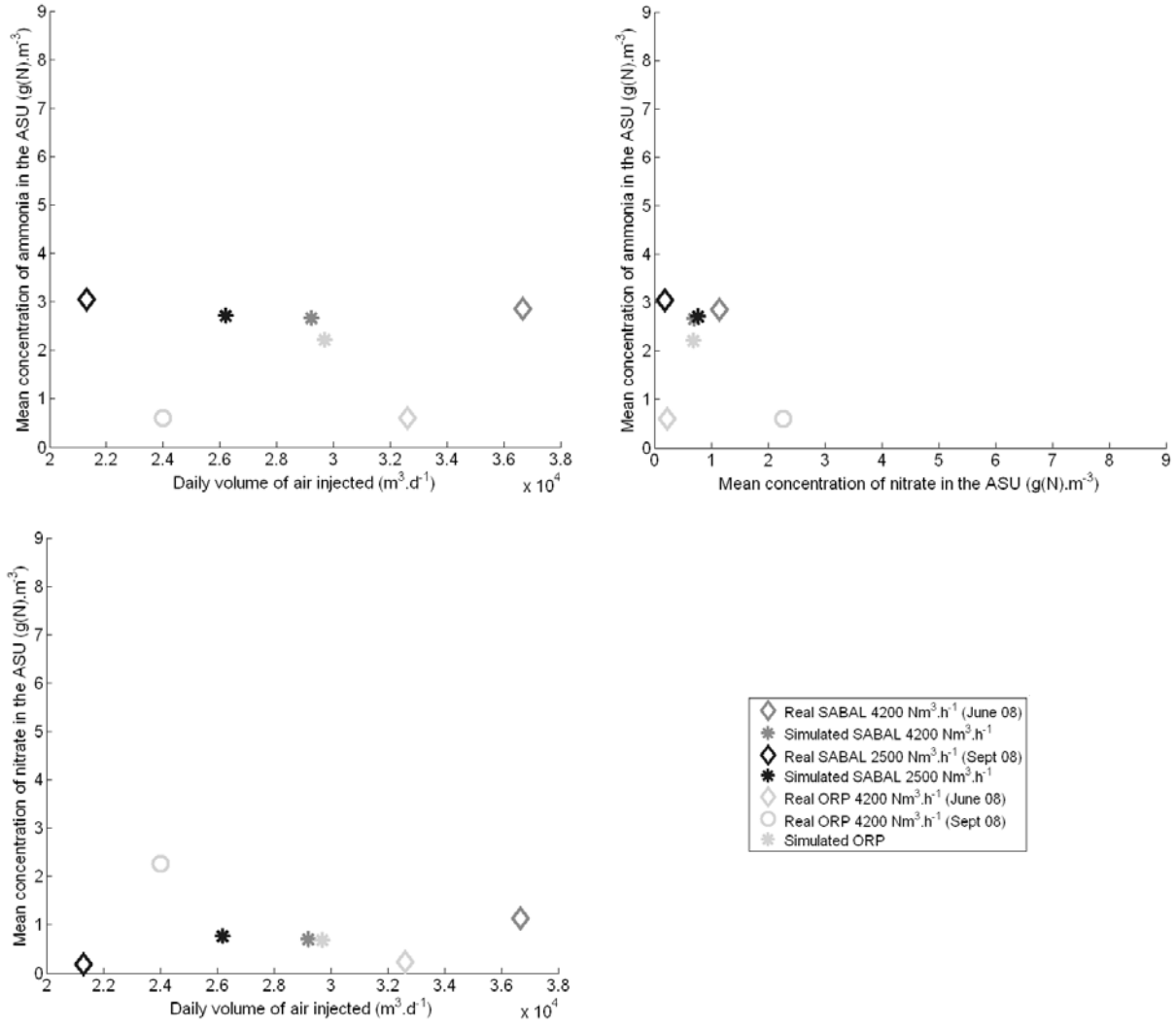

Figure 5.14: Comparison of simulated and real performance of ORP and SABAL control laws.

Considering ammonia and nitrate concentrations, the difference between simulated and real performance is much smaller, except for ORP simulations, for which the values of both concentrations are inverted.

These deviations between real and simulated performance is hence clearly due to deviations of the loading or composition of the influent between both periods which induces these deviations. In order to have a better fit, continuous daily measurements of the influent characteristics would have been required but this would have required measurement campaigns that were neither feasible nor affordable. These reference points are however adequate to provide tendencies and even if the order of magnitude is not correct, it will allow a better comparison of simulated performance obtained during the optimization that follows.

5.4 Optimization of the aeration control laws

Two control laws have been optimized in the case of Cambrai: SABAL with an air flowrate of 2500 $Nm^3.h^{-1}$ and SNDN with a freely evolving air flowrate. The first optimization will hence correspond to the maximum achievable performance with current air production system while the second one will give an insight into the performance that can be attained if this system is modified.

For these optimizations, the same procedure is used as in BSM1. For each solution, stabilization simulations are performed based on the dry weather influent dataset previously generated. A last week of simulation is then performed to evaluate the performance based on the rain weather influent dataset. It should be noted that these optimizations have been performed with aeration efficiency α to its default value, i.e. 0.5, since its calibrated value had not been found at the time the optimizations were launched. Since it is simply an efficiency ratio, it will not overly impact the Pareto front or the optimal solutions but mainly only the volume of air daily injected. Optimal solutions found will be simulated again with the calibrated value (i.e. 0.65) for their comparison with real performance obtained at the WWTP with the SABAL control law.

5.4.1 Optimal short-term performance

The optimized short-term performance obtained for the two control laws is presented in Figure 5.15. The difference between the two control laws is relatively small. In terms of

ammonia and total nitrogen concentrations in the effluent, the two control laws reach almost the same performance. The only difference considering these two objectives is that very low concentrations of total nitrogen in the effluent (below 5.6 g(N).m-3) can be reached only with SABAL but its own minimum is 5.2 g(N).m^{-3}. Considering the volumes of air daily injected, the difference between both control laws is located between 3 and 5 % of reduction for SNDN compared to SABAL for the same ammonia and total nitrogen concentrations in the effluent. This comparison is valid only for ammonia concentrations above 1 g(N).m^{-3}. Bigger reductions are achieved below this point but such solutions should probably be disregarded since their practical implementation would not be possible due to the uncertainty in the measurement of ammonia in such low levels. In future optimizations, a constraint should be added considering this limitation in order to avoid such solutions.

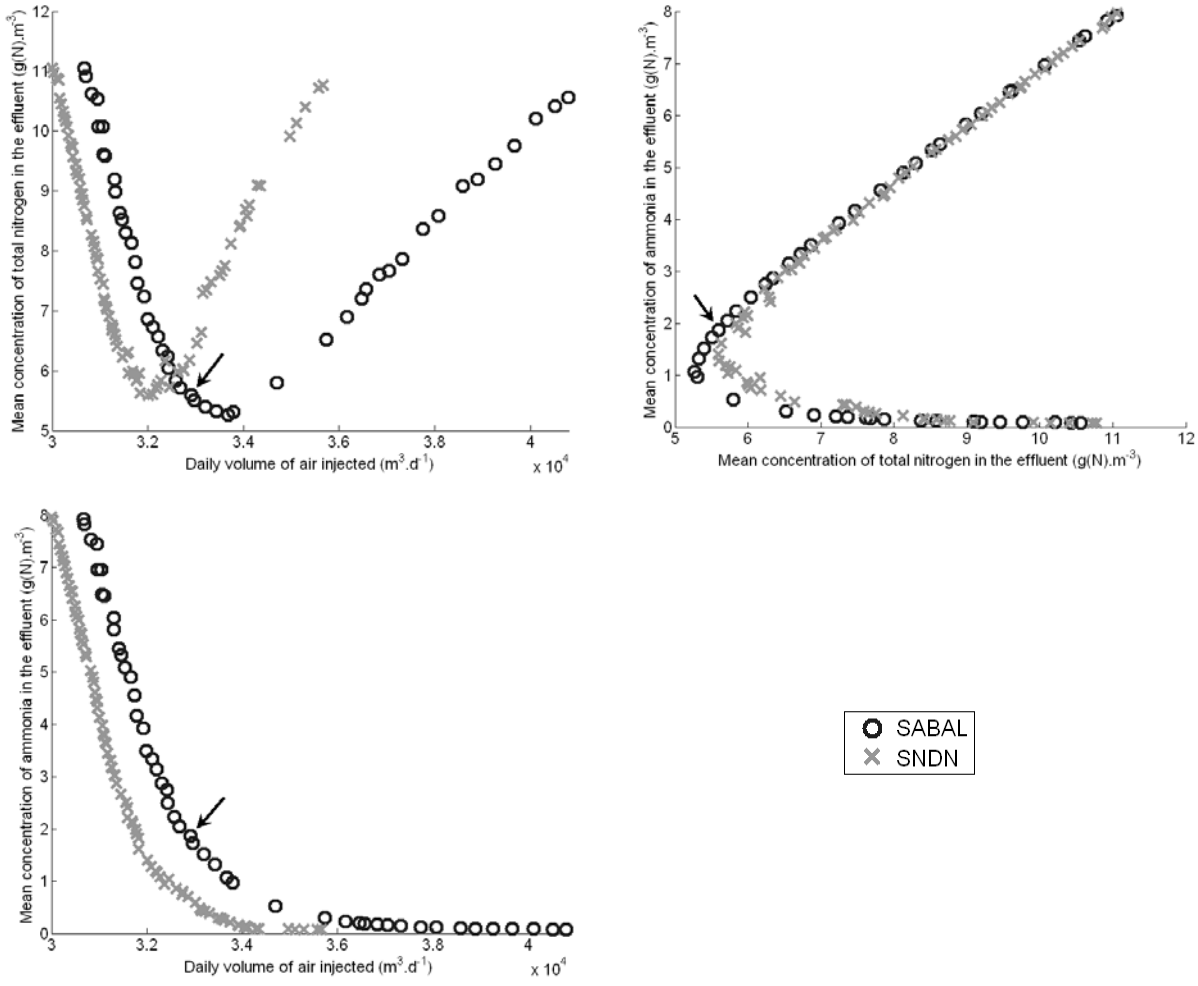

Figure 5.15: Comparison of short-term performance of SABAL and SNDN at Cambrai WWTP

On the figure, a point is indicated by an arrow. This point is an optimized setting of SABAL with ammonia levels of 2.56 and 3.06 g(N).m^{-3} to control the aeration. This point is hence very close to the real settings that are currently tested at the WWTP and has been found by the optimization procedure as an optimal point.

5.4.2 Comparison of optimized and real performance

In order to further compare the performance obtained with the two control laws with the real ones, the final optimized solutions are simulated again with the correct aeration efficiency α of 0.65. Since the total nitrogen effluent concentration is not available for the real measurements, the comparison is based on ammonia and nitrate mixed liquor mean concentrations instead of ammonia and total nitrogen effluent mean concentrations. The optimized solutions results as well as the real performance obtained at the WWTP are depicted in Figure 5.16.

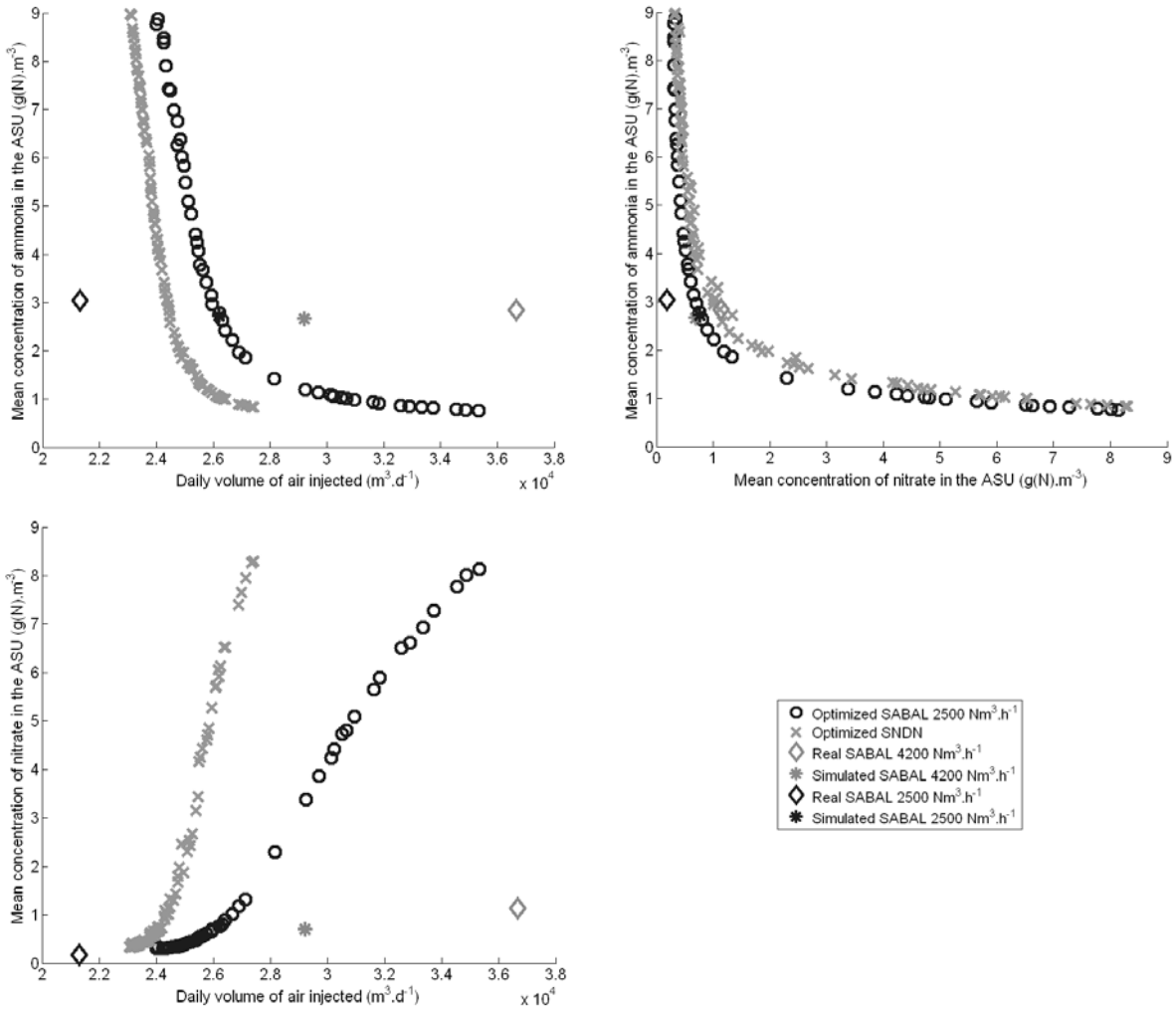

Figure 5.16: Comparison of simulated and real performance of the control laws at Cambrai WWTP

The first remark concerns the fact that the deviation between simulated and real mean air volume injected is much higher than the same deviation between SABAL and SNDN optimized performance. However, the second remark is that this deviation of real performance has been assessed in the previous subsection as the consequence of loading modifications more than the air flow rate change switch. The reduction in daily volume of air injected assessed in the previous subsection is hence probably accurate. Considering the ammonia and nitrate concentrations in the ASU, the same kind of operating conditions can be reached by both control schemes and the difference between simulated and real values is very small, as can be seen in the upper right portion of the figure.

5.4.3 Settings obtained for the optimized control laws

As for the optimization based on the BSM1, one of the significant interests of the optimization methodology is that it also provides insight into the settings of the control laws and their evolution. Each solution setting resulting from the optimization of SABAL and SNDN are presented in Figure 5.17 and Figure 5.18. The x-axis of these graphs is the mean daily volume of air injected for each solution. This allows a better visualization than the use of the solution number. The spacing between solutions is related to their objective values instead of being constant.

Concerning SABAL, the main observation is that the difference between the two ammonia levels is constant at 0.5 $g(N).m^{-3}$. The optimization was hence probably very easy since only two parameters are considered and one of them can be constant. This observation is in accordance with the practical experiments which say that the aeration phases should be as short as possible. The only limitation of this reduction is the number of cycles per day which has to be limited in order to prevent rapid deterioration of the air production system. In the optimized SABAL solutions, between 11 and 30 cycles per day occur (a greater number of cycles is achieved when the ammonia concentration in the effluent is very low). Solutions with more than 24 cycles per day on average should be avoided since they induce a little more than one cycle per hour on average, which means two cycles per hour at the peak of the incoming load. This is currently not acceptable by most air production systems. Such a constraint should be added in future optimization and will induce more variations of the difference between the two ammonia levels used in the control law.

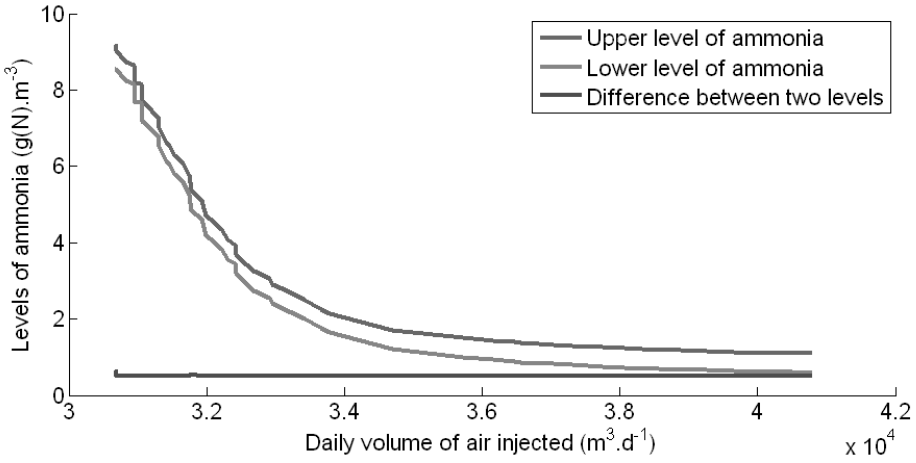

Figure 5.17: SABAL settings resulting from the optimization

Concerning SNDN, the same types of curves are obtained as in BSM1. The bounds of ammonia decrease when the set point of nitrate increases. The values are, however, a little more jittered in this application. This is apparently due to poor setting of the solver tolerances which avoided the convergence of the solutions to the true Pareto front. Another remark on these results is that for a low level of ammonia, the difference between the upper and lower bound of the ammonia control is probably too small since it would be very difficult to detect such small variations with usual sensors and it may induce instability of the controller. A constraint must be added to the optimization problem concerning this point.

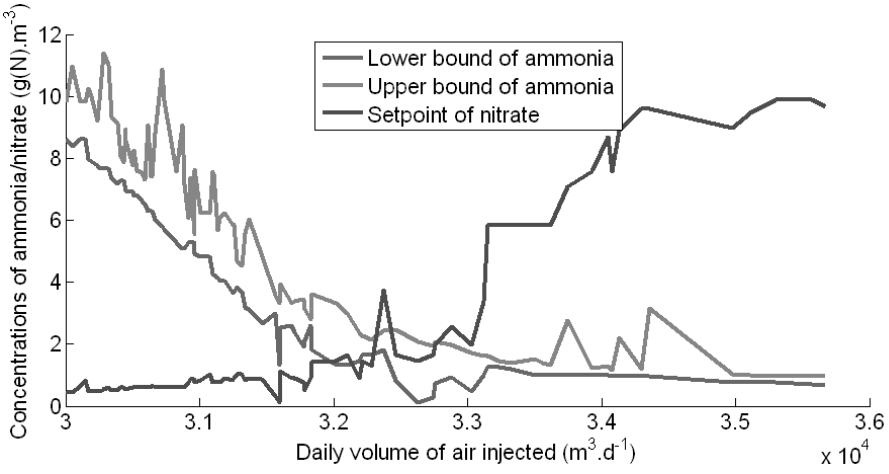

Figure 5.18: SNDN settings resulting from the optimization

Finally, to conclude on this application, it can be interesting to look at the optimized solution of SNDN from the perspective of the choice of an adequate air production system. The 1st, 5th 50th, 95th and 99th percentiles of the instantaneous air flowrate are plotted in Figure 5.19 for

each optimized solution. For most solutions, the air flowrate oscillates between 17000 $Nm^3.d^{-1}$ (710 $Nm^3.h^{-1}$) and 35000 $Nm^3.d^{-1}$ (1500 $Nm^3.h^{-1}$). These values are much lower than the range currently achievable by the air production system (2500-4250 $Nm^3.h^{-1}$) and this clearly indicates that the air production system would need to be modified if SNDN needed to be tested. The over-sizing of the current air production system is, however, normal since it is made to be capable of handling the increase of load that is expected in the next 10 years. Moreover, considering the deviation between the real and simulated mean volumes of air injected, the accuracy of the simulated values may be questioned.

Another remark about the figure is that the range of air flowrate variations increases significantly for low mean daily volume of air injected (i.e. for significant concentrations of ammonia in the effluent). This is in fact due to the rain event which disturbs the controller since the high bounds of ammonia cannot be respected anymore during this period and the aeration is almost stopped during this period. Then, when the rain stops, the control suddenly increases the air flowrate as the upper bound of ammonia is quickly reached. Such solutions should maybe be avoided in the future and must at least be further analyzed in terms of long-term performance.

To conclude, a clear perspective of the methodology perceived in the analysis of this figure is the optimization of air production system sizing, with objectives of installation, maintenance, operation costs and treatment performance. The GA can for instance choose between different sizes of standard air production systems. Their performance can then be computed based on datasets acquired in cooperation with manufacturers of air production systems.

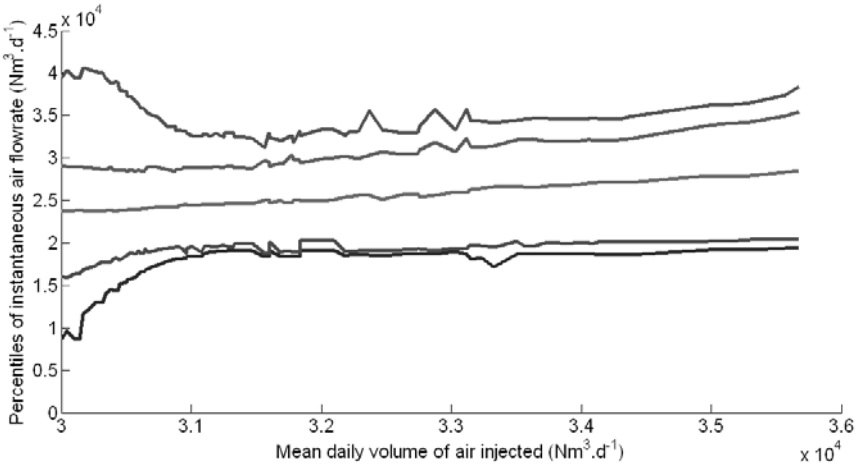

Figure 5.19: 1st, 5th, 50th, 95th and 99th percentiles of instantaneous air flow rate for each optimized solution.

5.5 Conclusion

This chapter has presented the application to the real case study of Cambrai WWTP.

Preparatory work concerning the modeling of the influent and WWTP and the computation of reference performance was first presented. This first step is very important since it will influence all future results and might cause problems during the optimization. The influent model is required for the generation of dry, rain and long-term influent datasets. The WWTP model is required for the simulations and evaluation of the problem objectives. The reference points are required for comparison with optimization results. They have to be based on current functioning of the WWTP but must be calculated based on simulations in order to avoid risks of deviation. In our application, this was the most difficult step since the simulation of ORP control laws is not yet a current practice and additional development is needed in this research area. Huge deviations were observed between the various real performance levels but they are in fact mostly the consequence of influent loading modifications. The impact of phosphorus removal is also maybe underestimated and could be enhanced with a model including this phenomena. The calibration of the model could also be enhanced with a more detailed measurement campaign and/or a study of the sensitivity of the calibration to each model parameters.

The work then focused on the comparison of two control laws aimed at simultaneous nitrification and denitrification. The first one, SABAL, is a degraded mode of SNDN based on sequenced aeration. This control law is required due to the current sizing of the air production system which is too much for SNDN. The other optimized control law is SNDN with continuous air production. Results obtained are indicating a reduction in 3 to 5 % of the daily volume of air injected with SNDN. The impact in terms of energy consumption however still has to be assessed since the modification of the air production system will probably induce greater savings. Long-term evaluations of optimized solutions also still need to be conducted for an unbiased comparison of the two control laws and better quantification of their performance. Finally, this real application opens a wide range of opportunities for more complex optimization problems such as the choice of an adequate air production system.

Chapter 6 - Conclusion - contributions and perspectives for future development

Chapitre 6 - Conclusion finale sur les contributions et perspectives de développements futurs

This last chapter presents a summary of the key findings of this thesis as well as the requirements, limitations and outcomes of the methodology proposed. Three perspectives of developments have been partially tested and are also presented in this chapter. Finally, thoughts for future research are summarized as a conclusion of this thesis.

6.1 Conclusion

6.1.1 Summary of key findings

This thesis has presented a new methodology for the optimization and comparison of WWTP control laws based on the combination of traditional WWTP models with the multiobjective genetic algorithm NSGA-II. The use of WWTP model simulations allows unbiased evaluation and comparison of control laws, which is very difficult to achieve on real processes due to the ever-changing characteristics of the influent and external operating conditions (e.g. population or industrial changes, temperature variations, etc.). The use of a multiobjective approach allows clear visualization of the trade-offs among multiple objectives and can hence provide more information before the choice of a specific control law and its settings.

The methodology developed first includes a new approach for the evaluation of WWTPs and control law performance based on simulations. This approach is intended to stabilize the simulations before starting an evaluation which is thus representative of the stabilized functioning. This stabilization is based on the repetition of one week of data representing dry weather. The convergence of the simulation to a stabilized functioning is checked after each simulated week and the repetition ends when the stabilization criterion has been reached. A final week of simulation is then performed to evaluate the performance, based on an influent dataset containing a rain event of two days. This influent dataset hence represents a good range of perturbations at the cost of only one week of simulation. A refinement of this simulation procedure consists in storing the stabilized states values together with the parameters used for each evaluation. They are then used to initialize the following simulation, based on the search of the nearest previous one in parameter space, with the assumption that close simulations in the parameter space represent almost the same operating points. This refinement reduces the computing time required for simulation convergence. This new simulation procedure therefore ensures that the system has converged to a stabilized

functioning before performing the evaluations and these evaluations are performed with a lower total computing demand than the procedure originally proposed in BSM1.

The methodology then highlights the importance of suitable objectives and constraint choices. The main criterion for the objective choices is that they must be closely related to the decision-maker's point of view. Aggregations of objectives with weighting schemes should be prohibited whenever there might be tradeoffs between these objectives. A wrong choice of the optimization problem objectives might lead to overlooking important information. With regard to the constraints, two types must be used: constraints ensuring relevant objective values and constraints ensuring relevant operation of the control laws. The second type of constraints is as important, like the first one (which is obvious) since good short-term performance might be achieved with a controller which is almost not acting on the system due to unachievable set points for instance. The long-term performance of such solutions will however be very bad and such solutions are hence not interesting at all.

Finally, the methodology proposed an evaluation of the final optimized solutions with long-term simulations representing many different events. This allows a true exploration of the real performance. The comparison of control laws and WWTP performance should mostly be based on this long-term performance, since the rain weather input dataset used during the optimization represents only a limited range of perturbations. These long-term evaluations provide insight into the real performance of the various alternatives as well as their variability. They are important, since the final choice of the solution to implement on the real process will partially be based on expected objective values.

After the definition and development of the methodology, this thesis has presented its application on two case studies. BSM1, a literature case study, was considered first, since its behavior is well documented. Then, the application to the real WWTP of Cambrai then presented and allowed us to have more insight about the current limitations of the methodology as well as about the work required before the application of the methodology: modeling of the influent typical concentrations and flow rate as well as their variations, modeling of the WWTP processes and estimation of reference points corresponding to current performance for the comparison with optimized ones. The methodology proposed hence allowed the unbiased comparison and evaluation of two controls laws for both case studies.

6.1.2 Scope of the methodology- limitations and perspectives

The methodology proposed and the applications performed have allowed us to identify some limitations as well as perspectives for new developments that have to be addressed in the future for a better use of this new tool.

Requirement and limitation #1

First, to be applied, this methodology requires a model of the WWTP with a domain of validity that corresponds to the domain to be explored by manipulating the parameters of the control law. This is probably the main limitation since the domain of validity of such models used is usually not well known. It is, for instance, not clear if a model calibrated for a given WWTP with activated sludge units functioning in sequenced aeration with almost sequenced oxygen concentrations (either 0 or 2 mg/L typically) can work for the same WWTP functioning in continuous aeration with very low oxygen concentrations. Further studies certainly should address this topic.

Perspective #1

This limitation is illustrated in the Appendix D, by an example from the Cambrai case study. Two activated sludge parameters have been measured via respirometry for the two treatment lines. One treatment line is controlled with sequenced aeration based on ORP levels, thus ensuring high levels of oxygen, while the other is controlled with sequenced aeration based on ammonia level and ensures almost simultaneous nitrification and denitrification with very low levels of oxygen. The parameters measured are quite different on the two lines. However, their impact on the simulation results remains to be evaluated and compared with results from default parameters. Another limitation of these developments is that these parameters may even be time- and space-varying and hence the respirometric measurements performed on a local sample at a specific sample time may not represent the overall characteristics of the activated sludge.

Considering this potential limited domain of model validity, another solution could be the development of “extended” models, that would provide insight into the impact of the control laws on the stability of WWTP functioning. For instance, modeling biomass diversity could be interesting as higher biomass diversity is assumed to help stabilize WWTP functioning. An illustration of such a model and the interest of these developments are shown in Appendix E.

Requirement and limitation #2

Another limitation of the methodology is the need for precise knowledge of the influent. This is required for the generation of dry weather, rain weather and long-term influent datasets. The example of Cambrai, however, clearly shows that even medium-size WWTPs have some data available that can allow the generation of these datasets when combining them with a phenomenological influent model. The validity of these generated datasets is naturally limited but this is an initial approach which can be refined as knowledge of the characteristics of the influent increases.

Perspective #2

From an industrial point of view, a specific tool should be developed to help the user of this influent model by clearly indicating the different levels of calibration to be considered and the parameters to be assessed.

Requirement and limitation #3

The third limitation of the methodology presented is that, so far, only set points or limits of the control law have been considered as parameters to optimize. Internal controller parameters defining the stability and the speed of the response of the controller have not been optimized. It is in fact considered that, in our case study, these parameters are fine enough to provide good performance for the whole range of set points explored by the genetic algorithm. A more detailed look into the effect of these parameters could, however, provide more precision into the values of the Pareto front.

Perspective #3

One solution to the limitation identified above could be to optimize these parameters at the same time as the set points and limits. This is not the most convenient way, since the objective of the optimization of the internal parameters is to provide good response time and stability, whereas the objective of the optimization of the set points and limits of the control law are effluent quality, energy consumption, and so forth. There is very little crossed-influence of the internal parameters on the second set of objectives, and very little influence of the set points and limits on the first set of objectives. Optimizing both at the same time is

therefore not reliable, as the multiobjective genetic algorithm would be confused with these small influences as well as the absence of tradeoffs between the two sets of objectives.

On the contrary, a good solution could be to make a two-step optimization. First, overall optimization is performed that proposes new set points and limits for the control law. Then, lower-level optimization is performed for each solution tested in the overall optimization to find best internal parameters of the controllers. A single objective and local optimization is certainly sufficient for this lower level. This is probably a promising solution that should be investigated in the future.

Requirement and limitation #4

The final requirement for the application of this methodology is experience with the use of genetic algorithms. This is necessary in order to choose the best parameterization of the problem, the right parameters of the genetic algorithm and the adequate objectives and constraints. Only this experience can lead to precise results corresponding to the objective of the study. Analyzing the result of the optimization is also a very important task to avoid inaccurate interpretation of the results. This methodology is just a tool; it does not replace an expert on WWTP models and control laws, but it can help to make true comparison of control laws.

Perspective #4

Considering the practical use of the methodology in industry, a specific tool is under development that will be easy to use for anyone and will provide direct binding with traditional modeling software without any requirements of programming skills. The development of such a tool was begun during this PhD program but its finalization in usable form will require further work. This is of practical importance, as nowadays even the best theoretical tool will rarely be used if handy software is not available. This tool will also help the users take the right decisions for their optimizations based on existing expert knowledge of GA functioning and will include modules that will help analyze the optimization results.

Perspective #5

Another perspective considering this limitation is the use of other genetic algorithms requiring less experience. A promising approach has been identified in the work of Reed *et al.* (2003) and Kollat and Reed (2006). This adaptation of NSGA-II, named epsilon-NSGA-II, tackles the main difficulties in the use of NSGA-II which are the choice of the algorithm parameters and the choice of the termination criteria. Four techniques are combined to solve this challenge: epsilon-dominance archiving (Laumanns *et al.*, 2002 ; Deb *et al.*, 2003), dynamic population sizing (Harik and Lobo, 1999) and automatic termination. This algorithm can hence remove the need of choosing any parameter of the genetic algorithm. This will, of course, not provide the adequate problem formulation (in terms of choice of objectives and constraints) but can still enhance the mean performance of the methodology by providing good results faster.

Perspective #6

A final perspective of this thesis, which is not linked to any limitation or requirement, is the extension of the methodology developed to other optimization problems.

Considering only the WWTP optimization case study, many other objectives, decision variables or control handles can be considered. For instance, the adequate sizing and amount of equipment for air production could be optimized from an overall point of view, considering at the same time the consequences in terms of investment and maintenance costs, in terms of energy consumption and in terms of effluent quality. Systems other than the aeration of activated sludge can also be considered. For instance, when the computing power increases in a few years, the optimization of an entire plant such as BSM2 with long-term influent datasets will probably be possible and could certainly provide interesting new solutions. Finally, the methodology developed here can surely be used to develop new plant layouts capable of better disturbance rejections. Other objectives can also be considered for the optimization, such as the risks of WWTP malfunctioning, which could be evaluated thanks to the work of Flores-Alsina *et al.* (2008).

In the water industry, many other systems can also be optimized. An example is presented in Appendix F which concerns the operation of the sewer network during storm events. Many

other optimizations could be performed concerning, for instance, the construction or rehabilitation of drinking water or sewer networks, optimal positioning of sensors in the network (for water quality monitoring, clogging detection, etc.) and the operation of drinking water networks.

6.2 Conclusion about future research

To conclude on the perspectives presented in this section, two main areas for future research can be identified:

- development of knowledge of the impact of simultaneous nitrification/denitrification on models from a specific point of view or, more generally speaking, identifying the impact of the control laws and other operating conditions on the bacteria population and its macroscopic characteristics. Two research approaches are possible in this area, either an experimental one based on observations of these macroscopic characteristics or a more theoretical ones based on the development of specific models for the estimation of the bacterial diversity and/or bacterial macroscopic characteristics. Both approaches will probably be necessary and will have to co-operate. Microbiologists and ecologists will certainly have to share knowledge on this topic

- development of the optimization methodology application for more complex problems and more diverse applications. Many problems can benefit from this methodology in all domains of the water industry. The methodology also should be developed to be more robust, to limit the number of parameters that have to be assessed by the end-user (or even remove any parameterization) and to allow the study of more complex uncoupled problems like the simultaneous assessment of controller set points and internal parameters.

A final smaller area of development concerns the simulation of ORP control laws. As mentioned in this thesis, so far very techniques allow the simulation of such control laws. These simulations will probably need the development of an adequate model capable of representing the overall behavior of the ORP signal and targeted at adequate simulations of the associated control law.

From an industrial point of view, the methodology now has to be disseminated in the enterprise. To achieve this goal, a specific tool needs to be developed, with a specific module targeted at the generation of influent profiles. This tool should be designed from an overall point of view to allow interfacing with many simulation software applications for each specific field of water industry processes and models.

Bibliography

Abraham A., Jain L. (2005). *Evolutionary Multiobjective Optimization.* In Abraham A., Jain L., Goldberg R. (Eds), *Evolutionary Multiobjective Optimization – Theoretical Advances and Applications.* Springer Verlag London, pp. 1-6.

Alex J., Benedetti L., Copp J.B., Gernaey K.V., Jeppsson U., Nopens I., Pons M.N., Rosen C., Steyer J.P., Vanrolleghem P.A. (2008). *Benchmark Simulation Model no. 1 (BSM1).* Report of the IWA Taskgroup on Benchmarking of Control Strategies for WWTPs (http://www.benchmarkwwtp.org), 61 pp.

Balku S, Berber R (2006). *Dynamics of an activated sludge process with nitrification and denitrification: Start-up simulation and optimisation using evolutionary algorithm.* Computer and Chemical Engineering, 30(3), pp. 490-499.

Batstone D.J., Keller J., Angelidaki I., Kayuznyi S.V., Pavlostathis S.G., Rozzi A., Sanders W.T.M., Siegrist H., Vavilin V.A. (2002). *Anaerobic Digestion Model No 1.* Scientific and Technical Report No. 13, IWA Publishing, London, 92 pp.

Bekele E.G., Nicklow J.W. (2007). *Multi-objective automatic calibration of SWAT using NSGA-II.* Journal of Hydrology, 341(3-4), pp. 165-176.

Beraud B., Steyer J.P., Lemoine C., Gernaey K.V., Latrille E. (2007). *Model-based generation of continuous influent data from daily mean measurements available at industrial scale.* 3rd International IWA Conference on Automation in Water Quality Monitoring - AutMoNet2007, September 5-7, 2007, Gent, Belgium,

Booker L. (1987) *Improving search in genetic algorithms.* In: Davis L. (Eds) *Genetic Algorithms and Simulated Annealing.* Morgan Kaufmann, pp. 61-73.

Camacho E.F., Bordons C. (2004). *Model Predictive Control.* Advanced Textbooks in Control and Signal Processing, Springer Verlag, London, 405 pp.

Cecil D. (2008a). *The Control of Denitrification Time in Full Scale by the Automatic Detection of the Low Nitrate Bend in the Redox Curve.* Water Science and Technology, 57(7), pp. 1095-1101.

Cecil D., Skou E. (2008b). *A Model of the Redox Potential Measurement in Anoxic Activated Sludge.* SIDISA conference, June 2008, Florence, Italy.

Chachuat B, Roche N, Latifi M.A. (2005). *Long-term optimal aeration strategies for small-size alternating activated sludge treatment plants.* Chemical Engineering and Processing, 44(5), pp. 591-604.

Charpentier, J. (1992). *La régulation Ecoredox.* L'Eau, l'Industrie, les Nuisances, 154, pp. 50-53.

Charpentier J., Martin G., Wacheux H., Gilles P. (1998). *ORP Regulation and Activated Sludge: 15 years of experience.* Water Science and Technology, 38(3), pp. 197-208.

Comas J., Rodriguez-Roda I., Poch M., Gernaey K.V., Rosen C., Jeppsson U. (2006). *Extension of the IWA/COST simulation benchmark to include expert reasoning for system performance evaluation.* Water Science and Technology, 53(4-5), pp. 331–339.

Copp J.B. (2002). *The COST Simulation Benchmark: Description and Simulator Manual.* Office for Official Publications of the European Community, Luxembourg, 154 pp..

Copp J.B., Jeppsson U. and Rosen C. (2003). *Towards an ASM1-ADM1 state variable interface for plant-wide wastewater modelling.* Proceedings of Water Environment Federation Conference, WEFTEC 2003, Los Angeles, California, USA, Oct. 11-15 2003, pp. 498-510.

Corne D.W., Jerram N.R., Knowles J.D., Oates M.J. (2001). *PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization.* Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001), pp. 283-290.

Cosgrove W.J., Rijsberman F.R. (2000). *World Water Vision: Making Water Everybody's Business.* Earthscan Publications Ltd, London, 128 pp.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, Londres, John Murray.

De Clercq, B. (2003). *Computational Fluid Dynamics of Settling Tanks: Development of Experi-ments and Rheological, Settling and Scraper Submodels.* PhD thesis, Ghent University, 338 pp.

De Jong K.A. (1975). *Analysis of the behaviour of a class of genetic adaptive systems*. PhD thesis, , University of Michigan, Ann Arbor, Michigan, 271 pp.

Deb K., Agrawal S., Pratap A., Meyarivan T. (2000). *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II.* Proceedings of 6th International Conference on Parallel Problem Solving from Nature, 16th-20th Sep 2000, Paris, France, pp. 849-858.

Deb K. (2001). *Multi-Objective Optimisation Using Evolutionary Algorithms.* John Wiley & Sons, 518 pp.

Deb K., Jain S. (2002). *Running Performance Metrics for Evolutionnary Multi-Objective Optimization.* Kanpure Genetic Algorithms Laboratory Report 2002004, India, 18 pp.

Deb K., Mohan M., Mishra S. (2003). *A fast multi-objective evolutionary algorithm for finding well-spread Pareto-optimal solutions.* Technical Report KanGAL 2003002, Indian Institute of Technology, Kanpur, India, 18 pp.

Dytczak M.A., Londry K.L., Oleszkiewicz J.A. (2008). *Activated sludge operational regime has significant impact on the type of nitrifying community and its nitrification rates.* Water Research, 42(8-9), pp. 2320-2328.

Ekama G.A., Barnard J.L., Günthert F.W., Krebs P., McCorquodale J.A., Parker D.S. Wahlberg E.J. (1997). *Secondary Settling Tanks: Theory, Modelling and Operation.* Scientific and Technical Report No. 6, IWA Publishing, London, 236 pp.

Eshelman L.J., Caruana R.A., Schaffer J.D. (1989) *Biases in the crossover landscape.* In J.D. Schaffer (Ed.), Proceedings of 3rd International Conference on Genetic Algorithms, Morgan Kaufman, San Mateo, CA, pp. 10-19.

Eshelman L.J. (1991). *The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination.* In: Rawlins G. (Ed.), Proceedings of the First Workshop on Foundations of Genetic Algorithms, Morgan Kaufmann, pp. 265-283.

Exler O., Antonelo T.L., Egea J.A., Alonso A.A., Banga J.R. (2008). *A Tabu search-based algorithm for mixed-integer nonlinear problems and its application to integrated process and control system design.* Computer and Chemical Engineering, 32(8), pp. 1877-1891.

Fikar M., Chachuat B., Latifi M.A. (2005). *Optimal Operation of Alternating Activated Sludge Processes.* Control Engineering Practice, 13(7), pp.853-861.

Flores-Alsina X., Rodriguez-Roda I., Sin G., Gernaey K.V. (2008). *Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty.* Water Research, 42(17), pp. 4485-4497.

Fuerhacker M., Bauer H., Ellinger R., Sree U., Schmid H., Zibuschka F., Puxbaum H. (2000). *Approach for a novel control strategy for simultaneous nitrification/denitrification in activated sludge reactors*. Water Research, 34(9), pp. 2499-2506.

Gernaey K.V., Vanrolleghem P.A., Lessard P. (2001). *Modeling of a reactive primary clarifier.* Water Science and Technology, 43(7), pp. 73-81.

Gernaey K.V., Jeppsson U., Batstone D.J., Ingildsen P. (2006a). *Impact of reactive settler models on simulated WWTP performance.* Water Science and Technology, 53(1), pp. 159-167.

Gernaey K.V., Rosen C., Jeppsson U. (2006b). *WWTP dynamic disturbance modelling - an essential module for long-term benchmarking development.* Water Science and Technology, 53(4-5), pp. 225-234

Gilles, P. et Pellas, L. (2000). *Utilisation d'un outil de simulation pour la conception d'une station.* Techniques Sciences Méthodes, 4, pp. 49-56.

Goldberg D.E. (1989). Genetic Algorithms in Search, Optimisation and Machine Learning. Addison-Wesley, Reading, Massachusetts, 432 pp.

Grau P., de Gracia M., Vanrolleghem P.A., Ayesa E. (2007). *A new plant-wide modelling methodology for WWTPs.* Water Research, 41(19), pp. 4357-4372.

Grijspeerdt K., Vanrolleghem P., Verstraete W. (1995). *Selection of one-dimensional sedimentation: models for on-line use.* Water Science and Technology, 31(2), pp. 193-204.

Gujer W., Henze M., Mino T., Van Loosdrecht M.C.M. (1999). *Activated Sludge Model No. 3.* Water Science and Technology, 39(1), pp. 183-193.

Gujer W. (2002). *Microscopic versus macroscopic biomass models in activated sludge systems.* Water Science and Technology, 45(6), pp. 1-11.

Halim I., Srinivasan R. (2008). *Designing sustainable alternatives for batch operations using an intelligent simulation-optimization framework.* Chemical Engineering Research and Design, 86(7), pp. 809-822.

Harik G.R., Lobo F.G. (1999). *A parameter-less genetic algorithm.* Technical Report IlliGAL 99009, University of Illinois, Urbana-Champaign, USA, 14 pp.

Hazen, A. (1904). *On Sedimentation.* Transactions of the American Society of Civil Engineers, 53, pp. 45-71.

Héduit A., Thévenot D.R. (1989). Relation between redox potential and oxygen levels in activated sludge reactors. Water Science and Technology, 21(8-9), pp. 947-956.

Henze M., Grady C.P.L. Jr, Gujer W., Marais G.v.R., Matsuo T. (1987). *Activated Sludge Model No. 1.* Scientific and Technical Report No. 1, IAWPRC, London.

Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C., Marais G.v.R. (1995). *Activated Sludge Model No. 2.* Scientific and Technical Report No. 3, IAWQ, London.

Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C., Marais G.v.R., van Loosdrecht M.C.M. (1999). *Activated Sludge Model No. 2d.* Water Science and Technology, 39(1), pp. 165-182.

Henze M., Gujer W., Mino T., Van Loosdrecht M.C.M. (2000). *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3.* IWA Scientific and Technical Report No. 9, IWA Publishing, London, 130 pp.

Herrera F., Lozano M., Verdegay J.L. (1998). *Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioural Analysis*. Artificial Intelligence Review, 12(4), pp.265-319.

Holenda B., Domokos E., Rédey A., Fazakas J. (2007). *Aeration optimisation of a wastewater treatment plant using genetic algorithm.* Optimal Control Applications and Methods, 28(3), pp. 191-208.

Holland J.H. (1975). *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, Michigan, 228 pp.

Horn J., Nafpliotis N., Goldberg D.E. (1994). *A Niched Pareto Genetic Algorithm for Multiobjective Optimization.* In Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Piscataway, new Jersey, June 1994, IEEE Service Center, 1, pp. 82-87.

IFEN (2006). *L'environnement en France – Edition 2006.* Report of 'Institut Français de l'Environnement, 504 pp.

Ingildsen P., Olsson G. (2002). *Exploiting online in-situ ammonium, nitrate and phosphate sensors in full-scale wastewater*

plant operation. Water Science and Technology, 46(4-5), pp. 139-147.

Ingildsen P., Wendelboe H. (2003). *Improved nutrient removal using in situ continuous on-line sensors with short response time.* Water Science and Technology, 48(1), pp. 95-102.

Izquierdo J., Montalvo I., Pérez R., Fuertes V. (2008). *Design optimization of wastewater collection networks by PSO.* Computers and Mathematics with Applications, 56(3), pp. 777-784.

Jeppsson U., Alex J., Pons M.N., Spanjers H., Vanrolleghem P.A. (2002). *Status and future trends of ICA in wastewater treatment - a European perspective.* Water Science and Technology, 45(4-5), pp. 485-494.

Jeppsson U., Pons M.N., Nopens I., Alex J., Copp J., Gernaey K.V., Rosen C., Steyer J.P., Vanrolleghem P.A. (2007). *Benchmark Simulation Model No 2 – General protocol and exploratory case studies.* Water Science and Technology, 56(8), pp. 67-78.

Jones R., Takács I. (2004). *Modelling the impact of anaerobic digestion on the overall performance of biological nutrient removal wastewater treatment plants.* Proceedings of the Water Environment Federation, WEFTEC 2004, Session 24.

Kaelin D., Rieger L., Eugster J., Rottermann K., Banninger C., Siegrist H. (2008). *Potential of in-situ sensors with ion-selective electrodes for aeration control at wastewater treatment plants.* Water Science and Technology, 58(3), pp. 629-637.

Kappeler J., Gujer W. (1992). *Estimation of kinetic parameters of heterotrophic biomass under aerobic conditions and characterization of wastewater for activated sludge modelling.* Water Science and Technology, 25(6), pp. 125-139.

Knowles J.D., Corne, D.W. (2000). *Approximating the non-dominated front using the Pareto Archived Evolution Strategy*. Evolutionary Computation, 8(2), pp. 149-172.

Kollat J.B., Reed P.M. (2006). *Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design*. Advances in Water Resources, 29(6), pp.792-807.

Krebs P., Armbruster M., Rodi W. (2000). *Numerische Nachklärbecken-Modelle.* KA-Wasserwirtschaft, Abwasser, Abfall, 47(7), pp. 985-999.

Kynch G.J. (1952). *A theory of Sedimentation.* Transactions of the Faraday Society, 48, pp. 166-176.

Laumanns M., Thiele L., Deb K., Zitzler E. (2002). *Combining convergence and diversity in evolutionary multiobjective optimization.* Evolutionary Computation, 10(3), pp.263-282.

Lemoine C., Grelier P. (2008). *Simultaneous Nitrification/Denitrification Control for Activated Sludge Waste Water Treatment.* IWA World Water Congress 2008, Vienna, Austria.

Lesouëf, A. (1990). *SIMBAD: un modèle mathématique pour systèmes de boues activées.* Techniques Sciences Méthodes - L'Eau, 85(7-8), pp. 371-378.

Majumdar S., Mitra K., Raha S. (2005). *Optimized species growth in epoxy polymerization with real-coded NSGA-II.* Polymer, 46(25), pp. 11858-11869.

Mathias K.E., Whitley D. (1994). *Transforming the Search Space with Gray Coding.* IEEE Conference on Evolutionary Computation, 1, pp. 513-518.

McMahon K. D., Garcia Martin H., Hugenholtz P. (2007). *Integrating ecology into biotechnology.* Current Opinion in Biotechnology, 18(3), pp. 287-292.

Meijer S. (2004). *ORP Measurements in Activated Sludge – A Literature Review.* In *Theoretical and practical aspects of modelling activated sludge processes.* PhD thesis, Delft university of Technology, The Netherlands.

Nielsen M.K., Önnerth T.B. (1995). *Improvement of a recirculating plant by introducing STAR control.* Water Science & Technology, 31(2), pp. 171-180.

Nopens I., Benedetti L., Jeppsson U., Pons M.N., Alex J., Copp J.B., Gernaey K.V., Rosen C., Steyer J.P., Vanrolleghem P.A. (2008). *Benchmark Simulation Model No 2 – Finalisation of plant layout and default control strategy.* IWA 6th World Water Congress and Exhibition (IWA2008), Sept. 7-12, 2008, Vienna, Austria.

Nopens I., Batstone D., Copp J.B., Jeppsson U., Volcke E., Alex J., Vanrolleghem P.A. (2009). *A new interface for ASM-ADM for use in the Benchmark Simulation Model No. 2.* Water Research (in press).

Olsson G., Nielsen M., Yuan Z., Lynggaard-Jensen A., Steyer J.P. (2005). *Instrumentation, Control and Automation in Wastewater Systems.* IWA Scientific and Technical Report No. 15, IWA Publishing, London, 264 pp.

Olsson G., Jeppsson U. (2006). *Plant-wide control: dream, necessity or reality ?* Water Science and Technology, 53(3), pp. 121–129.

Otterpohl R., Freund M. (1992). *Dynamic models for clarifiers of activated sludge plants with dry and wet weather flows.* Water Science and Technology, 26(5/6), pp. 1391-1400.

Platte F., Kuzmin D., Fredebeul C.H., Turek S. (2005). *Novel Simulation Approaches for Cyclic Steady-state Fixed-bed Processes Exhibiting Sharp Fronts and Shocks.* International Series of Numerical Mathematics, 151, pp.207-223.

Printemps C. (2004). *Apports de la modélisation pour la gestion des systèmes d'assainissement urbains.* PhD Thesis, Université de Poitiers, France.

Printemps C., Manic G. Lemoine C., Payraudeau M., Onnerth T., Ame C. (2006). *Development of a model based tuning tool for simulation of STAR control at Blois WWTP.* 7th International Conference on Hydroinformatics, Sept. 4-8, 2006, Nice, France.

Ramirez, I., and Steyer, J. P. 2008. *Modeling microbial diversity in anaerobic digestion.* Water Science and Technology, 57(2), pp. 265-270.

Reed P, Minsker BS, Goldberg DE.(2003) *Simplifying multiobjective optimization: an automated design methodology for the nondominated sorted genetic algorithm-II.* Water Resources Research, 39(7), 1196.

Reed P., Minsker B. (2004). *Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives.* Journal of Water Resources Planning and Management, 130(2), pp. 140-149.

Reeves C.R. (1993). *Using genetic algorithms with small populations.* In S. Forrest (Ed.), Proceedings of 5th International Conference on Genetic Algorithms, Morgan Kaufman, San Mateo, CA, pp. 92-99.

Reeves C.R. (2003). *Genetic Algorithms.* In Glover F, Kochenberger GA (Eds.) *Handbook of metaheuristics, International series on Operations Research and management science.* Kluwer Academic Publisher, pp. 55-82.

Richalet J., Rault A., Testud J.L., Papon J. (1978). *Model Predictive Heuristic Control : Applications to Industrial processes*. Automatica, 14, pp. 413-428.

Rivas A., Irizar I., Ayesa E. (2008). *Model-based optimisation of wastewater treatment plants design.* Environmental Modelling and Software, 23(4), pp. 435-450.

Rosen C., Jeppsson U., Vanrolleghem P.A. (2004). *Towards a common benchmark for long-term process control and monitoring performance evaluation.* Water Science and Technology, 50(11), pp. 41-49.

Sakizlis V., Perkins J.D., Pistikopoulos E.N. (2004). *Recent advances in optimization-based simultaneous process and control design.* Computers and Chemical Engineering, 28(10), pp. 2069-2086.

Sasaki K., Yamamoto Y., Tsumura K., Hatsumata S., Tatewaki M. (1993). *Simultaneous Removal of Nitrogen and Phosphorus in Intermittently Aerated 2-Tank Activated Sludge Process Using DO and ORP-Bending-Point Control.* Water Science and Technology, 28(11-12), pp. 513-521.

Savic D. (2002). *Single-objective vs. Multiobjective Optimisation for Integrated Decision Support Integrated Assessment and Decision Support.* In: Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society, 24-27 Jun 2002, pp. 7-12.

Spérandio M., Queinnec I. (2004). *Online estimation of wastewater nitrifiable nitrogen, nitrification and denitrification rates, using ORP and DO dynamics.* Water Science and Technology, 49(1), pp. 31-38.

Srinivas N., Deb K. (1995). *Multi-objective function optimization using non-dominated sorting genetic algorithms.* Evolutionary Computation, 2(3), pp. 221-248.

Steyer J.P., Harmand J. (2000). *Economical design of biological systems: new tools for advanced integrated process and control design.* INRA Narbonne (France), laboratory report PUB0200030140092553, 6 pp.

Takács I., Patry G.G., Nolasco D. (1991). *A dynamic model of the clarification-thickening process.* Water Research, 25(10), pp.1263–1271.

Thauré D., Lemoine C., Daniel O., Moatamri N., Chabrol J. (2008). *Optimisation of aeration for activated sludge treatment with simultaneous nitrification denitrification.* Water Science & Technology, 58(3), pp. 639-645.

Vera J., Torres N.V., Moles C.G., Banga J. (2003). *Integrated nonlinear optimization of bioprocesses via linear programming.* AIChE Journal, 49(12), pp. 3173-3187.

Wareham D.G., Hall K.J., Mavinic D.S. (1993). *Real-Time Control of Wastewater Treatment Systems Using ORP.* Water Science and Technology, 28(11-12), pp.273-282.

Weiss M., Plosz B., Essemiani K., Meinhold J. (2006). *Sedimentation of Activated Sludge in Secondary Clarifiers.* Fifth World Congress on Particle Technology, 23-27 April 2006, Orlando, Florida, USA.

Whitley D. (1999). *A Free Lunch Proof for Gray versus Binary Encodings.* Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999), pp. 726-733.

Zaher U., Grau P., Benedetti L., Ayesa E. and Vanrolleghem P.A. (2007). *Transformers for interfacing anaerobic digestion models to pre- and post-treatment processes in a plant-wide modelling context.* Environmental Modelling & Software, 22(1), pp. 40-58.

Zipper T., Fleischmann N., Haberl R. (1998) *Development of a new system for control and optimization of small wastewater treatment plants using oxidation-reduction potential (ORP).* Water Science and Technology, 38(3), pp. 307-314.

Zitzler E., Laumanns M., Thiele L. (2002). *SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization*. In: K.C. Giannakoglou *et al.* (Eds.). *Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001)*. International Centre for Numerical Methods in Engineering (CIMNE), pp. 95-100

Appendixes

Annexes

Appendix A - Verification of BSM1 implementation

The following tables presents the comparison between the simulation results of our implementation (1) and the original BSM1 (2) (provided in the report of Copp, 2002). For dynamic simulations, only results of the dry and rain weather influent datasets are computed since they are the only ones used during the present study.

The conclusion is that the differences are within the tolerances expected and observed between various simulators tested in the BSM1 report.

Table A.1: Steady state results in the activated sludge units and effluent

	Anoxic tank #1		Anoxic tank #2		Aerobic tank #1		Aerobic tank #2		Aerobic tank #3		Effluent	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
SI	30	30	30	30	30	30	30	30	30	30	30	30
SS	2.808	2.808	1.459	1.459	1.150	1.150	0.995	0.995	0.889	0.889	0.889	0.889
XI	1149.125	1149.125	1149.125	1149.125	1149.125	1149.125	1149.125	1149.125	1149.125	1149.125	4.392	4.392
XS	82.135	82.135	76.386	76.386	64.855	64.855	55.694	55.694	49.306	49.306	0.188	0.188
XBH	2551.766	2551.766	2553.385	2553.385	2557.131	2557.131	2559.186	2559.182	2559.344	2559.344	9.782	9.782
XBA	148.389	148.389	148.309	148.309	148.941	148.941	149.527	149.527	149.797	149.797	0.573	0.573
XP	448.852	448.852	449.523	449.523	450.418	450.418	451.315	451.315	452.211	452.211	1.728	1.728
SO	0.004	0.004	0.000	0.0001	1.718	1.718	2.429	2.429	0.491	0.491	0.491	0.491
SNO	5.370	5.367	3.662	3.662	6.541	6.541	9.299	9.299	10.415	10.415	10.415	10.415
SNH	7.918	7.918	8.344	8.344	5.548	5.548	2.967	2.967	1.733	1.733	1.733	1.733
SND	1.217	1.217	0.882	0.882	0.829	0.829	0.767	0.767	0.688	0.688	0.688	0.688
XND	5.285	5.285	5.029	5.029	4.392	4.392	3.879	3.879	3.527	3.527	0.013	0.013
SALK	4.928	4.928	5.080	5.080	4.675	4.675	4.294	4.294	4.126	4.126	4.126	4.126

Table A.2: Steady state results in the various layers of the secondary settler

	Layer #	1	2	3	4	5	6	7	8	9	10
TSS	(1)	12.50	18.11	29.54	68.98	356.07	356.07	356.07	356.07	356.07	6393.98
	(2)	12.50	18.11	29.54	68.98	356.07	356.07	356.07	356.07	356.07	6393.98

Table A.3: Dynamic open-loop results – Effluent concentrations and loads

	Dry weather				Rain weather			
	Effluent average concentration based on load		Effluent average load		Effluent average concentration based on load		Effluent average load	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Q	18061.33	18061.33			23808.18	23808.18		
SI	30	30	541.8400	541.84	22.8388	22.8389	543.7504	543.7515
SS	0.9736	0.9732	17.5848	17.5776	1.1345	1.1338	27.0100	26.9940
XI	4.5779	4.5796	82.6833	82.7147	5.6372	5.6394	134.2121	134.2656
XS	0.2229	0.2228	4.0255	4.0235	0.3448	0.3445	8.2094	8.2022
XBH	10.2206	10.2207	184.5980	184.5999	12.8567	12.8556	306.0952	306.0689
XBA	0.5420	0.5424	9.7891	9.7962	0.6426	0.6433	15.2998	15.3168
XP	1.7560	1.7575	31.7151	31.7427	2.0666	2.0692	49.2025	49.2630
SO	0.7463	0.7461	13.4791	13.4751	0.8472	0.8466	20.1699	20.1560
SNO	8.8231	8.8267	159.3566	159.4226	6.9585	6.9635	165.6694	165.7872
SNH	4.7632	4.7490	86.0291	85.7741	4.9862	4.9700	118.7122	118.3275
SND	0.7291	0.7288	13.1682	13.1641	0.8157	0.8154	19.4204	19.4123
XND	0.0157	0.0157	0.2834	0.2833	0.0236	0.0236	0.5618	0.5614
SALK	4.4565	4.4553	80.4910	80.4681	5.1435	5.1420	122.4574	122.4214

Table A.4: Dynamic open-loop results – Performance indexes

		Dry weather		Rain weather	
		(1)	(2)	(1)	(2)
EQ-index	kg poll.units/d	7066.72	7063.10	8840.37	8835.19
P_sludge	kg SS	17051.79	17055.43	16469.13	16479.51
P_sludge per day	kg SS/d	2435.97	2436.49	2352.73	2354.22
P_sludge_eff	kg SS	1642.26	1642.60	2693.35	2693.86
P_sludge_eff per day	kg SS/d	234.61	234.66	384.76	384.84
P_total_sludge	kg SS	18694.05	18698.04	19162.48	19173.38
P_total_sludge per day	kg SS/d	2670.58	2671.15	2737.50	2739.05
Aeration energy	kWh/d	6476.11	6476.11	6476.11	6476.11
Pumping energy	kWh/d	2966.76	2966.76	2966.76	2966.76
Max effluent *N_tot* level (18 g(N).m^{-3}) was violated during:	d	0.57	0.56	0.31	0.29
i.e.:	% of the time	8.18	8.03	4.46	4.17
The limit was violated at:	occasions	5	5	3	3
Max effluent *SNH* level (4 g(N).m^{-}) was violated during:	d	4.38	4.36	4.44	4.43
i.e.:	% of the time	62.50	62.35	63.39	63.24
The limit was violated at:	occasions	7	7	7	7

Table A.5: Dynamic closed-loop results – Effluent concentrations and loads

	Dry weather				Rain weather			
	Effluent average concentration based on load		Effluent average load		Effluent average concentration based on load		Effluent average load	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Q	18061.26	18061.27			23808.12	23808.13		
SI	30	30	541.8378	541.8381	22.8387	22.8387	543.7459	543.7467
SS	0.8817	0.8803	17.9247	15.8992	1.0296	1.0283	24.5120	24.4819
XI	4.5718	4.5716	82.5732	82.5681	5.6271	5.6256	133.9717	133.9357
XS	0.2007	0.2002	3.6252	3.6159	0.3110	0.3103	7.4050	7.3876
XBH	10.2308	10.2258	184.7804	184.6907	12.8810	12.8719	306.6719	306.4578
XBA	0.5783	0.5787	10.4439	10.4528	0.6856	0.6863	16.3236	16.3390
XP	1.7548	1.7560	31.6946	31.7157	2.0610	2.0624	49.0695	49.1027
SO	1.9997	1.9997	36.1178	36.1178	1.9998	1.9998	47.6122	47.6122
SNO	12.4394	12.4257	224.6717	224.4240	9.1748	9.1650	218.4344	218.2026
SNH	2.5287	2.4968	45.6707	45.0966	3.2172	3.1917	76.5961	75.9875
SND	0.7066	0.7058	12.7621	12.7479	0.7875	0.7870	18.7496	18.7378
XND	0.0144	0.0143	0.2603	0.2597	0.0215	0.0215	0.5122	0.5110
SALK	4.0387	4.0374	72.9446	72.9211	4.8589	4.8578	118.6819	115.6547

Table A.6: Dynamic closed-loop results – Performance indexes

		Dry weather		Rain weather	
		(1)	(2)	(1)	(2)
EQ-index	kg poll.units/d	7556.91	7539.80	9038.69	9020.61
P_sludge	kg SS	17085.46	17091.68	16504.47	16517.12
P_sludge per day	kg SS/d	2440.78	2441.67	2357.78	2359.59
P_sludge_eff	kg SS	1643.87	1643.48	2695.57	2694.42
P_sludge_eff per day	kg SS/d	234.84	234.78	385.08	384.92
P_total_sludge	kg SS	18729.32	18735.15	19200.04	19211.54
P_total_sludge per day	kg SS/d	2675.62	2676.45	2742.86	2744.50
Aeration energy	kWh/d	7241.27	7239.39	7169.77	7168.76
Pumping energy	kWh/d	1488.14	1492.72	1927.53	1934.39
Max effluent *N_tot* level (18 g(N).m^{-3}) was violated during:	d	1.28	1.25	0.79	0.76
i.e.:	% of the time	18.30	17.86	11.31	10.86
The limit was violated at:	occasions	7	7	5	5
Max effluent *SNH* level (4 g(N).m^{-3}) was violated during:	d	1.21	1.19	1.90	1.86
i.e.:	% of the time	17.26	16.96	27.08	26.63
The limit was violated at:	occasions	5	5	8	8

Table A.7: Dynamic closed-loop results – Nitrate controller performance

		Dry weather		Rain weather	
		(1)	(2)	(1)	(2)
Controller type		continuous PI with antiwindup			
Proportional gain (*K*)	$(m^3.d^{-1})/(g(N).m^{-3})$	15000	15000	15000	15000
Integral time constant (*Ti*)	d	0.05	0.05	0.05	0.05
Anti-windup time constant (*Tt*)	d	0.03	0.03	0.03	0.03
Manipulated variable lower limit	$m^3.d^{-1}$	?	0	?	0
Manipulated variable upper limit	$m^3.d^{-1}$	?	92230	?	92230
Controlled variable, *SNO*					
Setpoint	$g(N).m^{-3}$	1	1	1	1
Integral of absolute error (*IAE*)	$(g(N).m^{-3}).d$	1.4818	1.3567	1.8182	1.7016
Integral of squared error (*ISE*)	$(g(N).m^{-3})^2.d$	0.59844	0.5069	0.84205	0.7525
Max. dev. from set-point (max *E*)	$(g(N).m^{-3})$	0.88729	0.8512	0.9092	0.8799
Standard deviation of error (std *E*)	$(g(N).m^{-3})$	0.29234	0.2691	0.3468	0.3279
Variance of error (var *E*)	$(g(N).m^{-3})^2$	0.085463	0.072415	0.12027	0.1075
Manipulated variable (MV), *Qintrec*					
Max deviation of MV (max-min)	$m^3.d^{-1}$	36691.4502	32574.74	77424.58	74738.66
Max dev in change of MV (delta)	$m^3.d^{-1}$	8077.7893	4976.31	8897.2944	5179.9715
Std deviation in change of MV (delta)	$m^3.d^{-1}$	1661.9725	856.69	1641.798	947.2714
Variance in change of MV (delta)	$(m^3.d^{-1})^2$	2762152	733923	2695501	897323

Table A.8: Dynamic closed-loop results – Oxygen controller performance

		Dry weather		Rain weather	
		(1)	(2)	(1)	(2)
Controller type		continuous PI with antiwindup			
Proportional gain (*K*)	$(m^3.d^{-1})/(g(-COD).m^{-3})$	500	500	500	500
Integral time constant (*Ti*)	d	0.001	0.001	0.001	0.001
Anti-windup time constant (*Tt*)	d	0.0002	0.0002	0.0002	0.0002
Controlled variable, *SNO*					
Setpoint	$g(-COD).m^{-3}$	2	2	2	2
Integral of absolute error (*IAE*)	$(g(-COD).m^{-3}).d$	0.0075065	0.006150	0.0069977	0.005324
Integral of squared error (*ISE*)	$(g(-COD).m^{-3})^2.d$	1.75E-05	1.4E-05	5.83E-05	1.02E-05
Max. dev. from set-point (max *E*)	$(g(-COD).m^{-3})$	0.0069017	0.006477	0.065537	0.006222
Standard deviation of error (std *E*)	$(g(-COD).m^{-3})$	0.0015756	0.001391	0.0028853	0.001208
Variance of error (var *E*)	$(g(-COD).m^{-3})^2$	2.4824E-06	1.9347E-06	8.33E-06	1.46E-06
Manipulated variable (MV), *Qintrec*					
Max deviation of MV (max-min)	d^{-1}	186.4117	185.0865	220.2254	187.9100
Max dev in change of MV (delta)	d^{-1}	36.7723	34.2673	35.7611	32.1224
Std deviation in change of MV (delta)	d^{-1}	5.7745	5.5769	5.2715	4.8601
Variance in change of MV (delta)	$(d^{-1})^2$	33.3452	31.1024	27.7882	23.6202

Appendix B - Model-based mass balances for the determination of reaction amounts

In order to determine the mass of pollutants eliminated in the example of AMSTAR and SNDN presented in section 2.4, mass balances computations are performed, based on the simulation results.

First, the masses of COD, TKN, S_{NO} and S_O received in the influent (subscript *i*), released in the effluent and sludge wastage (subscripts *e* and *w*) and present in the system at time t (subscript *s*) are computed for the whole simulation considered, whose duration is *T*.

Based on this data, the mass of TKN oxidized is computed ($M_{TKN,oxid}$):

$$M_{TKN,oxid} = M_{TKN,i} - M_{TKN,e} - M_{TKN,w} + M_{TKN,s}(0) - M_{TKN,s}(T) \quad \text{(B.1)}$$

This gives the amount of nitrate formed ($M_{SNO,formed}$), since for 1g(N) of TKN oxidized, 1 g(N) of nitrate is formed:

$$M_{SNO,formed} = M_{TKN,oxid} \quad \text{(B.2)}$$

The total amount of nitrate reduced in gaseous nitrogen can then be computed:

$$M_{SNO,reduced} = M_{SNO,i} + M_{SNO,formed} - M_{SNO,e} - M_{SNO,w} + M_{SNO,s}(0) - M_{SNO,s}(T) \quad \text{(B.3)}$$

The amount of COD oxidized by the denitrification ($M_{COD,SNO}$) is then based on a stoichiometry of 2.85 g(COD) of COD consumed for 1 g(N) of nitrate reduced:

$$M_{COD,SNO} = 2.86 \cdot M_{SNO,reduced} \quad \text{(B.4)}$$

The amount of COD oxidized with oxygen as the electron acceptor ($M_{COD,SO}$) can then be computed:

$$M_{COD,SO} = M_{COD,i} - M_{COD,e} - M_{COD,w} - M_{COD,SNO} + M_{COD,s}(0) - M_{COD,s}(T) \quad \text{(B.5)}$$

The total mass of *COD* oxidized ($M_{COD,oxid}$) is hence:

$$\begin{aligned} M_{COD,oxid} &= M_{COD,SNO} + M_{COD,SO} \\ &= M_{COD,i} - M_{COD,e} - M_{COD,w} - M_{COD,SNO} + M_{COD,s}(0) - M_{COD,s}(T) \end{aligned} \quad \text{(B.6)}$$

From the stoichiometry of the oxidation of COD with oxygen as the electron acceptor (1g(COD) of COD oxidized for 1 g(-COD) of oxygen consumed), the consumption of oxygen for COD oxidation ($M_{SO,oxid,COD}$) can be computed:

$$M_{SO,oxid,COD} = M_{COD,SO} \quad \text{(B.7)}$$

The consumption of oxygen for nitrification ($M_{SO,oxid,SNH}$) can also be computed based on equation B.1, with a stoichiometry of 4.57 g(-COD) of oxygen consumed for 1 g(N) of TKN oxidized:

$$M_{SO,oxid,SNH} = 4.57 \cdot M_{TKN,oxyd} \quad \text{(B.8)}$$

The total amount of oxygen transferred in water by the air production system ($M_{SO,trans}$) be can then be computed:

$$\begin{aligned} M_{SO,trans} &= M_{SO,oxid,COD} + M_{SO,oxid,SNH} + M_{SO,e} + M_{SO,w} - M_{SO,i} - M_{SO,s}(0) \\ &\quad + M_{SO,s}(T) \end{aligned} \quad \text{(B.9)}$$

and compared with the amount of oxygen produced by the aeration system:

$$M_{SO,prod} = \int_0^T \sum_{aerated\ units} K_L a_i(t) \cdot dt \quad \text{(B.10)}$$

The masses of oxygen produced ($M_{SO,prod}$) and transferred ($M_{SO,trans}$), of COD oxidized ($M_{COD,oxid}$), of nitrate reduced ($M_{SNO,reduc}$) and of TKN oxidized ($M_{TKN,oxid}$) can then be compared for various control laws or various settings of the same control law.

Appendix C - Calibration and validation of Cambrai influent model for COD, TSS, TKN and S_{NH} concentrations

The figures presented on the next pages represent the calibration and validation of the Cambrai influent model for COD, TSS, TKN and SNH (see section 5.2). Model and measurement values are first plotted together for both cases and are followed by figures of the relative error between model and measurement values.

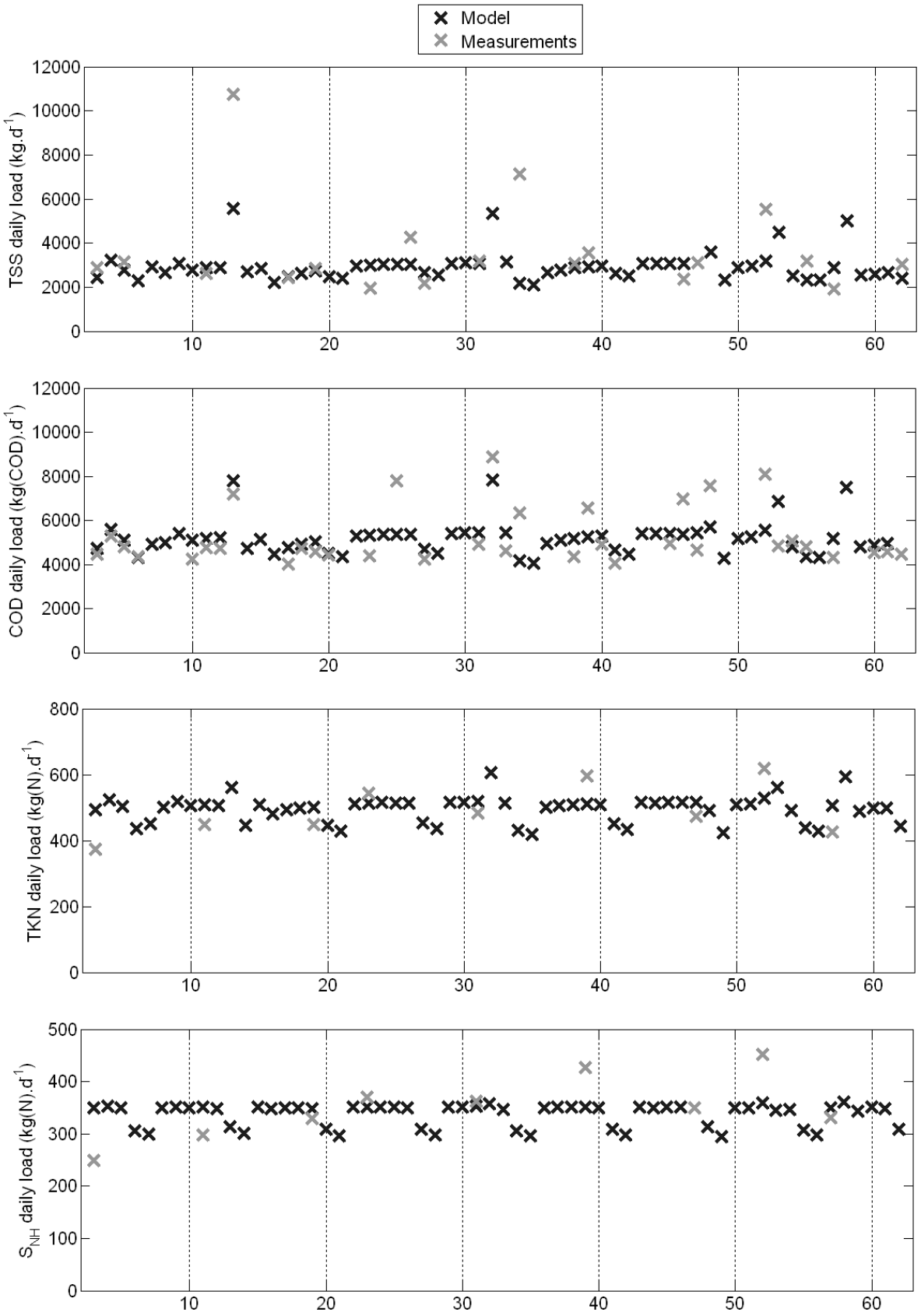

Figure C.1: Calibration of TSS, COD, TKN and S_{NH}

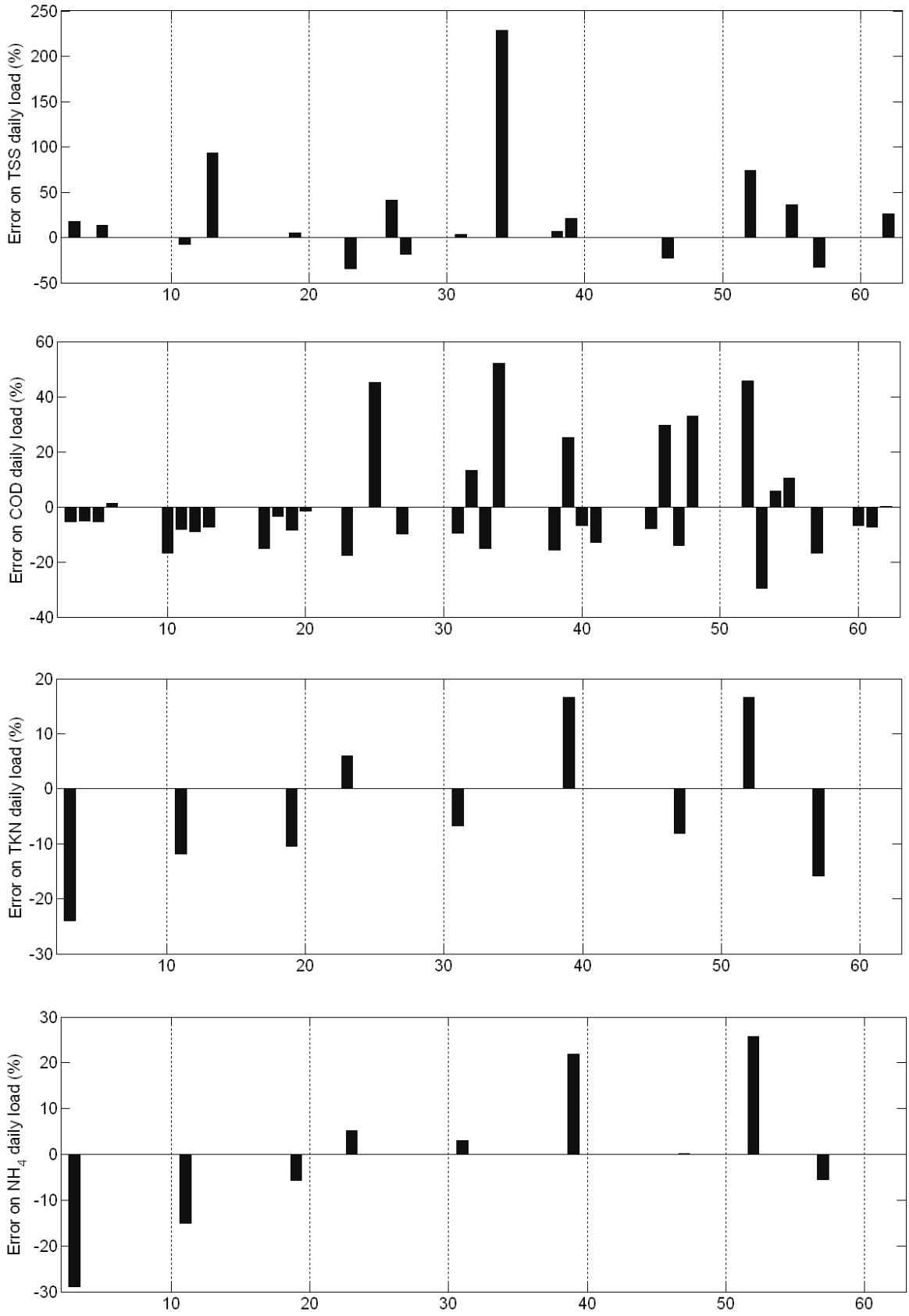

Figure C.2: Calibration errors for TSS, COD, TKN and S_{NH}

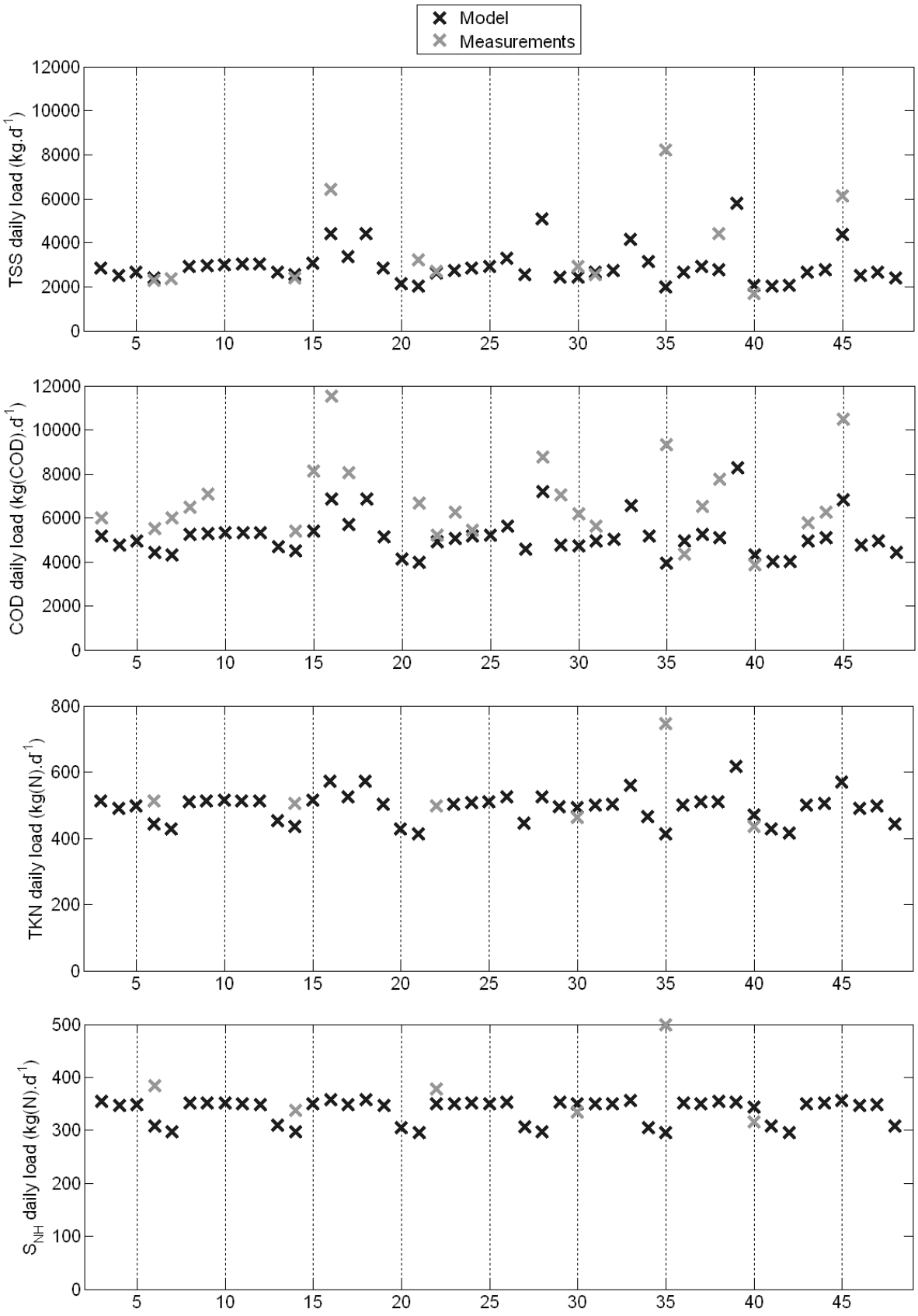

Figure C.3: Validation of TSS, COD, TKN and S_{NH}

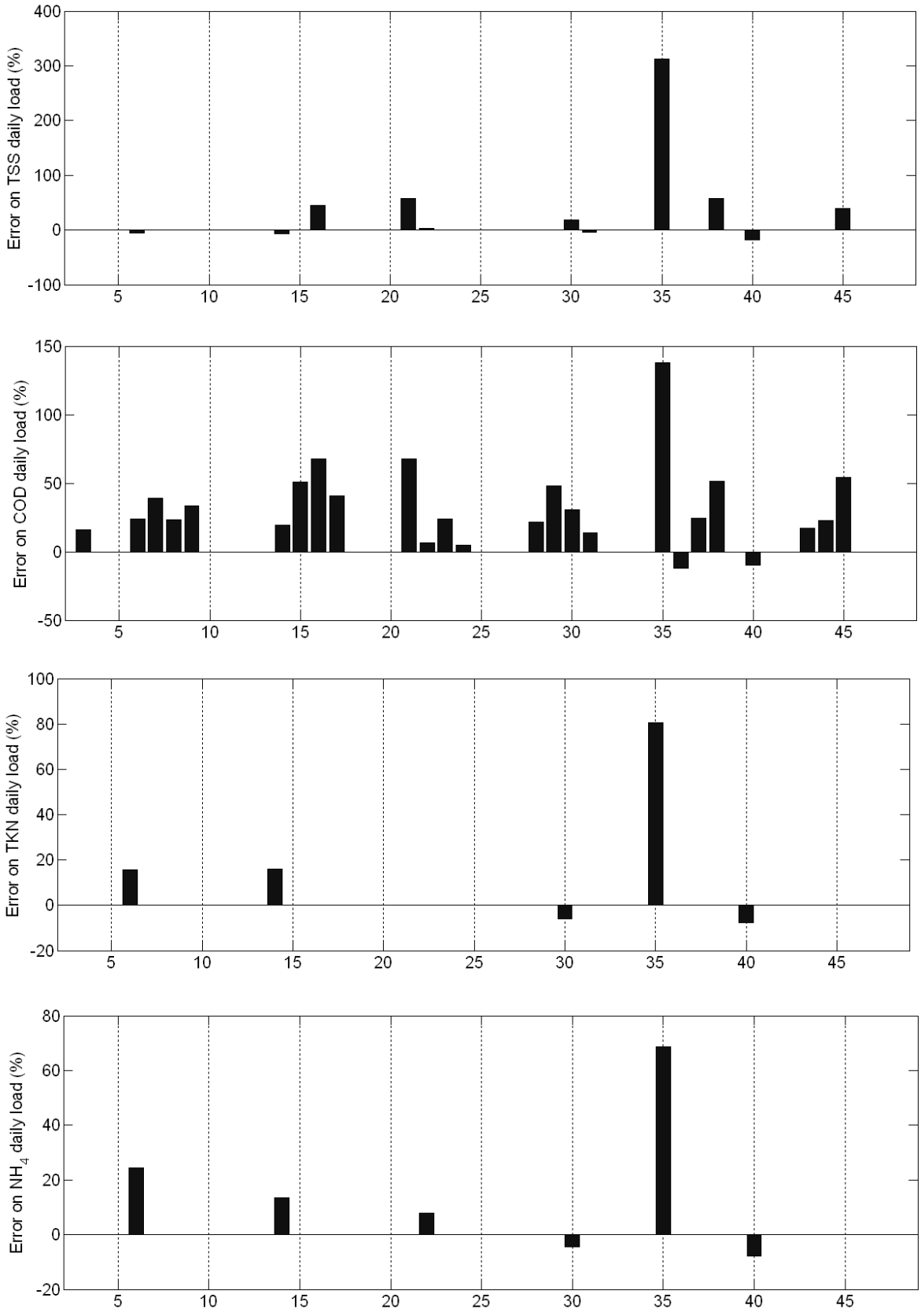

Figure C.4: Validation errors for TSS, COD, TKN and S_{NH}

Appendix D - Perspective #1: use of respirometry for model calibration

In this section, respirometric measurements are performed on the activated sludge of the two treatment lines of the WWTP of Cambrai. Two parameters are determined for the heterotrophic bacteria: the yield and the maximum growth rate. The samples were made on the 10th of September 2008. The first treatment line was controlled with sequenced NDN and low levels of oxygen for three months. The second treatment line was still controlled with the original ORP control and high level of oxygen during aeration phases. The two treatment lines are completely separated and the same influent is equally distributed in the two lines.

D.1 Methodology

Principles of measurement in respirometric tests

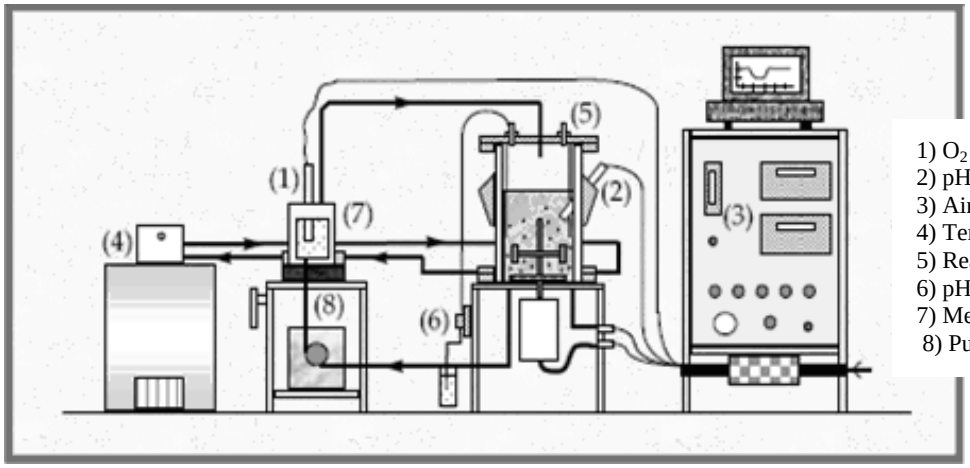

Figure D.1: Layout of a respirometer

For respirometric tests, the activated sludge to characterize is introduced in the reactor (5). This reactor is maintained at a constant temperature thanks to its double envelop controlled with the module for temperature control (4). The pH in the reactor is also controlled (6) with addition of base (NaOH) or acid (HCl) based on the measured value (2). The reactor is mixed and an aeration system continuously injects oxygen. The air flow rate can be adjusted thanks to the air flow meter (3). The measurement cell is sequentially fed in loop with sludge from the reactor thanks to the pump of recirculation (8). Oxygen measurements (1) are performed in this cell which is also continuously mixed but not aerated.

A respirometric measurement is performed in a cyclic fashion:

- the measurement cell is fed with aerated sludge from the reactor with the pump for recirculation,

- the recirculation is stopped and the decrease in oxygen concentration is measured in the cell,

- this decrease is the instantaneous consumption of bacteria (corresponding to this sample of sludge of the reactor at time *t*), named OUR (oxygen uptake rate) in g(-COD).m^{-3}.h^{-1},

- recirculation is then started again and the cycle is repeated.

At the end of the experiment, the concatenation of all points give the graph of the evolution of OUR and hence the bacterial activity.

Measurement of the yield of heterotrophic biomass

The yield of heterotrophic biomass is a macroscopic parameter. Noted Y_H, usually expressed in g of biomass cell COD formed per g of COD substrate oxidized, it is linked with the transfer of *COD* by heterotrophic biomass. For *1* g of COD oxidized by bacteria, $(1-Y_H)$ g of oxygen is consumed and the mass of bacteria is increased by Y_H g. The typical values of this parameter provided in activated sludge models are 0.67 in ASM1, 0.63 in ASM2, 0.625 in ASM2d and 0.6 in ASMAnjou.

The methodology used for the measurement of this parameter is based on the protocol of Kappeler and Gujer (1992) which consists in the addition of a low amount of synthetic carbon substrate based. The detailed protocol is as follows.

The biomass is first placed in endogenous respiration and 20 minutes before the injection of the substrate an inhibitor of autotrophic bacteria (ATU) is injected. The substrate is then injected. The OUR is measured during the whole experiment which finishes when all substrate is consumed, i.e. when the OUR is back to its original value corresponding to endogenous respiration.

A mass balance of COD then provides the yield coefficient by means of the following equation:

$$Y_H = 1 - \frac{\int_0^{T\,\exp} OUR(t) - OUR_{endogenous}\,dt}{COD_{oxyd}} \qquad \text{(D.1)}$$

where *OUR(t)* is the instantaneous measurement of the oxygen uptake rate, $OUR_{endogenous}$ is the consumption of oxygen corresponding to endogenous respiration (i.e. the value of OUR before the injection of the substrate) and COD_{oxyd} is the mass of COD oxidized, i.e. the mass of COD injected.

Measurement of the maximum growth rate of heterotrophic biomass

The maximum growth rate of heterotrophic biomass is also a macroscopic parameter representing the bacteria consortium in its whole. Noted μ_{mH}, usually expressed in d^{-1}, it characterizes the capacity of bacteria to develop quickly. The bigger this parameter, the quicker the substrate can be oxidized. The typical values of this parameter provided in activated sludge models are 6 d^{-1} in ASM1, 6 d^{-1} in ASM2, 6 d^{-1} in ASM2d and 6 d^{-1} in ASMAnjou (all values are corresponding to a mixed liquor temperature of at 20°C, since this parameter is dependant of the temperature).

This parameter is qualified as maximum because the instantaneous growth rate also depends on the concentrations of substrates. When the concentrations of substrates decreases, the growth rate decreases as well. This is characterized by another parameter, named the half saturation concentration, noted K_S. Depending on the bacterial group, different parameters can be found but the two parameters are quite linked.

Two main groups of bacteria have been identified so far in the literature considering this growth rate: r-strategist and K-strategist (Dytczak, 2008). In the first group, the maximum growth rate is high but such bacteria are sensitive to low concentration of substrate and have therefore a high half saturation concentration. *Nitrobacter* is one of these bacteria. On the contrary, K-strategist have a low maximum growth rate but are less sensitive to low concentration of substrate, like *Nitrospira*.

An illustration of these two groups is made on Figure D.2, with two random characteristics. On this graph, it can be seen that K-strategists will be the dominants species if the substrate concentration is mostly lower than 10 $g.m^{-3}$ while r-strategists will be in the other case.

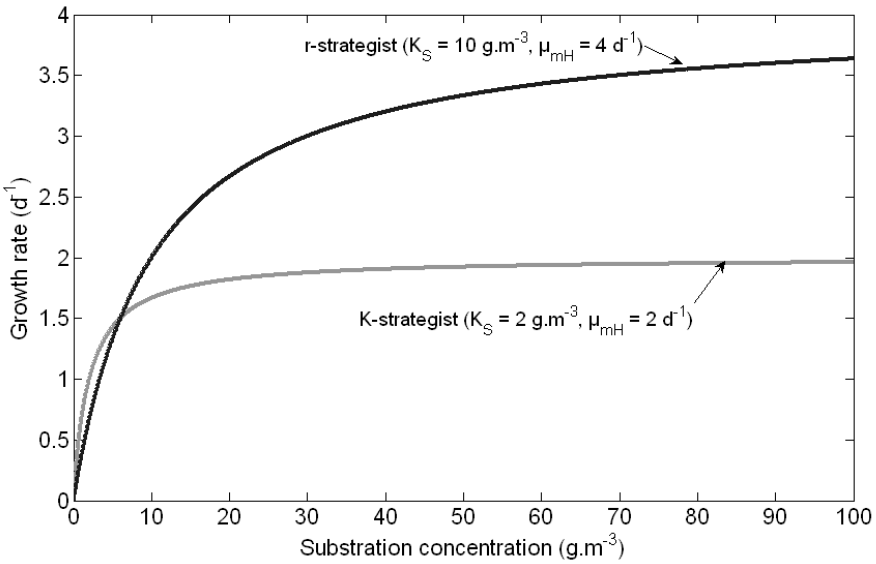

Figure D.2: Example of two characteristics of growth rates

The protocol to determine the maximum growth rate is also based on a proposition of Kappeller and Gujer (1992). This time, a large amount of carbon substrate is injected in the reactor. As with the measurement of the previous parameter, the biomass is first placed in endogenous respiration and an inhibitor of autotrophic bacteria (ATU) is injected 20 minutes before the injection of the substrate.

During the experiment, OUR is measured. Only the period where there is no limitation of substrate is considered. During this period, the OUR can be described by the following equation:

$$OUR(t) = -\left[\frac{1-Y_H}{Y_H}\right]\mu_{mH} X_H(t) - (1-f_p) b_H X_H(t) \qquad \text{(D.2)}$$

Where Y_H is the yield of heterotrophic biomass (in g(biomass cell COD formed).g(COD, substrate oxidized)$^{-1}$), μ_{mH} is the maximum growth rate (d^{-1}), X_H is the concentration of biomass (g(COD).m^{-3}), f_P is the fraction of inert product arising from biomass decay and b_H is the true decay rate of the biomass (i.e. not the decay rate of ASM1 but of the one ASM3, with a default value of 0.2 d^{-1}).

It is then possible to describe the variations of heterotrophic biomass with equation D.3:

$$\frac{dX_H}{dt} = (\mu_{mH} - b_H) X_H(t) \qquad \text{(D.3)}$$

which can be integrated, with X_{H0} the initial biomass concentration (g(COD).m^{-3}):

$$X_{H(t)} = X_{H0} \cdot e^{(\mu_{mH} - b_H)\cdot t} \quad \text{(D.4)}$$

Combining equation D.2 and D.4 gives:

$$OUR(t) = \left[\frac{1 - Y_H}{Y_H} \mu_{mH} + (1 - f_e) b_H \right] X_{Ho} e^{(\mu_{mH} - b_H)t} \quad \text{(D.5)}$$

or

$$\ln\left(OUR(t)\right) = \ln\left[\left(\frac{1 - Y_H}{Y_H} \mu_{mH} + (1 - f_e) b_H \right) X_{HO} \right] - (\mu_{mH} - b_H)t \quad \text{(D.6)}$$

The first term does not depend on the time, so the relation between the logarithm of the OUR and the time is a line, whose directing coefficient is μ_{mH} - b_H. If b_H is known or measured with another experiment, μ_{mH} can be determined. Otherwise, an empirical relation can be used to identify this parameter:

$$\frac{b_H}{\mu_{mH}} = 5\% \quad \text{(D.7)}$$

D.2 Results and discussion

Only the final results of the experiments performed are presented here. They represent the respirometric tests performed between 11 and 15 September 2008 on the sample of the 10 September. For treatment line #1, the TSS concentration and MLSS concentration are respectively measured at 4.30 g(COD)/L and 1.65 g(COD)/L. For the second line, they are respectively 2.28 g(COD)/L and 0.83 g(COD)/L. The reason for this huge difference is unknown. This does not represent typical concentrations observed on the second treatment line. The results obtained are presented in Table D.1 below.

It can be observed in this table that the parameters are really different for both lines. The yield of the first line is 11% lower than the one of the second treatment lines. The maximum growth rate is also lower on the first line compared to the second one. Both differences indicate that a change in the composition or metabolism of the bacteria consortium occurred between the

two lines. This can easily be explained by the low concentrations of oxygen applied in the first line which favors K-strategists, unlike the second line which favors r-strategists.

Table D.1: Results of the measurement of biomass parameters

		Empirical relation between μ_{mH} and b_H		Default value of b_H in ASMAnjou	
	Y_H	μ_{mH}	b_H	μ_{mH}	b_H
Line #1	0.615	2.076	0.099	2.478	0.5
Line #2	0.691	3.125	0.149	3.476	0.5

D.3 Conclusion

The accuracy of the values provided in this example still has to be verified with repetitions of the experiments. However this example clearly shows that there might be a big impact of the control law on the characteristics of the bacteria consortium and hence on the macroscopic model parameters characterizing this biomass. Such modifications are currently not considered in the models used in the optimization methodology proposed in this thesis and should probably be the topic of further research. The experiments presented here should be expanded to other parameters with repetitions of the tests. The real impact of the change of parameter values on the model predictions also should be assessed since it might be limited.

Appendix E - Perspective #2: models taking into account the evolution of the biomass

E.1 Introduction

Activated sludge processes rely on the development of microbial organisms. These organisms are very abundant in wastewater and therefore in activated sludge from WWTPs. They are highly diverse not only in terms of species but also in terms of metabolisms.

So far, microbiologists have done many studies and developed many techniques to try to describe these micro-organism communities as well as possible by making cultures of them. Recent advances, however, have shown that there is still a very limited percentage of the whole population that is known.

On another hand, the performance of the plants depends on the performance of the microbial organisms which have to degrade as many pollutants as possible using as little time and space as possible. Recent studies like that of Lemoine and Grelier (2008) or the one presented in previous section show that changes in operating conditions (with a change in control law, for instance) has a clear impact on the microbial population.

This impact is however poorly known, especially its quantification, considering the distribution and abundance of micro-organisms. To date, most efforts concerned the description of the impact for some given operating conditions based on the qualification of present bacteria but with very limited quantification. Almost no possibilities of modeling have been investigated so far by microbiology engineers (McMahon *et al.*, 2007). On the other hand, ecology engineers have developed many tools to study and model the micro-organisms and their interactions.

Between these two fields (microbiology and ecology), modeling engineers have developed the Activated Sludge Models, which are macro-models relying on the assumption that each transformation of the pollutants is done by a single group of micro-organisms averaging the behavior of the entire community. This is a huge simplification which, for instance, does not take into account competition between micro-organisms for the same substrate. These models have, however, proven their predictive quality for investigation of WWTP functioning.

When using these ASM models for prediction, the main assumption is that the changes in operating conditions investigated have no impact on the macroscopic behavior of the microbial community. It is assumed that just the amount of organisms present in the reactor may change. For instance, in ASM1, the micro-organisms have been defined by two groups: autotrophic bacteria which degrade ammonia in nitrate in the presence of oxygen, and heterotrophic bacteria which degrade COD either in the presence of oxygen or nitrate as electron acceptor. These two groups have given macroscopic characteristics such as maximum growth rate, yield rate and half-saturation constants.

Since these characteristics are established by experimental measurements, they are therefore valid only for a given composition of the microbial community at a given time. It may therefore be interesting to investigate the real composition of these two groups of bacteria. This is of practical interest since the methodology for the optimization of WWTP control laws presented in this thesis targets modifications of the operating conditions and the real impact of these modifications on the macroscopic parameters is not considered. Modifications of treatment efficiency may thus currently be really underestimated.

Two attempts at using modified ASM or ADM models (both are mass-balanced models based on the same kind of macroscopic assumptions) to analyze the impact of such models have been found so far. The first work is presented in Gujer (2002) and is based on ASM3, where cell internal storage is considered. This parameter is assumed to influence kinetics parameters. Therefore, two bacteria can have different kinetics if they don’t have the same amount of internal storage. These internal cell storage modifications are typically induced by the operating conditions encountered by the bacteria. The model developed in this paper hence considers the probability of each bacterium going from one reactor to another one, with different operating conditions in each tank. The impact of this microscopic model on the macroscopic parameters is then assessed.

Another solution has been presented in Ramirez and Steyer (2008) considering the inclusion of a digester biodiversity model in ADM1. The procedure proposed consists in dividing each group of bacteria (acidogens, acetogens and methanogens in ADM1) in a number of smaller groups. Each of these smaller groups is capable of the same transformations, but their characteristics (i.e. kinetic parameters) are not the same.

In this section, the application of the second approach to ASM1 is considered to investigate the impact of some control laws on the composition of bacteria groups. For now, the application is limited to autotrophic bacteria and only ten smaller groups are used, each one with specific characteristics.

E.2 Methodology

A simplification of the BSM1 is used for the investigation presented in this section. The five original activated sludge units are grouped in one anoxic and one aerobic tank. The settler is modeled with a fixed point model instead of Takács's complicated model of (i.e. defined and fixed fractions of the incoming TSS mass goes to the overflow and underflow of the settler). The model of the activated sludge unit is based on ASM1 which is modified to allow the division of the autotrophic bacteria population into 10 groups capable of the same transformations but having different characteristics.

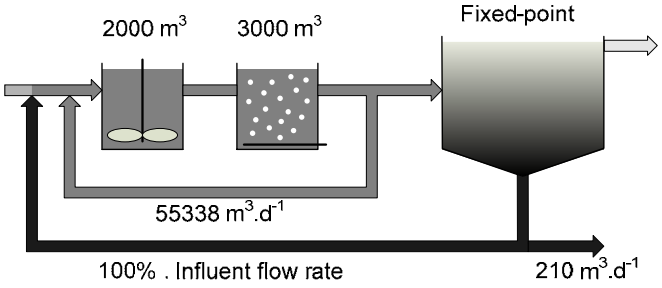

Figure E.1: Simplified layout of BSM1

In this study, only the maximum growth rate (μ_a in day^{-1}) and the half saturation concentration for the oxygen ($K_{o,a}$ in g(-COD).m^{-3}) are modified for each groups. Other parameters are set at their default value in BSM1. These two selected characteristics of the biomass influence the kinetic of the autotrophic bacteria growth rate (or process rate in ASM1, ρ, in day^{-1}):

$$\rho = \mu_a \cdot \left(\frac{S_{NH}}{S_{NH} + K_{NH}} \right) \cdot \left(\frac{S_O}{S_O + K_{OA}} \right) \tag{E.1}$$

In this equation, S_O and S_{NH} are respectively the concentration of oxygen (g(-COD).m^{-3}) and ammonia (g(N).m$^{-3)}$ in the activated sludge, and K_{NH} is the half saturation concentration for the ammonia. This last parameter is considered as constant in this study (the default value in BSM1 is 1 g(N).m^{-3}).

In this study, ten groups of autotrophic bacteria are used. Their characteristics are manually chosen so that each one has a small domain where their growth rate is the biggest of the entire population. These characteristics are depicted in Figure E.2 with a simplified growth rate (ρ_2) since the Monod-term of equation E.2 representing the dependence on ammonia concentration is constant for all groups (at the right side of the figure the curves are ordered from bacteria group number 1 to 10 – simply a reference number that will also be used in other figures):

$$\rho_2 = \mu_a \cdot \left(\frac{S_O}{S_O + K_{OA}} \right) \tag{E.3}$$

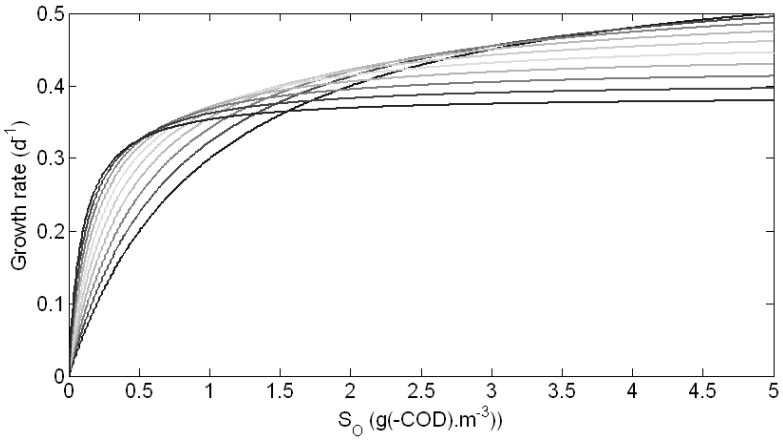

Figure E.2: Characteristics of the ten groups of autotrophic bacteria

The influent dataset used for the simulations comes from BSM1_LT (Rosen *et al.*, 2004). It is composed of 609 days of data representing various events typically occurring at a WWTP. Its main characteristics for this study are a reduction of load between day 280 and day 294 and the total absence of autotrophic bacteria (negligible in BSM simulations).

A first simulation is performed in open loop, i.e. without any controller, the oxygen transfer coefficient being set to 10 h^{-1} in the aerobic compartment. A second simulation is performed in closed loop, a simple PI controller being used to control the oxygen concentration in the aerated unit at 2 g(-COD).m^{-3}. Finally, a third simulation is performed, based on a repetition of the influent dataset. Open loop control is applied during the first half of this third simulation and closed loop control during the second half. The first two simulations will present the typical consequences of both operations while the third simulation will give insight into the consequences of a change in control law during the operation.

E.3 Results and discussion

The results of the first two simulations are presented in Figure E.3. For the open loop simulation, the bacteria groups are in the following order at the end of the simulation, started with the most present one: 8 9 7 10 6 5 4 3 2 1. For the closed loop simulation, the bacteria groups are ordered like this: 4 5 3 6 2 7 8 1 9 10. There is clearly a huge difference between the consequences of the two operating conditions on the evolution of the different bacteria groups. The total mass of bacteria is however almost constant in both simulations, as can be seen in Figure E.4. The difference in total mass between the two simulations is induced by a small difference in terms of pollutant removal (the closed loop case provides more oxygen and the bacterial community is therefore more developed). It can also be observed that the best-performing bacteria groups in the closed loop case are more sensitive to the reduction of the incoming load than the best ones in the open loop case.

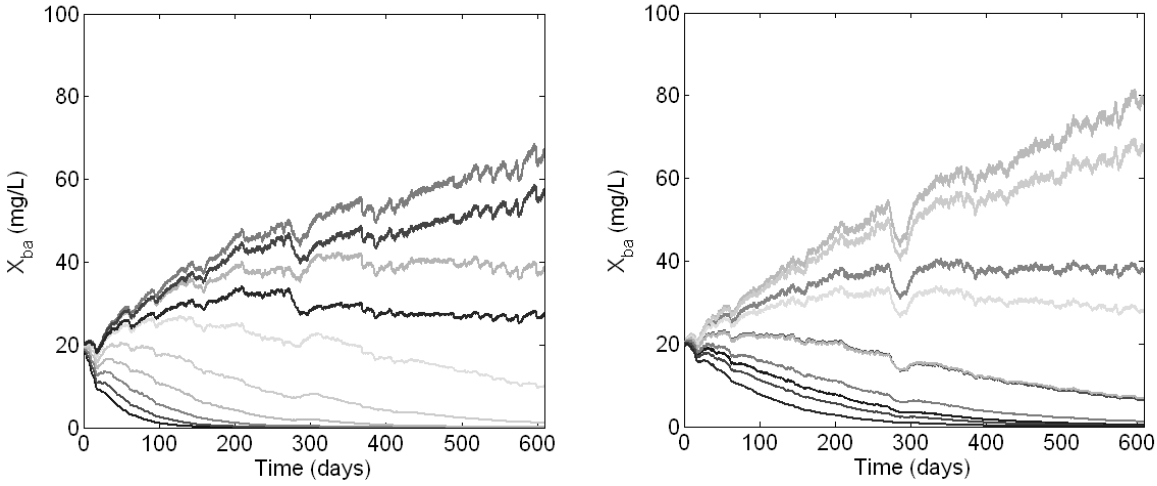

Figure E.3: Results of open loop (left) and closed loop (right) simulations

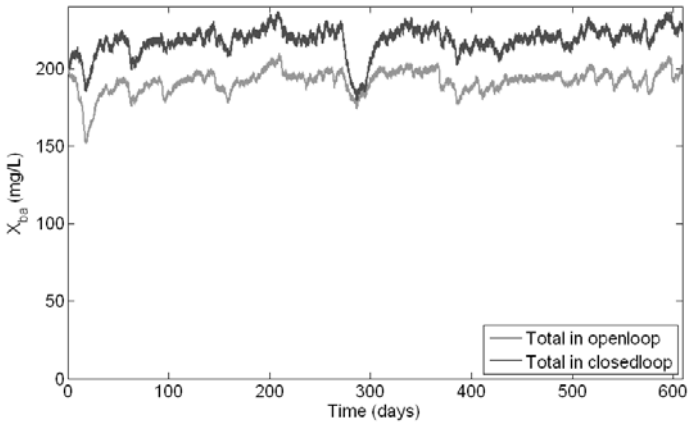

Figure E.4: Total concentrations of bacteria

In order to observe the diversity of the two populations, ecologists frequently use a diversity index named the Shannon index, defined as follows (with n the number of groups and p_i the ratio of the group i over the whole population):

$$I = -\sum_{i=1}^{n} p_i \cdot \log_{10}(p_i) \tag{E.4}$$

The impact on our simulation of using this index is presented in Figure E.5. Diversity decreases steadily for both simulations, with only a very small increase for the open loop case during the reduction of the incoming load. This is easily explained by the fact that no autotrophic bacteria are present in the influent so only species present can develop, and as the conditions do not change, there is no reason for an increase in diversity, except during the reduction of the incoming load.

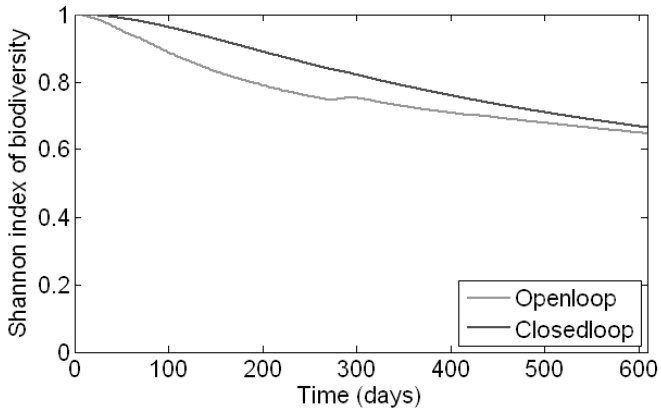

Figure E.5: Shannon index of biodiversity for the open loop and closed loop simulations

When the third simulation is considered with a succession of an open loop and a closed loop control, the results obtained are illustrated in Figure E.6. This time, almost the same bacteria are present at the end of the closed loop simulation as in the closed loop case previously presented, except for one group, which almost disappeared during the open loop period. As seen earlier, diversity decreases steadily (Figure E.7) except during the reduction of the incoming load of the open loop period and at the beginning of the switch between the two control laws. The final diversity is lower than in previous simulations. This is also assumed to be the consequence of no incoming bacteria but this aspect still has to be further qualified.

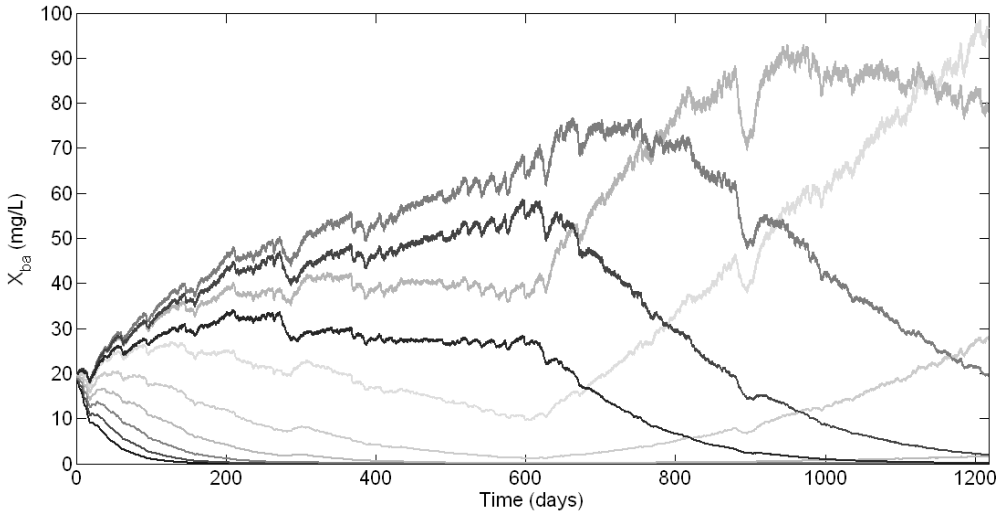

Figure E.6: Results of the succession of open loop and closed loop simulations

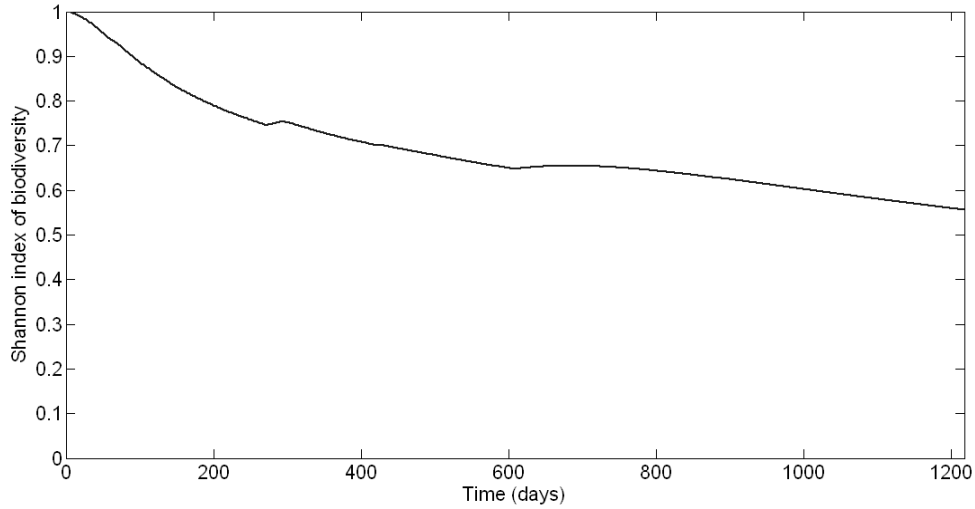

Figure E.7: Shannon index of biodiversity for the combined simulation

E.4 Conclusion and perspectives

This study was just a brief attempt to propose a solution that includes biodiversity in activated sludge models. It is of practical importance to have more information concerning optimization of various control laws based on WWTP models. If the macroscopic characteristics of the bacteria population change due to modifications of the control laws, this has to be assessed in the evaluation of the objectives during the optimization. The methodology proposed here allows this impact to be evaluated and can lead to new estimations of the macroscopic characteristic of the whole bacteria community.

Nevertheless, many points remain to be explored. First, the impact of a small amount of autotrophic bacteria in the influent has to be assessed for all simulations performed here. It

may provide different results, especially when testing the succession of the open loop and closed loop controls. Then, it should be possible to carry model-based respirometric measurements of the bacteria population at various stages of the simulations. This would give us the macroscopic parameters that would have been measured with the real activated sludge at various points in times. These macroscopic parameters can then maybe be used in the simulations for the evaluation of objectives during an optimization procedure which may still be based on the simple ASM1.

Another point which can be of high interest for the evaluation of the performance of a given control law is the evaluation of the quality of the diversity of the population, or more exactly the consequence of the current diversity on a toxic event or a change of incoming load. This is of practical importance for the daily operation of a WWTP, as a control law capable of better handling toxic events could be preferred to one which has better performance in terms of effluent quality and/or energy consumption but takes more time to recover from a toxic event.

Finally, such a tool could also be used to develop strategies to adapt the bacteria population before a known change of operating conditions. The typical interest of such an application is the abrupt increase or decrease in incoming load during holidays, which is not easy to handle, analyze or simulate for now.

Appendix F - Perspective #6: optimization of the operation of sewer networks during storm events

The methodology developed in this thesis for the optimization of WWTP control laws is based on the use of a multiobjective genetic algorithm. This methodology should therefore produce good results for other optimization problems.

An example is detailed here that was developed during the present study to illustrate the benefits of using this same methodology on another optimization problem. The case studied here concerns the optimization of the operation of a combined sewer network during a storm event, based on a very simple case study. In this illustration, the functioning of a pumping station during a storm event is considered.

This pumping station is located at the lower point of the sewer network and is pumping wastewater to the WWTP by mean of a variable speed pump. It is also equipped with a storm basin which is used to limit the volume of combined sewer overflow. The storm basin is located above the pumping tank and is filled by means of a single speed pump. Its emptying is made by gravity back to the pumping tank and is controlled by mean of a valve, which can be either open or closed (no intermediate position is possible).

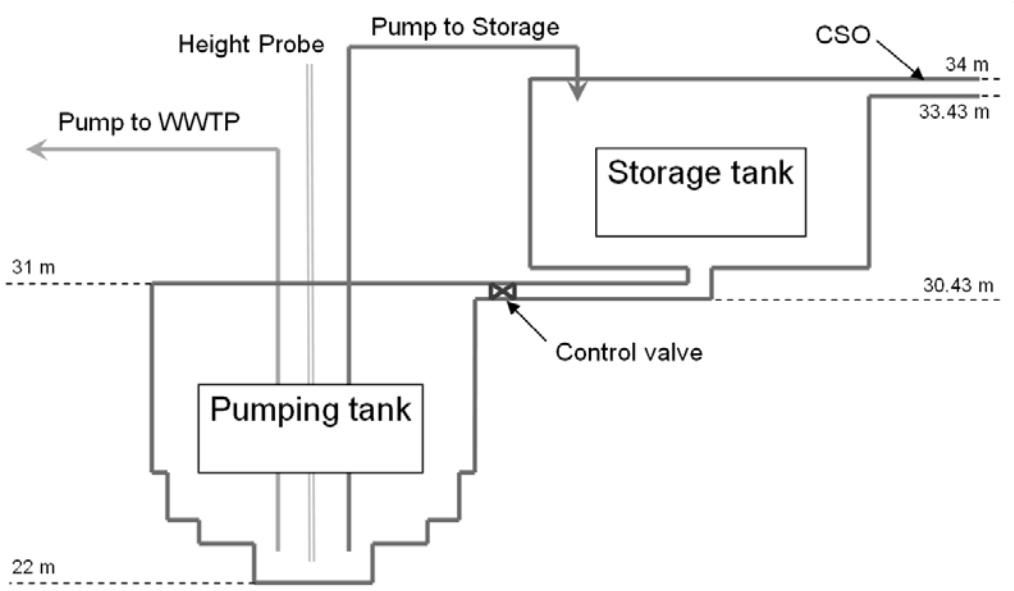

Figure F.1: Layout of the pumping station studied

The pumps and the valve are controlled based on the height of water measured in the pumping tank. The objectives chosen for the optimization are the minimization of the total volume of

CSOs and the maximization of the volume of wastewater stored in the pumping and storage tanks (which will indeed minimize the use of the pump to the WWTP and therefore minimize the energy consumption).

The variables that can be manipulated by the genetic algorithm are the thresholds for the various switches of the speeds of the pump to the WWTP as well as the two thresholds controlling the opening and closing of the control valve. The optimization of the start and stop levels of the pump to the storage basin are not considered in this application. The bounds of the manipulated variables are summarized in Table F.1, together with the values actually used. All heights are formulated as a difference compared to the previous one (except the first one for each equipment). This is meant to make the optimization problem easier. Constraints are also considered in the problem formulation. They ensure that all the heights are in the feasible range of 22 to 31 meters.

For each candidate solution, a single simulation of the sewer network is performed. No stabilization of the simulations is performed like in the methodology presented in this thesis since it is not required in this specific case study. A single run is sufficient because 3 days of simulations are performed and the rain event is located in the middle of the second day. The sewer network is hence already in pseudo steady-state at the beginning at the simulation with a storm tank empty and it is assumed to go back to this steady state at the end of the simulations. The simulations are performed directly with the usual software used for the modeling of the sewer network (Infoworks CS in this case).

Table F.1: Parameters for the optimization of CSO events

		Minimum	Maximum	Actual value
Pump to WWTP	Turn pump off when height below … m	22	31	24.4
	Set speed to 0.13 m3/s when height is … m above the height of turning pump off	0	9	4.55
	Set speed to 0.18 m3/s when height is … m above the height of setting speed to 0.13 m3/s	0	9	0.15
	Set speed to 0.35 m3/s when height is … m above the height of setting speed to 0.18 m3/s	0	9	0.15
	Set speed to 0.69 m3/s when height is … m above the height of setting speed to 0.35 m3/s	0	9	0.35
Valve of storage	Open the valve when the height is below … m	22	31	29.1
	Close the valve when the height is above the opening height plus … m	0	9	0.1

The results of the optimization are presented in Figure F.2. Compared to the actual operation of the pump and valve, a reduction of 21% of the volume of CSO can be obtained with the same mean volume stored, meaning the same utilization of pumps. This increase of performance is however specific to the rain event studied. This event is a big storm event. Such events are happening less than once per year in average. The long term performance hence still has to be studied with longer datasets representing many more events, just like it is done for the WWTP optimization with the long term simulations.

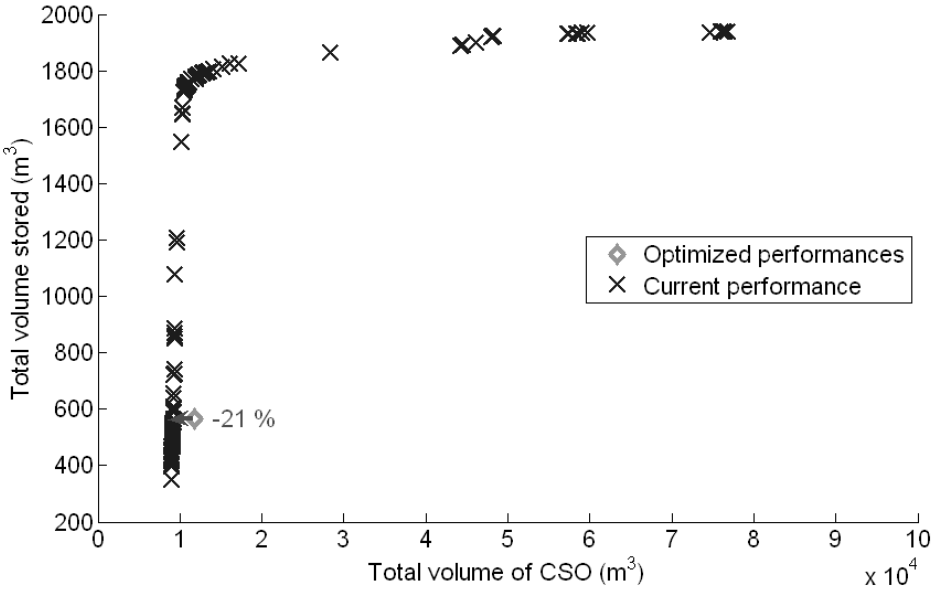

Figure F.2: Results of the optimization of sewer network operation

This short example clearly shows that the methodology developed on the specific case of the optimization of the control law of WWTPs can easily be transposed to other case studies in the field of water, like the operation of a sewer network during storm events.

Appendix G - Extended abstract in French

Résumé étendu en français

Afin de permettre aux lecteurs uniquement francophones d'avoir un aperçu approfondi du contenu de cette thèse, un résumé étendu en français est ici proposé.

G.1 Introduction

L'augmentation continue de la population mondiale, sa concentration dans les villes et l'industrialisation ont eu pour effet d'augmenter continuellement les besoins en eau ainsi que la pollution des milieux naturels durant les dernières décennies. L'utilisation de stations d'épuration a permis de réduire l'impact des eaux usées sur le milieu naturel en les dépolluant avant de les rejeter dans les rivières, lacs, mers et océans.

Afin de continuer à améliorer la qualité de ces traitements, deux solutions sont principalement utilisées. Tout d'abord, de nouveaux procédés plus compacts et plus performants sont inventés pour remplacer les procédés traditionnels très extensifs. D'autre part, les procédés actuellement utilisés sont optimisés afin d'améliorer leurs performances, aussi bien du point de vue de la qualité de l'eau rejetée que de la consommation d'énergie et de réactifs. La réduction de ces consommations est aussi importante que l'amélioration de la qualité d'eau rejetée car elle permet de réduire l'impact environnemental global et le coût de fonctionnement des procédés de traitement.

Afin d'améliorer les procédés existants, le recours à l'automatisation à l'aide de capteurs est de plus en plus courant. Ces derniers permettent d'avoir un suivi en continu de la pollution présente dans les procédés. Il est ainsi possible d'ajuster le traitement à son juste niveau grâce à des lois de commande de plus en plus sophistiquées.

Avant de mettre en place ces lois de commande, il est néanmoins nécessaire d'évaluer leurs performances afin de choisir celle qui sera optimale pour l'usine considérée. En effet, le dimensionnement des procédés et des actionneurs ainsi que les caractéristiques de l'eau usée reçue induisent des compromis différents pour chaque unité de traitement.

Le principal problème concernant cette comparaison est qu'elle est très difficile à réaliser sur des procédés réels. Outre le coût financier du test de ces lois de commande, de longues

périodes d'évaluation sont nécessaires pour chacune d'entre elle. En effet, les procédés de traitement sont principalement basés sur l'utilisation de bactéries effectuant la dépollution de l'eau. Les différentes lois de commande vont exercer une pression sélective sur ces espèces microbiennes et il est nécessaire d'attendre la stabilisation des populations avant de pouvoir véritablement évaluer les performances d'une loi de commande. Cette stabilisation nécessite en général entre 1 et 6 mois avant d'être atteinte. Or, les caractéristiques de l'eau usée reçue ainsi que les conditions opératoires sont très fluctuantes en raison des modifications démographiques et des conditions climatiques. Il est alors très complexe d'effectuer une comparaison non biaisée de diverses lois de commande sur des procédés réels.

Afin de résoudre ce problème, la solution proposée dans cette thèse se base sur les progrès réalisés durant les vingt dernières années dans la modélisation des procédés de dépollution des eaux usées. Il est ainsi maintenant possible de simuler le fonctionnement d'une usine de traitement. Ces simulations permettent ainsi d'évaluer de façon rigoureuse les performances de diverses lois de commande et de divers paramétrages de ces lois.

Afin d'assurer une comparaison non biaisée, il est nécessaire de trouver les meilleures performances possibles des différentes alternatives. Les simulations sont alors couplées à une méthode d'optimisation permettant la recherche automatique des meilleurs réglages des lois de commande. Cette optimisation est basée sur une classe de méthodes d'optimisation nommée « algorithmes génétiques » qui a pour avantage de permettre de rechercher les optimums globaux des problèmes considérés. Un autre avantage de ces méthodes est leur possibilité de couplage direct avec les logiciels de simulations déjà couramment utilisés pour la modélisation des procédés.

L'approche d'optimisation proposée est enfin dite « multi-objectifs ». En effet, de nombreux objectifs antagonistes sont généralement rencontrés lors du choix d'une loi de commande. A titre d'exemple, il y a généralement un compromis à trouver entre la qualité de l'eau rejetée et l'énergie consommée. En effet, la recherche d'une qualité d'eau de sortie la plus propre possible engendre des surcouts énergétiques qui ne sont généralement pas souhaitables du point de vue global de l'impact environnemental et du coût de fonctionnement du procédé de traitement. Un compromis entre ces deux objectifs est généralement souhaité et peut être visualisé avec l'approche multi-objectifs proposée. Enfin, la comparaison des lois de commande sur un seul critère global n'est en général pas satisfaisante. En effet, la plupart des lois développées ont un domaine d'application précis dans lequel elles sont les plus

performantes. Il est alors particulièrement utile d'identifier ces différents domaines et plus particulièrement les frontières qui font passer d'une loi optimale à une autre.

G.2 Méthodologie développée

Afin de permettre le couplage d'un algorithme génétique multi-objectifs avec un modèle de simulation des procédés de traitement, il a été nécessaire de définir, durant la thèse, un protocole permettant le bon fonctionnement de l'optimisation. Cette méthodologie a été développée sur le cas d'école du Benchmark Simulation Model 1 (BSM1).

La méthodologie d'optimisation développée repose tout d'abord sur le principe de l'optimisation dynamique (cf. Figure G.1). Pour chaque solution proposée par l'algorithme d'optimisation, une simulation complète du modèle de station est effectuée. Les performances de la solution sont alors calculées à partir des résultats de cette simulation et sont renvoyées à l'algorithme génétique qui proposera alors de nouvelles solutions à tester. Cette procédure est répétée jusqu'à convergence, c'est-à-dire jusqu'à ce que l'ensemble des meilleures solutions trouvées ne puisse être amélioré.

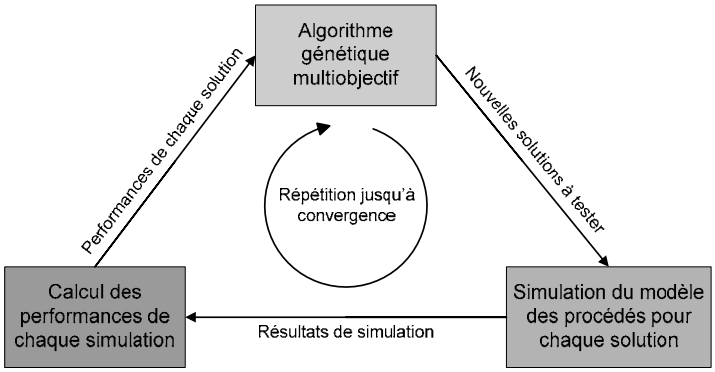

Figure G.1 : Procédure d'optimisation dynamique basée sur l'algorithme génétique NSGA-II et des simulations du modèle des procédés considérés.

Dans ce principe d'optimisation, cinq points principaux ont été définis dans la méthodologie.

Le premier concerne la méthode utilisée pour effectuer les simulations des procédés de la filière de traitement. En effet, tout comme pour les tests réels des lois de commande, ces simulations doivent représenter un état stabilisé du système avant qu'une évaluation des performances ne puisse être effectuée. L'optimisation et les simulations étant effectuées de façon automatisée, un critère a été développé dans cette thèse pour s'assurer de la convergence d'une simulation vers un état stabilisé. Ce critère est tout d'abord basé sur la répétition d'un jeu de données d'entrée représentant une semaine de fonctionnement. La

différence entre l'état des procédés au début et à la fin de la semaine est ensuite observée. Si cette différence est très faible, cela signifie que le procédé a convergé vers un état stabilisé et les performances peuvent alors être évaluées.

Le second point défini dans la méthodologie concerne les jeux de données à utiliser pour les simulations. Pour la stabilisation du procédé précédemment mentionnée, l'utilisation d'un jeu de données d'une semaine représentant un temps sec s'est avérée pertinente. En effet, il permet de représenter les perturbations moyennes que les procédés subissent tout en représentant la charge et le débit moyen typiquement rencontrés. Pour l'évaluation des performances après stabilisation, il a été déterminé que l'utilisation d'un jeu de données représentant un temps de pluie (i.e. deux jours de pluie sur la semaine de données) est la plus pertinente. En effet, ce jeu de données permet d'évaluer la fiabilité de la loi de commande pour des perturbations plus contraignantes que le simple temps sec.

Le choix des objectifs de l'optimisation fait l'objet du troisième point de la méthodologie. Ce choix doit être basé sur les objectifs réels du problème. En effet, tout l'intérêt de l'approche multi-objectifs réside dans l'observation des compromis entre différents objectifs. Ces objectifs doivent donc être liés et aussi opposés que possible. La manipulation des variables de l'optimisation doit de plus avoir un impact sur leurs valeurs. D'un point de vue industriel, l'utilisation de cette approche multi-objectifs permet de proposer un ensemble de solutions aux preneurs de décision. En effet, dans l'approche mono-objectif, quand plusieurs objectifs sont formulés, il est nécessaire de les agréger en un seul à l'aide d'un système de pondérations. Ce travail est généralement fait par la personne qui réalise l'optimisation, que nous nommeront ici analyste. Cette décision se fait généralement avec très peu de liens avec le preneur de décision. L'approche mono-objectif n'offre alors au décisionnaire qu'une seule solution et les compromis ayant mené à celle-ci ont généralement un impact très difficile à évaluer. Le décisionnaire est alors généralement gêné de n'avoir à choisir qu'une seule alternative, sans information sur les compromis ayant mené à cette solution ni sur les autres possibilités au voisinage de cette solution. Dans l'approche multi-objectifs, il est au contraire possible de proposer au décisionnaire un ensemble de compromis parmi lesquels celui-ci peut choisir sa solution en toute connaissance de cause.

Il est ensuite apparu important d'introduire la notion de contrainte dans le quatrième point de la méthodologie et de la distinguée de la notion d'objectif. En effet, les contraintes correspondent à des limites acceptables pour que les solutions soient jugées viables. Au

contraire, les objectifs se rapportent aux performances pour lesquelles un compromis est envisagé et sur lequel on souhaite obtenir des informations.

Il est essentiel de ne pas oublier cette notion de compromis. En effet, dans le cadre de l'optimisation des procédés de dépollution, on pourrait être tenté, à titre illustratif, de considérer les trois critères suivants : la qualité de l'effluent en sortie, la consommation énergétique et la qualité du suivi de consigne du contrôleur. Les deux premiers critères peuvent effectivement être qualifiés « d' objectifs » puisqu'il existe réellement un compromis entre ces deux critères. Au contraire, le troisième critère n'est nécessaire que pour s'assurer que le contrôleur soit bien capable d'agir sur le système au point de fonctionnement choisi. Il n'existe donc pas de compromis entre ce dernier critère et les deux premiers. Celui-ci doit alors être assimilé à une contrainte, avec une valeur maximale admissible. Des contraintes peuvent également être placées sur les valeurs minimales ou maximales des deux objectifs proposés ici. En effet, il est important de limiter l'espace de recherche de l'algorithme d'optimisation afin de favoriser uniquement les solutions véritablement pertinentes sans pour autant limiter les opportunités de découvrir de nouvelles alternatives.

La principale limitation de la méthode d'optimisation proposée jusqu'à présent réside dans l'utilisation d'un jeu de données d'une seule semaine représentant un temps de pluie. Ceci demeure néanmoins l'une des seules solutions pertinentes en raison du temps de calcul associé à l'utilisation de jeux de données plus longs. Cependant, afin d'analyser la robustesse des solutions trouvées, celles-ci sont simulées a posteriori avec le jeu de données de BSM1_LT de 609 jours. Cette technique permet d'éprouver la robustesse de la loi de commande et d'obtenir des valeurs de performances médianes plus fiables que celles obtenues avec le seul temps de pluie. En outre, il est possible de considérer les $5^{ième}$ et $95^{ième}$ percentiles des performances journalières afin d'avoir une visualisation du comportement extrême des lois de contrôle testées et de leur variabilité. Cette évaluation à long terme permet enfin de détecter des solutions aberrantes, qui seraient viables uniquement sur le temps de pluie considéré lors de l'optimisation mais non adaptées sur le long terme. Ainsi, ce raffinement de la méthode évite la sélection d'une solution qui au final est non optimale, mais permet aussi de détecter de nouvelles contraintes à introduire voir de proposer de nouvelles améliorations de la méthodologie d'optimisation ou l'ajout de nouvelles variables d'optimisation.

G.3 Application de la méthode sur le cas d'école du Benchmark Simulation Model 1

Une fois la méthode développée, celle-ci a tout d'abord été appliquée au cas d'école du Benchmark Simulation Model 1 (BSM1). Ce cas d'école a été développé dans le cadre d'un groupe de travail de l'IWA afin de permettre la comparaison et l'évaluation non-biaisée des lois de commande pour les stations d'épuration. Il comprend la définition complète du cas d'étude, avec un modèle de station d'épuration virtuelle, des critères de performances à étudier et des jeux de données à simuler.

L'usine virtuelle est modélisée à l'aide de cinq réacteurs à boues activées représentés par le modèle ASM1, et un décanteur secondaire représenté par le modèle dit de Takács (cf. Figure G.2). Seuls les trois derniers réacteurs sont pourvus d'un système d'aération. Une boucle de recirculation interne permet le retour des nitrates en tête de station tandis qu'une autre boucle de recirculation permet le retour en tête des boues décantées dans le clarificateur secondaire. Une partie de ces boues produites en excès est également extraite de la station.

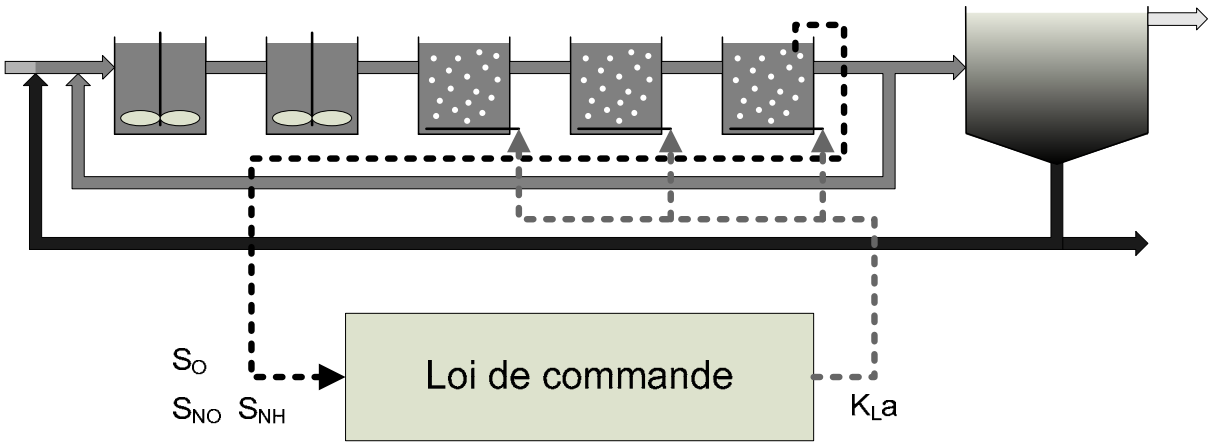

Figure G.2 : Schéma de la modélisation de l'usine de traitement virtuelle proposée dans BSM1 et implantation des lois de contrôle considérées.

Dans notre cas d'étude, deux lois pour le contrôle de l'aération de cette unité sont considérées : AMSTAR, une simplification du module de gestion de l'aération séquencée du système de contrôle STAR©, et SNDN, une loi de commande réalisant une nitrification/dénitrification simultanée. Ces deux lois de contrôle sont implantées de la façon suivante dans le modèle BSM1 (cf. Figure G.2) : les mesures nécessaires à ces lois sont effectuées dans le dernier réacteur à boues activées tandis que le même coefficient de transfert d'oxygène est appliqué pour les trois réacteurs aérés du BSM1.

Pour l'optimisation et la comparaison de ces lois de commande, trois objectifs ont été considérés : les concentrations moyennes en ammoniaque et en azote total dans l'effluent ainsi que la consommation énergétique (somme des énergies utilisées pour le pompage et pour l'aération). Les variables d'optimisation considérées sont les paramètres de réglage des deux lois de commande. Des contraintes sont également ajoutées au problème afin de garantir la fiabilité des réglages obtenus sur le long terme mais également afin de limiter le domaine de recherche à des solutions pertinentes.

Après obtention des solutions optimales grâce à l'algorithme génétique, celles-ci sont simulées sur le long terme à l'aide du jeu de données de 609 jours proposé dans BSM1_LT. Les performances à long terme des solutions sont calculées quotidiennement à partir des résultats de simulation. La médiane et les $5^{ième}$ et $95^{ième}$ percentiles de ces performances quotidiennes sont illustrées sur la Figure G.3, pour les deux lois de contrôle AMSTAR et SNDN ainsi que pour un point de référence correspondant au contrôle en boucle fermé proposé dans BSM1.

Les résultats de l'optimisation démontrent que globalement les performances sont meilleures pour la loi SNDN comparée à AMSTAR. Ainsi, pour toute concentration moyenne d'ammoniaque dans l'effluent supérieure à 3,2 $g(N).m^{-3}$ ou toute concentration moyenne d'azote total dans l'effluent supérieure à 12,3 $g(N).m^{-3}$, le système est plus économe. Les consommations d'énergie observées sont diminuées en moyenne de 8% pour une même concentration moyenne d'ammoniaque dans l'effluent ou de 8,5 % pour une même concentration moyenne d'azote total dans l'effluent. Néanmoins, la loi de commande SNDN ne permet pas d'atteindre des niveaux de qualité d'eau très élevés, c'est-à-dire de très faibles concentrations moyennes en ammoniaque et en azote total dans l'effluent.

Dans ce contexte, AMSTAR est la seule alternative et elle permet d'atteindre des concentrations médianes dans l'effluent de 2,0 $g(N).m^{-3}$ d'ammoniaque ou de 11,8 $g(N).m^{-3}$ d'azote total, au prix d'une consommation énergétique plus importante. Enfin, le contrôle en boucle fermé proposé dans le BSM1 permet d'atteindre des niveaux d'ammoniaque dans l'effluent encore plus bas, jusqu'à 1,7 $g(N).m^{-3}$, cependant au prix d'un rejet d'azote total (i.e. de nitrates principalement) beaucoup plus important ce qui n'est pas forcément l'objectif recherché.

De nombreux autres enseignements peuvent bien entendu être tirés de cette figure. Ce cas d'école a en effet permis de démontrer tout l'intérêt de l'approche proposée pour l'optimisation et la comparaison de ces deux lois de commande. Cette approche permet en effet une véritable visualisation des compromis existants entre les différents objectifs du problème, tout en considérant les performances à long terme de la station. L'application de cette méthodologie sur un cas réel a alors pu être réalisée et va maintenant être détaillée.

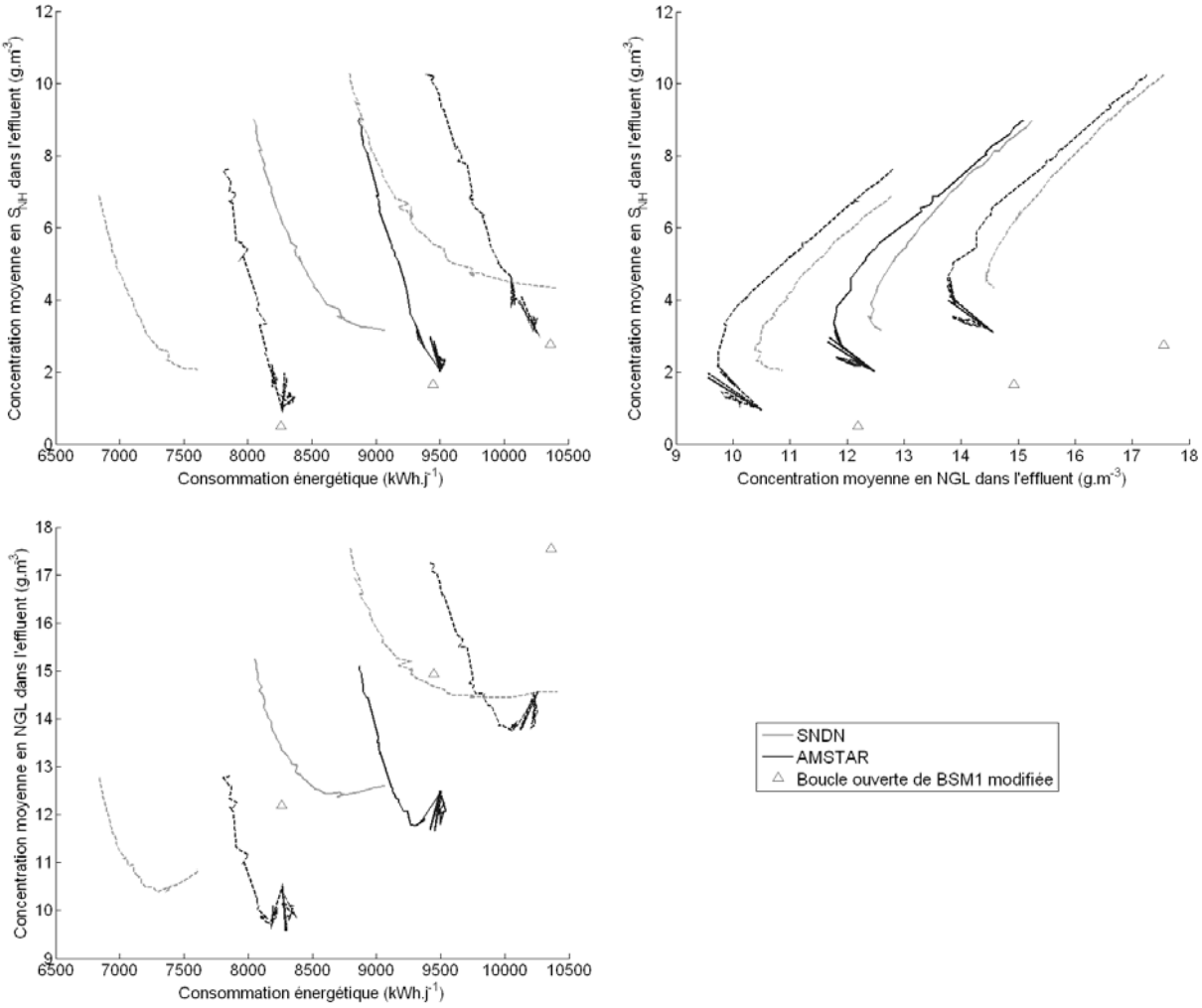

Figure G.3 : 5^{ième}, 50^{ième} et 95^{ième} percentiles des performances journalières obtenues au long terme pour le point de référence proposé dans le BSM1 et pour les solutions optimales des lois de contrôle AMSTAR et SNDN obtenues sur ce cas d'école du BSM1.

G.4 Application de la méthodologie sur le cas réel de la station de dépollution de Cambrai

Après le développement et l'application de la méthodologie sur le cas d'école du BSM1, celle-ci a été appliquée sur le cas réel de l'usine de traitement des eaux usées de Cambrai. L'objectif de cette application est de comparer les performances d'une loi de commande actuellement testée sur site avec celles de la loi de nitrification/dénitrification simultanée (SNDN) dont la mise en place est envisagée. La première loi de contrôle, nommée SABAL dans cette thèse, correspond à un fonctionnement dégradé de SNDN en raison des limitations du système d'aération. Cette loi est basée sur une aération séquencée pilotée par des niveaux d'ammoniaque déclenchant la mise en marche et l'arrêt du système d'aération. L'objectif de la comparaison de ces deux lois est d'évaluer les gains de performances à attendre de la mise en place de la SNDN et de comparer ceux-ci aux coûts de modifications du système d'aération.

Cette application a tout d'abord nécessité la mise au point préalable d'un modèle de l'affluent de la station de Cambrai permettant la génération de jeux de données fiables représentant les caractéristiques typiques à court terme et à long terme de cet affluent. Cette modélisation permet de générer des jeux de données dynamiques et continus à partir des données d'auto-surveillance journalière ou hebdomadaire de la station d'épuration et de l'enregistrement en continu de la pluie et du débit d'entrée de la station. Diverses campagnes de mesures ont été également réalisées sur site afin de permettre l'ajustement de certains paramètres et ratios du modèle d'affluent (profils journaliers, ratio ISS/TSS, etc.).

Dans un second temps, il a été nécessaire de procéder à la modélisation du traitement secondaire de cette usine de dépollution, basée sur le modèle ASMAnjou pour les réacteurs à boues activées et sur le modèle de Takács pour le décanteur secondaire. Le calage de cette modélisation a également été nécessaire. Enfin, divers raffinements ont été apportés au modèle afin de représenter la dénitrification dans le décanteur secondaire, l'oxygénation issue de la chute d'eau en sortie des réacteurs de boues activées et le traitement biologique et physico-chimique du phosphore.

L'optimisation des deux lois de commande (i.e. SABAL et SNDN) a ensuite été effectuée. La loi SABAL est basée sur deux consignes d'ammoniaque pour le démarrage et l'arrêt des suppresseurs avec un débit nominal de 2500 $Nm^3.h^{-1}$ pendant les phases d'aération. La loi

SNDN est basée sur le contrôle en continu des concentrations d'ammoniaque et de nitrate dans la boue activée, le débit d'air calculé étant supposé pouvoir être appliqué sans contraintes de bornes minimum ou maximum. Les objectifs considérés pour ces optimisations sont les concentrations moyennes en ammoniaque et en azote total dans l'effluent ainsi que le volume d'air total injecté. Afin de comparer les résultats obtenus avec les performances réelles des procédés, la Figure G.4 présente les performances selon trois critères disponibles expérimentalement : les concentrations en ammoniaque et en nitrate dans le bassin de boues activées et le volume journalier moyen d'air injecté. En plus des solutions optimales obtenues pour les deux lois de commande, les performances réelles et simulées de la loi SABAL sont illustrées sur la figure pour deux débits d'air de 4200 et 2500 $Nm^3.h^{-1}$.

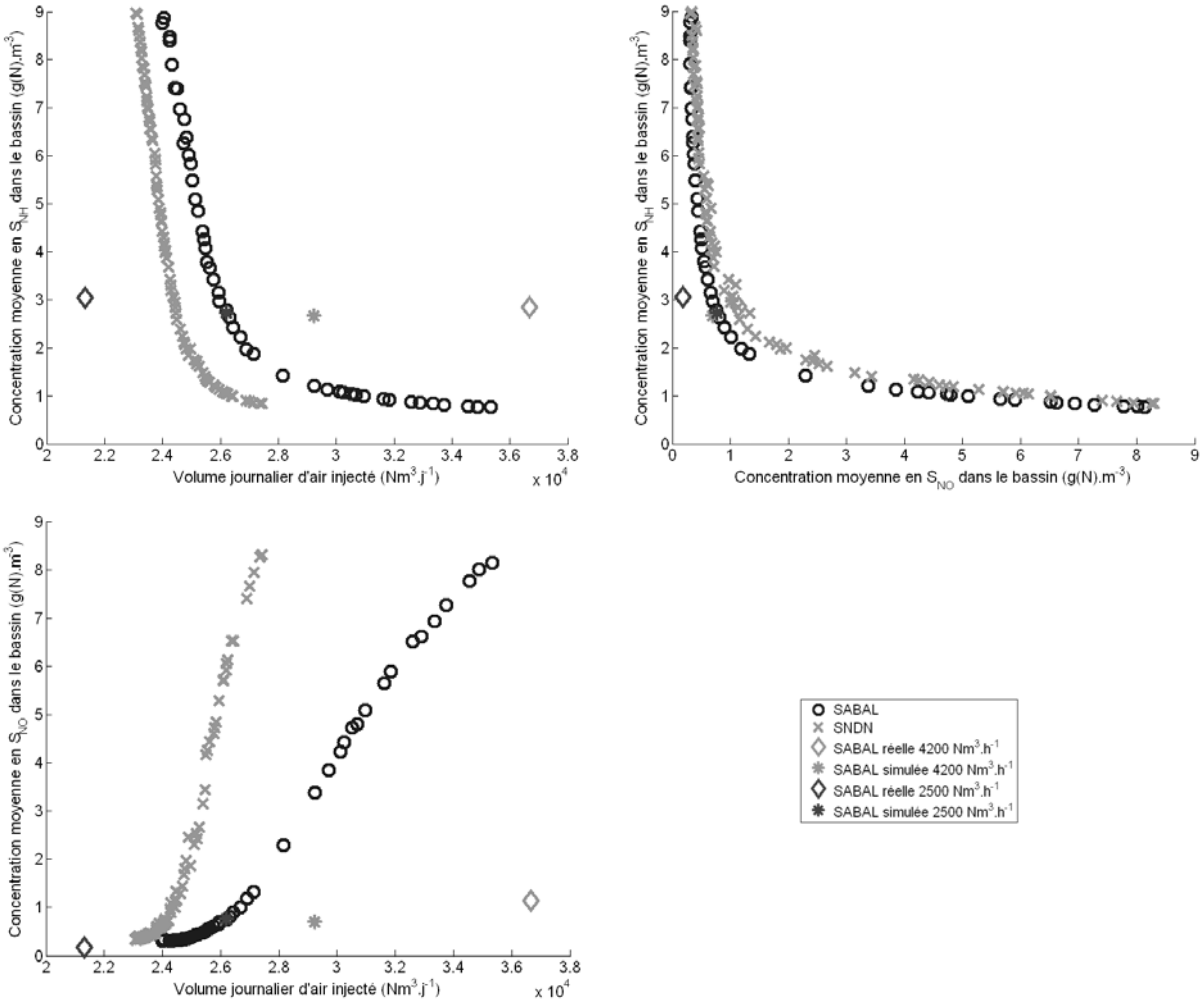

Figure G.4 : Comparaison des performances optimales et réelles des lois de contrôle SABAL et SNDN sur le cas de la station d'épuration de Cambrai

La Figure G.4 permet d'évaluer que l'installation de SNDN permettra d'économiser entre 3 et 5% d'énergie par rapport à SABAL, pour une même qualité d'eau de sortie. En effet, les performances optimales des deux lois de contrôle en termes de concentrations en ammoniaque

et en nitrate sont relativement identiques. Ce faible gain énergétique est toutefois à relativiser. Tout d'abord, la déviation entre les performances réelles et les performances simulées est très grande sur les deux cas de SABAL étudiés. Cela laisse envisager que de plus grandes économies sont peut-être réalisables en réalité avec l'installation de la SNDN. D'autre part, l'optimisation des conditions opératoires réalisée ici ne tient pas compte des modifications engendrées sur la biomasse et donc les paramètres du modèle de boues activées. Cette dernière perspective devra être particulièrement développée à la suite de la thèse pour permettre une meilleure analyse des résultats obtenus. Une autre perspective est l'analyse de l'impact de ces deux lois de contrôle sur le dimensionnement du système de production d'air. En effet, les débits d'air utilisés par celles-ci sont très différents et des compromis énergétiques, financiers et de qualité d'eau sont à rechercher.

Ce cas d'étude réel démontre ici tout l'intérêt de la méthode d'optimisation proposée en permettant une visualisation claire des performances des deux lois de contrôle étudiées. L'analyse des performances à long terme reste à effectuer et permettra d'obtenir des informations plus détaillées et précises sur les caractéristiques de ces deux lois de commande.

G.5 Conclusion

La méthodologie d'optimisation développée dans cette thèse permet l'étude de diverses lois de commande applicables à des stations d'épurations. Le travail effectué s'est essentiellement orienté sur le premier élément nécessaire, à savoir une méthodologie d'optimisation fiable et robuste. Grace à cette méthodologie et à l'usage de la simulation, il est maintenant possible de comparer les performances optimales de diverses lois de commande sur un même cas d'étude et de façon non biaisée. L'approche multi-objectifs utilisée permet une visualisation des compromis possibles entre les diverses lois de commande ainsi que la détermination de leurs domaines d'application respectifs. Les deux cas d'études réalisés ont permis de démontrer la pertinence de la méthode en apportant des résultats d'évaluation et de comparaison de deux lois de commande gérant l'aération des bassins de boues activées, tout d'abord sur un cas d'école (le BSM1) puis sur un cas réel (l'usine de dépollution de Cambrai).

Les perspectives des travaux réalisés dans cette thèse sont multiples. Tout d'abord, il semble nécessaire de développer la connaissance scientifique concernant l'impact des conditions opératoires sur les paramètres des modèles de boues activées. En effet, la procédure d'optimisation utilisée ici permet de modifier ces conditions opératoires et donc les populations bactériennes, les performances de ces dernières étant représentées par les

paramètres des modèles à boues activées. Il est cependant pour l'instant impossible de quantifier ces modifications et donc de le refléter dans les modèles de boues activées.

Un autre développement important concerne l'estimation de l'impact des conditions opératoires sur la biodiversité des boues activées, cette biodiversité pouvant en retour avoir un impact sur la robustesse du procédé. Ces deux impacts restent cependant à évaluer et à modéliser. Ces développements sont nécessaires afin d'introduire plus de réalisme dans les évaluations des performances en tenant également compte de ce critère.

Enfin, une dernière perspective de ce travail concerne le transfert de la méthodologie auprès de l'industriel ayant financé ces travaux et l'application de cette méthodologie à des cas d'étude plus complexes et plus divers. En restant dans le domaine des stations d'épuration, il peut être possible de considérer des problèmes plus complexes tels le choix d'un système d'aération en plus des paramètres de la loi de commande le régulant ou l'optimisation de l'agencement des cuves, procédés et flux dans une station d'épuration pour un traitement optimal. Enfin de nombreux autres problèmes d'optimisation multi-objectifs existent dans le domaine de l'eau et peuvent utiliser cette méthodologie. Il peut s'agir des procédés de production d'eau potable, mais également de la construction et de la gestion des réseaux de distribution d'eau potable ou de collecte des eaux usées.

Title: Methodology for the optimization of wastewater treatment plant control laws based on modeling and multi-objective genetic algorithms.

Abstract: The work presented in this thesis concerns the development of an optimization methodology for control laws of wastewater treatment plants. This work is based on the use of WWTP process models in order to simulate their operation. These simulations are used by a multi-objective genetic algorithm, NSGA-II. This optimization algorithm allows the search of optimal solutions when multiple objectives are considered (e.g. effluent quality and energy consumption). It also visualizes compromises between various control laws as well as their respective best domains of application.

In the first part of this work, the optimization methodology is developed around four main points: the conception of a robust simulation procedure, the choice of input datasets for the simulations, the choice of objectives and constraints to consider and the evaluation of long-term performance and robustness of control laws. This methodology is then applied to the literature case study of BSM1.

In the second part of the work, the methodology is applied to the real case study of the Cambrai wastewater treatment plant. This application includes the development of new aspects, such as generation of dynamic input datasets out of daily monitoring measurements of the wastewater treatment plant, as well as simulation of control laws based on oxydo-reduction potential measurements. This application allowed analysis of the compromises between the control law currently tested on the wastewater treatment plant and a new control law foreseen. The benefits of this modification could thus be clearly observed.

Keywords: optimization, multi-objective, biological treatment, activated sludge, modeling, control, genetic algorithm

Résumé : Le travail présenté dans cet ouvrage concerne le développement d'une méthodologie d'optimisation des lois de contrôle/commande des stations d'épuration des eaux usées urbaines. Ce travail est basé sur l'utilisation des modèles des procédés de traitement afin de réaliser la simulation de leur fonctionnement. Ces simulations sont utilisées par un algorithme d'optimisation multi-objectifs, NSGA-II. Cet algorithme d'optimisation permet la recherche des solutions optimales en fonction des différents objectifs considérés (qualité de l'effluent, consommation énergétique, etc.). Il permet également la visualisation claire des compromis entre diverses lois de contrôle ainsi que la détermination de leurs domaines d'application respectifs.

Dans une première partie de cet ouvrage, la méthodologie est développée autour de quatre axes principaux : la conception d'une méthode de simulation fiable et robuste, le choix des jeux de données d'entrée à utiliser en simulation, le choix des objectifs et contraintes à considérer et enfin l'évaluation des performances et de la robustesse à long terme des lois de contrôles. L'application de cette méthodologie sur le cas d'école du BSM1 est réalisée dans cette première partie.

Dans une seconde partie, la méthodologie développée est appliquée sur le cas réel de l'usine de dépollution de Cambrai. Cette application a nécessité le développement de nouveaux aspects que sont la génération de données d'entrée dynamiques à partir des données d'auto-surveillance de la station d'épuration et la simulation des lois de contrôles basées sur une mesure du potentiel redox. Cette application a permis de visualiser les compromis entre la loi de contrôle actuellement utilisée sur site et une nouvelle loi envisagée. Il a ainsi été possible d'évaluer le gain de performances à attendre de ce changement.

Mots-clés : optimisation, multi-objectif, traitement biologique, boues activées, modélisation, contrôle/commande, algorithme génétique