



HAL
open science

Identification de relations de redondance analytique pour le diagnostic de fonctionnement de capteurs d'une station d'épuration

Yvon Tharrault, Gilles Mourot, José Ragot, David Fiorelli, Serge Gillé

► **To cite this version:**

Yvon Tharrault, Gilles Mourot, José Ragot, David Fiorelli, Serge Gillé. Identification de relations de redondance analytique pour le diagnostic de fonctionnement de capteurs d'une station d'épuration. Nov 2006, pp.CDROM. hal-00118714

HAL Id: hal-00118714

<https://hal.science/hal-00118714>

Submitted on 6 Dec 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification de relations de redondance analytique pour le diagnostic de fonctionnement de capteurs d'une station d'épuration

Yvon THARRAULT*, Gilles MOUROT*, José RAGOT*, David FIORELLI** et Serge GILLÉ**

* Centre de Recherche en Automatique de Nancy (UMR 7039 CNRS-INPL-UHP)

Nancy-Université, INPL, 2, avenue de la Forêt de Haye, 5416 Vandoeuvre-lès-Nancy, France

** Centre de Recherche Public Henri Tudor

LTI, Laboratoire des Technologies Industrielles, 70, route de Luxembourg, L - 4009 Esch-Alzette

{yvon.tharrault, gilles.mourot, jose.ragot}@ensem.inpl-nancy.fr

{david.fiorelli, serge.gille}@tudor.lu

<http://www.cran.uhp-nancy.fr> <http://www.tudor.lu>

Résumé— Cette article traite de l'identification, en vu du diagnostic de fonctionnement de capteur, d'un modèle d'une station d'épuration des eaux usées. Il existe une équivalence entre modèle et relations de redondance analytique, en effet, le modèle défini des relations entre l'entrée et la sortie du système et les relations de redondances lient les variables entre elles, dont celle d'entrée et celle de sortie. Nous recherchons alors les équations de redondances du système, c'est-à-dire les relations liant entre elles uniquement les variables mesurées. Nous avons tout d'abord établi un modèle de connaissance de la station en simplifiant et en linéarisant le modèle ASM1 [5]. Pour compléter le modèle de connaissance ainsi déterminé, nous étudions les données afin de trouver d'autres équations de redondances. Une régression linéaire multiple a permis la détermination d'un modèle statique de type boîte noire, cohérent avec le modèle de connaissance. L'ajout de décalages temporels entre les signaux permet d'améliorer les performances de ce modèle. Nous montrons finalement l'utilisation de ce modèle pour donner le principe de détection-localisation des défauts de mesure.

Mots-clés— Diagnostic, ASM1, Validation de données, Identification, Détection et localisation de défauts capteurs.

I. INTRODUCTION

Le diagnostic de fonctionnement de station d'épuration des eaux usées a été très peu abordé, sur le plan expérimental ou sur le plan théorique. De façon générale effectuer le diagnostic de fonctionnement d'un système consiste à détecter et localiser les défaillances internes (processus lui-même), les défaillances externes (actionneurs, capteurs), les fonctionnements anormaux, puis à estimer les caractéristiques principales des défaillances mises en évidence. Le diagnostic de fonctionnement constitue donc un élément essentiel de toute procédure d'automatisation d'un système puisqu'il permet au système de contrôle de connaître, en permanence, l'état réel du processus (système et instrumentation associée) afin d'élaborer des décisions appropriées.

Dans le cas des stations d'épuration à boues activées, la complexité des mécanismes mis en jeu (cinétiques non-linéaires, paramètres variant dans le temps, absence de mesures fiables, variables non directement atteignables, fortes variations des conditions opératoires et des charges de pollution) impose le développement et l'utilisation de techniques avancées de l'automatique pour développer une stratégie de diagnostic de

fonctionnement en temps réel. Dans nos travaux, nous utilisons le diagnostic par redondance analytique, cela revient à effectuer un test de cohérence entre la mesure et l'estimation fourni par un modèle. Nous recherchons alors les différentes équations de redondances entre variables de la station, à partir de deux méthodes, la première à base d'un modèle de connaissance de type ASM (Activated Sludge Model), et la seconde à l'aide d'une régression linéaire multiple.

Dans une première partie, nous présentons la station d'épuration de Bleesbrück, et en particulier les éléments qui la composent, puis les mesures disponibles dans l'installation. Dans l'optique de mettre en évidence les relations existantes entre les variables caractérisant le processus de "biologie à boues activées" ainsi que leur structure, la seconde partie de l'exposé est consacrée à l'application à cette station du modèle ASM1 [5]. Dans un premier temps, on cherche un modèle linéaire qui dépend des grandeurs disponibles ; il a alors fallu réduire la complexité et linéariser le modèle ASM1. Nous cherchons ensuite les relations entre les variables de la station par régression linéaire multiple. Cette analyse nous permet, à la fois de vérifier dans les données les relations du modèle de connaissance et de mettre en évidence des relations de redondance utilisées pour déterminer un modèle de type boîte noire. Nous montrons ensuite que cette approche est utile pour la localisation de défauts. Dans la dernière partie nous concluons sur les résultats actuels de l'étude.

II. LA STATION DE BLEESBRÜCK

La station a une capacité maximale de traitement de 100 000 Équivalent Habitants et elle traite des eaux usées d'origines urbaines et industrielles. Elle est située à proximité de la Sûre, fleuve qui recueille les eaux épurées.

A. Descriptif du fonctionnement

Les eaux usées brutes passent par une vanne qui adapte automatiquement le débit d'entrée à la capacité hydraulique de la station (au maximum $1200\text{m}^3/\text{heure}$). Les eaux en excès sont déversées directement dans la rivière Sûre. Dans la procédure d'épuration des eaux, nous pouvons distinguer trois phases (Figure 1).

NOMENCLATURE ASMI

S_I	Concentration en matière organique soluble inerte
S_S	Concentration en substrat facilement biodégradable
S_{O_2}	Concentration en oxygène
$S_{O_2}^{sat}$	Saturation de la concentration en oxygène
S_{NO}	Concentration en azote sous forme de nitrites et de nitrates
S_{NH}	Concentration en azote sous forme d'ammonium
S_{ND}	Concentration en azote organique soluble biodégradable
S_{ALK}	Alcalinité
X_I	Concentration en matière organique particulaire inerte
X_S	Concentration en biomasse lentement biodégradable
X_H	Concentration en biomasse hétérotrophe active
X_A	Concentration en biomasse autotrophe active
X_{ND}	Concentration en azote organique particulaire biodégradable
X_P	Concentration en matière organique particulaire résultant du décès de la biomasse
b_A	Coefficient de mortalité des autotrophes
b_H	Coefficient de mortalité des hétérotrophes
f_{XI}	Fraction de biomasse générant les produits particuliers
i_{NBM}	Masse d'azote dans la biomasse
i_{NXI}	Masse d'azote dans la matière organique particulaire inerte
k_a	Taux d'ammonification
k_h	Taux spécifique maximal pour l'hydrolyse
k_{La}	Coefficient de transfert d'oxygène
$K_{(.)}$	Coefficient de demi-saturation :
$K_{NH,A}$	de l'ammoniac pour la biomasse autotrophe
$K_{NH,H}$	de l'ammoniac pour la biomasse hétérotrophe
K_{NO}	du nitrate pour la biomasse hétérotrophe
$K_{O_2,A}$	de l'oxygène pour la biomasse autotrophe
$K_{O_2,H}$	de l'oxygène pour la biomasse hétérotrophe
K_S	pour les hétérotrophes
K_X	pour l'hydrolyse du substrat lentement biodégradable
Y_A	Coefficient de rendement des autotrophes
Y_H	Coefficient de rendement des hétérotrophes
μ_A	Taux de croissance maximal des autotrophes
μ_H	Taux de croissance maximal des hétérotrophes
$\eta_{NO_{3g}}$	Facteur de correction pour μ_H sous condition anoxie
$\eta_{NO_{3h}}$	Facteur de correction pour l'hydrolyse sous condition anoxie

A.1 Le prétraitement.

Il comprend les phases de dégrillage, de dessablage et de dégraissage. Entre le dégrillage et le dessablage, les eaux usées sont relevées de 12 mètres afin qu'elles s'écoulent de manière gravitaire dans le reste de la station. Le dégrillage sert à éliminer les matières volumineuses de plus de 15mm en suspension dans l'eau, qui pourraient nuire aux traitements en aval. Le dessablage a pour objectif d'éliminer les sables, cailloux et autres éléments lourds et non organiques, qui pourraient obstruer ou provoquer une abrasion des conduites et des pompes. Le dégraissage-déshuilage sert à débarrasser l'eau des graisses et huiles, qui peuvent nuire aux traitements de dépollution. Moins denses que l'eau, elles s'accumulent à la surface des bassins et sont ensuite raclées et évacuées. L'opération de dessablage/dégraissage dure environ 15 minutes.

A.2 Le traitement primaire.

Après le prétraitement, l'eau arrive dans un bassin d'aération rectangulaire d'une capacité d'environ 1100m³, appelé pre-

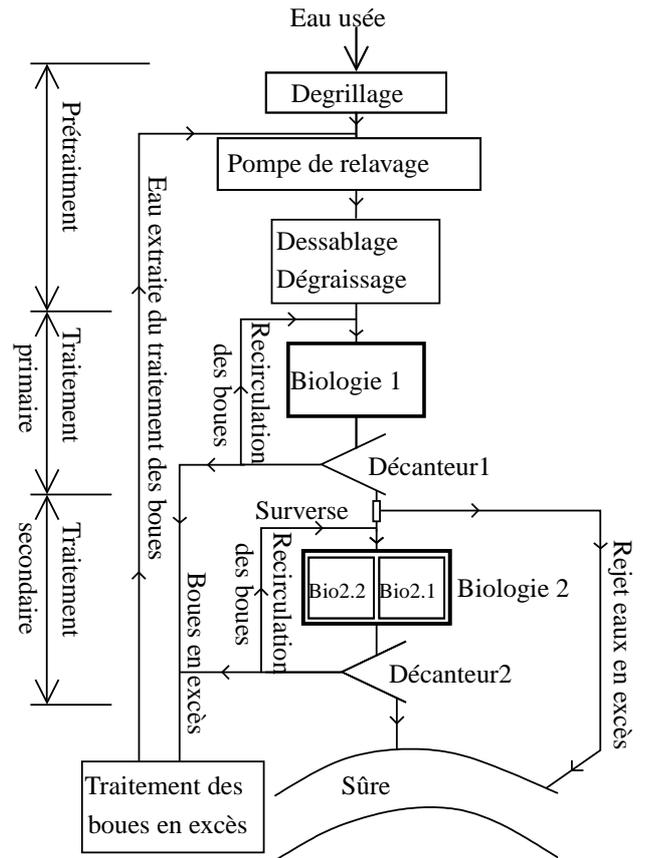


Fig. 1
STRUCTURE DE LA STATION D'ÉPURATION

mière biologie. Le but de cette biologie est d'éliminer principalement les pollutions organiques et le phosphore. Dans ce bassin a lieu l'épuration biologique, technique qui maîtrise et accélère les processus de purification naturelle par la formation de boues activées constituées de bactéries alimentées en oxygène et nourries de la pollution contenue dans l'eau. Le phosphore est éliminé chimiquement par l'adjonction d'un produit fourni par des pompes doseuses. L'eau traitée issue de la première biologie alimente par leur centre deux décanteurs, où elle est séparée des boues. Ces dernières, plus lourdes que l'eau, s'y décantent et sont raclées au fond du bassin dans un silo, où elles peuvent être recirculées soit en tête de station pour y épurer de nouvelles eaux usées, soit pour être éliminées. L'eau épurée, débarrassée des boues, se déverse en périphérie des bassins par débordement. Le temps de traitement des eaux dans la première biologie est d'environ 2,5 heures et dans les décanteurs d'environ 1,5 heures.

A.3 Le traitement secondaire.

La biologie 2 est constituée de deux bassins (Bio 2.1 et Bio 2.2) d'une capacité d'environ 560m³ chacun. C'est ici que se produit la deuxième étape d'affinage du traitement biologique de l'eau en éliminant l'azote par nitrification. Cette opération dure environ 1,5 heures. Une surverse, située juste avant ces bassins, limite le débit d'entrée de la seconde biologie à environ 8000m³/jour. Le surplus est directement déversé dans

Numéro	Nom	Position	Type de Mesure
1	Debit		Débit
2	Temperature		Température
3	Conductivité	Avant la biologie 1	Conductivité
4	AbsorptionUV		Absorption UV
5	PH		PH
6	MES		
7	Oxygène B1	Dans la Biologie 1	Oxygène dissous
8	Ammonium1	Avant le décanteur 1	Ammonium
9	Nitrate1		Nitrate/Nitrite
10	orthophosphate 1		Orthophosphate
11	Hauteur	Avant la Biologie 2	Hauteur de l'eau usée avant la surverse
12 et 14	oxygèneB21 et oxygèneB22	Dans la Biologie 2.1 et 2.2	Oxygène dissous
13 et 15	MES B21 et MES B22		Matières en Suspension
16	Ammonium2	Avant les décanteurs 2	Ammonium
17	Nitrate2		Nitrate/Nitrite
18	Orthophosphate2	Avant le rejet dans la Sûre	Orthophosphate

TABLE I
CAPTEURS DISPONIBLES

la Sûre. Les effluents de la seconde biologie alimentent deux décanteurs secondaires. Les eaux y séjournent pour décantation pendant 1,5 heures, pour être finalement déversées dans la Sûre.

La station possède aussi une partie de retraitement des boues en excès issues des deux biologiques, qui ne sera pas présentée dans cette article car, dans le cadre de notre étude, nous nous intéressons seulement à la partie traitement des eaux usées. Il faut cependant remarquer que les eaux extraites du traitement des boues en excès sont réinjectées dans la station juste avant les pompes de relevage.

B. Capteurs disponibles

Pour cette application, nous cherchons à valider les données des capteurs dont la liste est détaillée dans la table I. Les valeurs fournies par les capteurs sont enregistrées toutes les 30 secondes.

III. LE MODÈLE DE CONNAISSANCE

Dans le but de déterminer les relations entre les variables ainsi que leurs structures, nous allons modéliser un système composé d'une biologie (bassin à boues activées) et d'un décanteur.

La complexité des relations biologiques présentes rend difficile la modélisation phénoménologique. Dans la littérature, il existe différents types de modèles qui décrivent les mécanismes de dégradation biologique dans les stations d'épuration, comme l'ASM1 [5], l'AMS2 [6], l'ASM2d [7], et l'ASM3 [4]. Dans notre cas nous allons utiliser le modèle ASM1, les améliorations du modèle ASM3 concernent surtout le comportement anoxie [2], qui est non présent dans notre cas, alors que les modèles ASM2 et ASM2d prennent en compte l'élimination biologique du phosphore. Le modèle ASM1 décrit les mécanismes biologiques de dégradation des matières organiques, de nitrification et de dénitrification et il est essentiellement utilisé comme modèle de simulation. Dans notre cas, il y a un certain nombre de variables que nous ne

mesurons pas sur la station considérée. L'objectif est de simplifier le modèle ASM1 afin d'éliminer le plus possible les variables non mesurées. Pour cela nous allons utiliser des hypothèses généralement admises dans la littérature et validées par la pratique. Dans cette étude deux simplifications classiques sont apportées au modèle original [9] :

- l'alcalinité S_{ALK} est omise puisqu'elle n'intervient pas dans les dynamiques des autres variables du modèle ;
- les composés particulaires interne X_I et X_P ne sont pas distingués.

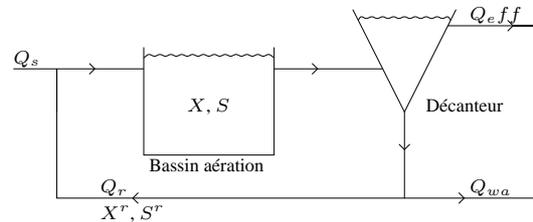


Fig. 2
BASSIN DE BOUES ACTIVÉES ET DÉCANTEUR

La figure 2 présente la biologie et le décanteur. On appelle Q_s le débit entrant dans la station, Q_r le débit de re-circulation des boues, Q_{wa} le débit d'extraction des boues en excès et Q_{eff} le débit à la sortie du décanteur. La modélisation des processus de décantation a fait l'objet de nombreux travaux [12] [8]. Nous considérons dans la suite, en raison de l'absence de mesure dans le décanteur, que celui-ci est parfait, c'est-à-dire qu'aucune boue ne sort par la surverse du décanteur. Cette hypothèse conduit aux deux équations suivantes :

$$S = S^r \quad (1)$$

$$(Q_s + Q_r)X = (Q_r + Q_{wa})X^r \quad (2)$$

où X et S sont les concentrations respectives de particules et de substrat dans le bassin d'aération, et X^r et S^r les

concentrations de particules et de substrat dans le circuit de re-circulation des boues.

A. Modélisation ASM1

Le modèle ASM1 comprend alors 11 variables d'états. C'est un modèle général, qui peut être utilisé pour modéliser la première ou la deuxième biologie.

Remarque : Dans la première biologie, nous ne prenons pas en compte la précipitation du phosphore par l'ajout de produit chimique car elle ne dépend pas de la biologie.

La figure 2 représente le réacteur biologique. Les équations de bilans de matière s'écrivent :

$$\dot{\xi}_\alpha = D_s \xi_\alpha^{in} + D_r \xi_\alpha^r - (D_s + D_r) \xi_\alpha + \mathcal{R}_\alpha \quad (3)$$

où $D_s = \frac{Q_s}{V}$, $D_r = \frac{Q_r}{V}$, V est le volume du bassin d'aération et \mathcal{R}_α correspond au terme de dégradation du composé α et ξ_α^{in} , ξ_α^r , ξ_α désignent la concentration du composé α dans l'influent, dans le recyclage et dans le bassin d'aération, c'est-à-dire :

$$\xi = [S_I, S_S, S_{NO}, S_{NH}, S_{ND}, S_{O_2}, X_I, X_S, X_H, X_A, X_{ND}]^T \quad (4)$$

A partir des équations (1) et (2), l'équation (3) peut s'écrire pour $\alpha = 1$ à 6 :

$$\dot{\xi}_\alpha = D_s (\xi_\alpha^{in} - \xi_\alpha) + \mathcal{R}_\alpha$$

et pour $\alpha = 7$ à 11 :

$$\dot{\xi}_\alpha = D_s \xi_\alpha^{in} - (D_s + D_r) \frac{D_{wa}}{D_r + D_{wa}} \xi_\alpha + \mathcal{R}_\alpha$$

où $D_{wa} = \frac{Q_{wa}}{V}$.

Les différentes valeurs de \mathcal{R}_α sont explicitées dans la table II et la table III. Les grandeurs Y_A , Y_H , f_{XI} , i_{NBM} , i_{NXI} ,

α	\mathcal{R}_α
1	0
2	$-\frac{1}{Y_H} (\rho_1 + \rho_2) + \rho_7$
3	$-\frac{1 - Y_H}{2.86 Y_H} \rho_2 + \frac{1}{Y_A} \rho_3$
4	$-i_{NBM} (\rho_1 + \rho_2) - \left(i_{NBM} + \frac{1}{Y_A} \right) \rho_3 + \rho_6$
5	$-\rho_6 + \rho_8$
6	$k_{La} (S_{O_2}^{sat} - S_{O_2}) - \frac{1 - Y_H}{Y_H} \rho_1 - \frac{4.57 - Y_A}{Y_A} \rho_3$
7	$f_{XI} (\rho_4 + \rho_5)$
8	$(1 - f_{XI}) (\rho_4 + \rho_5) - \rho_7$
9	$\rho_1 + \rho_2 - \rho_4$
10	$\rho_3 - \rho_5$
11	$(i_{NBM} - f_{XI} i_{NXI}) (\rho_4 + \rho_5) - \rho_8$

TABLE II

STECIOMETRIE DES RÉACTIONS DU MODÈLE DE DÉGRADATION

$\eta_{NO_{3g}}$, $\eta_{NO_{3h}}$, k_a , k_h ainsi que les termes $K_{..}$ sont constants.

Les valeurs de ces constantes sont données pour une température de 10°C et de 20°C [5]. Les autres grandeurs varient en fonction du temps. Le terme $k_{La} (S_{O_2}^{sat} - S_{O_2})$ correspond à l'apport d'air extérieur dans la biologie, c'est-à-dire au fonctionnement des aérateurs de la biologie. Les termes en ρ_j avec $j=1...8$, correspondent aux cinétiques du modèle de dégradation et sont explicités dans la table III.

Le modèle obtenu, issu du modèle général ASM1 fait intervenir 11 variables d'état dont de nombreuses qui ne sont pas mesurées. Pour pouvoir appliquer ce modèle à notre situation, il faut le réduire afin qu'il ne prenne en compte que les variables mesurées.

B. Réduction du modèle ASM1

Les hypothèses simplificatrices utilisées sont inspirées des thèses de B. Chachuat [1] et de C. Gomez-Quintero [3].

B.1 Absence de croissance anoxie.

La première hypothèse simplificatrice est l'élimination de la cinétique ρ_2 "croissance anoxie des hétérotrophes". En effet, la station d'épuration ne possède pas de période d'anoxie significative, donc le terme ρ_2 est supprimé des différentes valeurs de la dégradation \mathcal{R}_α (table II).

B.2 Simplification des dynamiques du modèle.

Les variations de la concentration en matière organique soluble inerte (S_I) n'ont aucune influence sur les autres composés et inversement, nous omettons alors ce composé dans le modèle, en supprimant la variable d'état de l'équation (3) correspondant à $\alpha = 1$.

La théorie des perturbations singulières (séparation des dynamiques lentes et des dynamiques rapides) permet de considérer que X_I , X_A et X_H ont des dynamiques lentes. On considère alors que ces grandeurs sont constantes sur quelques heures. On supprime alors les variables d'états de l'équation (3) correspondant à $\alpha = 7, 9$ et 10.

B.3 Simplification des composés organiques.

Le modèle ASM1 établit une distinction entre les composés biodégradables solubles (S_S), supposés directement assimilables par les micro-organismes et particuliers (X_S), qui doivent être préalablement hydrolysés en composés solubles pour pouvoir être assimilés. Cependant, la mesure de la demande chimique en oxygène (DCO), ne permet pas de distinguer la partie soluble et la partie particulaire [11]. On considère alors un composé organique unique : $X_{DCO} = S_S + X_S$

A partir de l'équation (3) avec les variables d'états correspondants à $\alpha = 2$ et 8, puis en additionnant les deux expressions obtenues, X_{DCO} peut alors s'écrire :

$$\begin{aligned} \dot{X}_{DCO} &= D_s X_{DCO}^{in} - D_s S_S - (D_s + D_r) \frac{D_{wa}}{D_r + D_{wa}} X_S \\ &\quad - \frac{1}{Y_H} \rho_1 + (1 - f_{XI}) (\rho_4 + \rho_5) \end{aligned} \quad (5)$$

j	Processus	Cinétique, ρ_j
1	Croissance aérobie des hétérotrophes	$\mu_H \frac{S_S}{S_S + K_S} \frac{S_{O_2}}{S_{O_2} + K_{O_2,H}} \frac{S_{NH}}{S_{NH} + K_{NH,H}} X_H$
2	Croissance anoxie des hétérotrophes	$\mu_H \frac{S_S}{S_S + K_S} \frac{S_{O_2}}{S_{O_2} + K_{O_2,H}} \frac{S_{NH}}{S_{NH} + K_{NH,H}} \frac{S_{NO}}{S_{NO} + K_{NO}} \eta_{NO_{3g}} X_H$
3	Croissance aérobie des autotrophes	$\mu_A \frac{S_{O_2}}{S_{O_2} + K_{O_2,A}} \frac{S_{NH}}{S_{NH} + K_{NH,A}} X_A$
4	Mortalité des hétérotrophes	$b_H X_H$
5	Mortalité des autotrophes	$b_A X_A$
6	Ammonification de l'azote organique soluble	$k_a X_H S_{ND}$
7	Hydrolyse de la matière organique absorbée	$k_h \frac{X_S}{K_X + X_S/X_H} \left[\frac{S_{O_2}}{S_{O_2} + K_{O_2,H}} + \eta_{NO_{3h}} \frac{K_{O_2,H}}{S_{O_2} + K_{O_2,H}} \frac{S_{NO}}{S_{NO} + K_{NO}} \right]$
8	Hydrolyse de l'azote organique absorbée	$k_h \frac{X_{ND}}{K_X + X_S/X_H} \left[\frac{S_{O_2}}{S_{O_2} + K_{O_2,H}} + \eta_{NO_{3h}} \frac{K_{O_2,H}}{S_{O_2} + K_{O_2,H}} \frac{S_{NO}}{S_{NO} + K_{NO}} \right]$

TABLE III
CINÉTIQUES DU MODÈLE DE DÉGRADATION

Dans la station, le débit de recirculation est supérieur au débit d'extraction des boues, on suppose alors que $\frac{D_{wa}}{D_r + D_{wa}} \ll 1$, on néglige donc le terme X_S . On définit ensuite la fraction soluble des composés organiques par $f_{SS} = \frac{S_S}{X_{DCO}}$. L'équation (5) s'écrit alors :

$$\dot{X}_{DCO} = D_s (X_{DCO}^{in} - f_{SS} X_{DCO}) - \frac{1}{Y_H} \rho_1 + (1 - f_{XI}) (\rho_4 + \rho_5) \quad (6)$$

Avec ce changement de variable, on remarque que les cinétiques des processus (table III) changent, en effet :

$$\frac{S_S}{S_S + K_S} = \frac{X_{DCO}}{X_{DCO} + K_{DCO}} \quad \text{où } K_{DCO} = \frac{K_S}{f_{SS}} \quad (7)$$

B.4 Simplification des composés azotés.

Le modèle ASM1 distingue quatre fractions azotées S_{NO} , S_{NH} , S_{ND} et X_{ND} . Dans la mesure où les concentrations en azote ammoniacal S_{NH} et en nitrate/nitrite S_{NO} constituent l'essentiel des rejets d'azote total et sont mesurables en ligne, il n'apparaît pas concevable de supprimer ces composés du modèle réduit. En revanche, les deux fractions d'azote organique soluble S_{ND} et particulaire X_{ND} , dont le rôle est de décrire la formation interne de S_{NH} par les processus d'hydrolyse et d'ammonification, ne constituent qu'une faible partie des rejets d'azote. On pose alors comme hypothèse que le processus d'ammonification de l'azote organique soluble (cinétique ρ_6) est constant. La concentration S_{ND} est considérée constante, la modélisation de X_{ND} est alors inutile, car cette concentration est seulement utilisée dans l'estimation de S_{ND} .

Avec cette hypothèse, on considère la formation interne d'ammoniaque constant. Cela induit de larges erreurs sur la concentration en ammoniacque [1], notamment lorsque les périodes d'anoxie sont longues. Dans notre cas, nous n'avons pas de périodes d'anoxie, nous allons donc utiliser cette hypothèse car elle permet de diminuer significativement le nombre d'inconnues.

On suppose également que les concentrations, à l'entrée de la station, en oxygène dissous sont négligeables devant les

concentrations dans le bassin ($S_{O_2}^{in} = 0$). On trouve alors, à partir des équations (3, 6, 7), comme le modèle réduit suivant :

$$\left\{ \begin{array}{l} \dot{X}_{DCO} = D_s \left(X_{DCO}^{in} - \frac{K_S}{K_{DCO}} X_{DCO} \right) - \frac{1}{Y_H} \theta_1 \\ \quad \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \frac{S_{O_2}}{K_{O_2,H} + S_{O_2}} \frac{S_{NH}}{S_{NH} + K_{NH,H}} + \theta_2 \\ \dot{S}_{NO} = D_s (S_{NO}^{in} - S_{NO}) + \frac{\theta_3}{Y_A} \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_{O_2}}{K_{O_2,A} + S_{O_2}} \\ \dot{S}_{NH} = D_s (S_{NH}^{in} - S_{NH}) - i_{NBM} \theta_1 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \\ \quad \frac{S_{O_2}}{K_{O_2,H} + S_{O_2}} \frac{S_{NH}}{S_{NH} + K_{NH,H}} - \left(i_{NBM} + \frac{1}{Y_A} \right) \theta_3 \\ \quad \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_{O_2}}{K_{O_2,A} + S_{O_2}} + \theta_4 \\ \dot{S}_{O_2} = -D_s S_{O_2} - \frac{1 - Y_H}{Y_H} \theta_1 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \\ \quad \frac{S_{O_2}}{K_{O_2,H} + S_{O_2}} \frac{S_{NH}}{S_{NH} + K_{NH,H}} - \frac{4,57 - Y_A}{Y_A} \theta_3 \\ \quad \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_{O_2}}{K_{O_2,A} + S_{O_2}} + k_{La} (S_{O_2}^{sat} - S_{O_2}) \end{array} \right. \quad (8)$$

avec comme constantes :

$$\left\{ \begin{array}{l} D_s = \frac{Q_S}{V} \\ \theta_1 = \mu_H X_H \\ \theta_2 = (1 - f_{XI}) (b_H X_H + b_A X_A) \\ \theta_3 = \mu_A X_A \\ \theta_4 = k_a X_H S_{ND} \end{array} \right. \quad (9)$$

Le modèle ainsi obtenu ne dépend maintenant que de neuf variables : D_s , X_{DCO}^{in} , X_{DCO} , S_{NO} , S_{NO}^{in} , S_{NH}^{in} , S_{NH} , S_{O_2} et du fonctionnement des aérateurs. Dans un premier temps, dans un souci de simplicité, on essaye de trouver des relations linéaires.

C. Linéarisation du modèle réduit

Nous linéarisons le modèle (8) en nous basant sur les considérations suivantes :

– Dans notre cas, la biologie 1 ne possède pas de période anoxie, on suppose alors que la concentration en oxygène

dissous est importante, c'est-à-dire que $S_{O_2} \gg K_{O_2,H}$ et donc $\frac{S_{O_2}}{S_{O_2} + K_{O_2,H}} \approx 1$; on utilise la même hypothèse avec $K_{O_2,A}$, donc $\frac{S_{O_2}}{S_{O_2} + K_{O_2,A}} \approx 1$.

- Pour linéariser les équations (8) par rapport aux variables d'états (X_{DCO} , S_{NH} et S_{NO}), on simplifie :
 - $\frac{X_{DCO}}{K_{DCO} + X_{DCO}} \rightarrow a_{DCO}X_{DCO} + b_{DCO}$ où a_{DCO} et b_{DCO} sont des constantes.
 - $\frac{S_{NH}}{K_{NH,A} + S_{NH}} \rightarrow a_{NH,A}S_{NH} + b_{NH,A}$ où $a_{NH,A}$ et $b_{NH,A}$ sont des constantes.
 - $\frac{S_{NH}}{K_{NH,H} + S_{NH}} \rightarrow a_{NH,H}S_{NH} + b_{NH,H}$ où $a_{NH,H}$ et $b_{NH,H}$ sont des constantes.
 - $\frac{S_{NO}}{K_{NO} + S_{NO}} \rightarrow a_{NO}S_{NO} + b_{NO}$ où a_{NO} et b_{NO} sont des constantes.
- En plus, les produits de variables d'état sont aussi remplacés par des termes de type $S_x S_y = \bar{S}_y S_x + \bar{S}_x S_y$ avec \bar{S}_x et \bar{S}_y des valeurs moyennes constantes des variables S_x et S_y .

A partir du système d'équation (8) et compte tenu de ces hypothèses, on trouve alors le système d'équations (10) :

$$\left\{ \begin{array}{l} \dot{X}_{DCO}(t) = D_S(t) \left(X_{DCO}^{in}(t) - \frac{K_S}{K_{DCO}} X_{DCO}(t) \right) \\ \quad - \frac{1}{Y_H} \theta_1 (a_{DCO} X_{DCO}(t) + b_{DCO}) \\ \quad (a_{NH,H} S_{NH}(t) + b_{NH,H}) + \theta_2 \\ \dot{S}_{NO}(t) = D_S(t) (S_{NO}^{in}(t) - S_{NO}(t)) \\ \quad + \frac{\theta_2}{Y_A} a_{NH,A} S_{NH}(t) + \frac{\theta_2}{Y_A} b_{NH,A} \\ \dot{S}_{NH}(t) = D_S(t) (S_{NH}^{in}(t) - S_{NH}(t)) \\ \quad - i_{NBM} \theta_1 (a_{DCO} X_{DCO}(t) + b_{DCO}) \\ \quad (a_{NH,H} S_{NH}(t) + b_{NH,H}) - \left(i_{NBM} + \frac{1}{Y_A} \right) \\ \quad \theta_3 (a_{NH,A} S_{NH}(t) + b_{NH,A}) + \theta_4 \\ \dot{S}_{O_2}(t) = -D_S(t) S_{O_2}(t) - \frac{1-Y_H}{Y_H} \theta_1 (a_{DCO} X_{DCO}(t) \\ \quad + b_{DCO}) (a_{NH,H} S_{NH}(t) + b_{NH,H}) \\ \quad - \frac{4.57-Y_A}{Y_A} \theta_3 (a_{NH,A} S_{NH}(t) + b_{NH,A}) \\ \quad + k_L a(t) (S_{O_2}^{sat} - S_{O_2}(t)) \end{array} \right. \quad (10)$$

On trouve alors un système linéaire à paramètres variants, en effet le taux de dilution relatif au débit incident ($D_S(t)$) et le fonctionnement des aérateurs ($k_L a(t) (S_{O_2}^{sat} - S_{O_2}(t))$) dépendent du temps.

La discrétisation du modèle (10) d'un point de vue numérique donne alors avec un pas d'échantillonnage de Δ :

$$\left\{ \begin{array}{l} X_{DCO}(k+1) = \Delta (D_S(k) X_{DCO}^{in}(k) \\ \quad + A_{DCO}(k) X_{DCO}(k) + B_{NH} S_{NH}(k) + k_1) \\ S_{NO}(k+1) = \Delta (C_{NO}(k) S_{NO}(k) + D_{NH} S_{NH}(k) + k_2) \\ S_{NH}(k+1) = \Delta (D_S(k) S_{NH}^{in}(k) \\ \quad + E_{NH}(k) S_{NH}(k) + F_{DCO} X_{DCO}(k) + k_3) \\ S_{O_2}(k+1) = \Delta (G_{O_2}(k) S_{O_2}(k) + H_{NH} S_{NH}(k) \\ \quad + I_{DCO} X_{DCO}(k) + k_L a(k) S_{O_2}^{sat} + k_4) \end{array} \right. \quad (11)$$

où les paramètres $A_{DCO}(k)$, B_{NH} , $C_{NO}(k)$, D_{NH} , $E_{NH}(k)$, F_{DCO} , $G_{O_2}(k)$, H_{NH} , I_{DCO} , k_1 , k_2 , k_3 et k_4 sont identifiés en fonction du système d'équations (10), comme par exemple $G_{O_2}(k) = \frac{1}{\Delta} - D_S(k) - k_L a(k)$.

Après avoir déterminé ce modèle réduit, il faut faire le lien entre les grandeurs du modèle et les capteurs de la station. Les grandeurs S_{NH} et S_{NO} correspondent aux concentrations en ammonium et en nitrate dans la biologie, or nous ne connaissons que la mesure de ces deux grandeurs à la sortie de la biologie. Pour prendre en compte cette différence, il faut ajouter des retards sur les mesures disponibles.

On ne dispose pas des mêmes informations pour la modélisation de la première biologie et la modélisation de la seconde biologie. La table IV résume les relations potentielles entre les inconnues du modèle et les mesures disponibles pour chaque biologie.

Pour la première biologie, on suppose que l'eau en entrée de la station possède une concentration négligeable de nitrate ($S_{NO}^{in} = 0$). Pour estimer X_{DCO}^{in} on peut utiliser la mesure de l'absorption UV qui donne une idée de la quantité de particules contenue dans l'eau. L'ammoniaque influence fortement la conductivité, on peut donc probablement lier S_{NH}^{in} et la mesure de conductivité [10]. On peut essayer d'utiliser la mesure de Matières en Suspension présente dans le bassin pour faire un lien avec X_{DCO} .

Pour la deuxième biologie, la qualité des eaux en entrée est mieux connue. Pour estimer X_{DCO}^{in} , on peut utiliser la mesure de Matières en Suspension présente dans la première biologie. On peut essayer, comme dans le cas de la première biologie, d'utiliser les mesures de Matières en Suspension présent dans les bassins pour établir un lien avec X_{DCO} .

	1 ^{ère} biologie	2 ^{nde} biologie
D_S	Débit	Débit
X_{DCO}^{in}	Absorption UV	MES1
X_{DCO}	MES1	MES21 et MES22
S_{NO}^{in}	0	Nitrate1
S_{NO}	Nitrate1	Nitrate2
S_{NH}^{in}	Conductivité	Ammonium1
S_{NH}	Ammonium1	Ammonium2
S_{O_2}	Oxygène1	OxygèneB21 et OxygèneB22

TABLE IV
MESURES POUR COMPLÉTER LE MODÈLE

La réduction de dimension et de complexité du modèle ASM1, nous a donc permis de déterminer les structures des relations de redondance, mais ces redondances ne concernent que les relations propres à chaque biologie. Pour à la fois compléter les liens entre les variables des biologies et obtenir des relations de redondance liant les variables des deux biologies ensemble, nous allons maintenant analyser les corrélations statistiques entre les signaux.

IV. LE MODÈLE BOÎTE NOIRE

Pour construire un modèle boîte noire par régression linéaire multiple, nous avons choisi la période du 18 mars 2004 et jusqu'au 6 avril 2004; pour la création du modèle les données du 18 mars au 28 mars sont utilisées, et pour la validation celles du 29 mars au 6 avril 2004. Nous avons choisi cette période en raison du peu de valeurs manquantes ou invalides et d'une température de l'eau variant peu permettant ainsi de

Var. expliquée	Crit.	Corr.	Var. explicatives
1	0.087	0.925	11 55
2	0.067	0.773	8 3 7 4 55
3	0.163	0.658	2 8 4 55
4	0.232	0.523	2 8 3 55
5	×	×	
6	0.132	0.360	11 1 55
7	0.266	0.798	2 8 9 16 14 55
8	0.307	0.882	2 10 7 16 55
9	0.333	0.825	7 8 16 10 55
10	0.479	0.763	8 16 9 2 55
11	0.049	0.925	1 55
12	0.154	0.973	14 16 55
13	0.042	0.976	15 55
14	0.165	0.973	12 7 55
15	0.031	0.976	13 58
16	0.340	0.852	18 8 9 7 12 10 55
17	0.208	0.258	1 55
18	0.367	0.668	16 55

TABLE V

VARIABLES EXPLICATIVES DANS LE CAS STATIQUE

s'affranchir de cette perturbation. Nous disposons d'une mesure toutes les 30 secondes, c'est-à-dire de 120 mesures par heure. Après avoir complété les zones où les capteurs sont en défaut par des valeurs issues des jours précédents où les comportements sont semblables, nous avons étudié les relations entre les variables.

A. Modèle statique

Dans un premier temps, les variables expliquées (table V) sont déterminées par régression linéaire multiple en ajoutant les variables les plus explicatives en premier (la variable 55 correspond à une constante). Si on appelle $y_i(t)$ la variable expliquée, $u_j(t)$ les variables explicatives et θ_j les paramètres associés aux variables explicatives, on trouve alors un modèle linéaire de la forme :

$$y_i(t) = \sum_{j=1}^p \theta_j u_j(t) \quad (12)$$

où p représente le nombre de variable explicatives.

Les estimations des paramètres $\hat{\theta}_j$ et de la mesure $\hat{y}_i(t)$ sont déterminées par minimisation du critère quadratique correspondant à la somme des carrés des écarts entre la mesure et l'estimation.

La table V précise le coefficient de corrélation de chaque variable estimée $\hat{y}_i(t)$ vis-à-vis de sa mesure $y_i(t)$. La seconde colonne "Crit." correspond à la valeur numérique du critère ϕ défini par l'équation (13). Notons que ces calculs ont été effectués sur les données n'ayant pas servi à l'identification.

$$\phi = \frac{\|y_i(t) - \hat{y}_i(t)\|}{\|\hat{y}_i(t)\|} \quad (13)$$

On remarque alors que dans le cas de la mesure du PH (5), il n'est pas possible de construire un modèle, en raison d'une

Var. expliquée	Crit.	Corr.	Var. explicatives
1	0.083	0.929	11 19 55
2	0.030	0.926	20 8 26 3 21 55
3	0.123	0.803	21 2 20 8 4 55
4	0.212	0.590	2 20 22 8 3 55
5	0.035	0.028	25 55
6	0.022	0.955	24 19 29 10 55
7	0.214	0.780	25 20 26 2 55
8	0.089	0.987	26 44 20 2 55
9	0.149	0.958	27 25 7 26 8 55
10	0.247	0.944	28 2 8 55
11	0.046	0.934	1 19 29 55
12	0.123	0.982	14 30 32 55
13	0.015	0.992	15 31 33 55
14	0.113	0.987	12 32 30 55
15	0.015	0.991	13 33 31 55
16	0.058	0.996	34 52 44 26 55
17	0.026	0.986	35 53 55
18	0.039	0.997	36 54 52 34 16 55

TABLE VI

VARIABLES EXPLICATIVES DANS LE CAS DYNAMIQUE

faible corrélation avec les autres grandeurs. A l'opposé, le débit (1) est expliquée seulement à partir de la mesure de la hauteur de la surverse (11) et vis-versa, ce qui traduit une forte corrélation entre les deux mesures.

Si on rapproche le modèle de connaissance de la seconde biologie et le modèle statique, on remarque que l'on trouve les mêmes types de relations. En effet on peut expliquer l'ammonium à la sortie de la deuxième biologie (16) à partir de la mesure d'ammonium à la sortie de la première biologie (8), qui correspond à la concentration d'ammonium à l'entrée de la station (S_{NH}^{in}) dans le modèle de connaissance, les autres termes de l'équation correspondent à l'estimation de la demande chimique en oxygène (X_{DCO}), en effet les mesures d'orthophosphates (18 et 10), de nitrate (9) et d'oxygène dissous (7 et 12) donnent une indication de la charge de polluants, ce qui est liée à la Demande Chimique en Oxygène.

B. Modèle dynamique

En ajoutant un décalage temporel à certaines variables, on peut améliorer la qualité de la modélisation. On considère, en plus, les mêmes variables décalées de 120 points et de 240 points, c'est-à-dire un décalage de 1h et de 2h, qui correspondent approximativement aux temps de séjour des eaux usées dans la première et dans la deuxième biologie. La table VI présente les mêmes informations que dans la table V mais dans le cas dynamique. La signification des numéros est explicitée dans la table VII.

Le modèle déterminé pour le nitrate1 est alors :

$$\begin{aligned} \text{nitrate1}(k) = & 0.3792 \text{oxygene1}(k) \\ & - 0.0280 \text{ammonium1}(k) \\ & - 0.0083 \text{oxygene1}(k - 120) \\ & + 0.0223 \text{ammonium1}(k - 120) \\ & + 0.7555 \text{nitrate1}(k - 120) + 0.2256 \end{aligned} \quad (14)$$

La figure 3 présente l'estimation de la concentration de nitrate1 avec le modèle statique ainsi que la différence entre son estimation et sa mesure, cette différence est nommée résidu. La figure 4 montre l'estimation et le résidu de cette même variable mais avec le modèle dynamique. Ces deux estimations sont réalisées sur un jeu de validation, on remarque une meilleure estimation dans le cas dynamique. Les résidus ainsi déterminés peuvent être utilisés pour effectuer le diagnostic de la station.

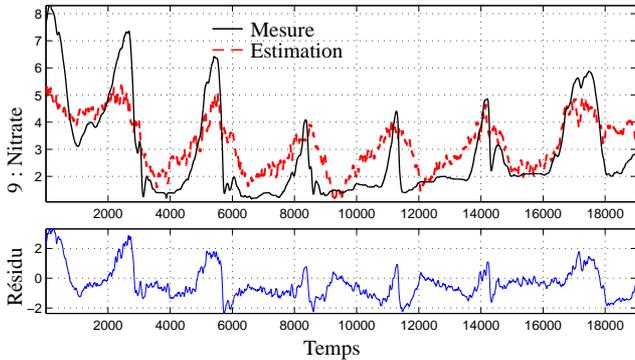


Fig. 3

ESTIMATION DE LA CONCENTRATION DE NITRATE1 AVEC LE MODÈLE STATIQUE

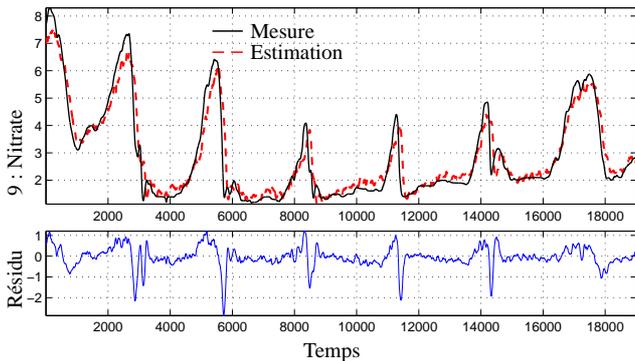


Fig. 4

ESTIMATION DE LA CONCENTRATION DE NITRATE1 AVEC LE MODÈLE DYNAMIQUE

C. Utilisation des modèles pour le diagnostic

La table VII représente les occurrences des différentes variables dans les modèles obtenus précédemment ; ces occurrences sont notifiées avec des cercles dans le cas du modèle statique et avec des croix dans le cas du modèle dynamique. Cette table d'occurrence permet de mettre en évidence la possibilité de localiser les défauts de capteurs. En effet, cette table représente les variables dont dépendent les résidus de chaque mesure.

Par exemple, si un défaut apparaît sur la mesure de la température (2), alors les tests de détection de défauts des résidus 3, 4, 7, 8 et 10 doivent être positifs, on peut ainsi localiser le défaut. Cette table permet aussi de mettre en évidence l'impossibilité de dissocier un défaut sur le capteur de MES B21

(13) et sur le capteur de MES B22 (15), car dans les deux cas les tests de détection de défauts des résidus 13 et 15 doivent être positifs, il faut alors avoir une information supplémentaire pour localiser le défaut. On se trouve dans le même cas pour localiser un défaut apparaissant sur le capteur d'Oxygène B21 (12) et celui d'oxygène B22 (14).

On remarque aussi que dans le cas dynamique, on estime la mesure de nitrate2 (17) seulement à partir de cette mesure à des instants antérieurs. Du point de vue du diagnostic, estimer une grandeur à partir de sa valeur aux instants précédents ne pose pas de problème. En effet, si un défaut apparaît entre une mesure et cette même mesure à un instant précédent alors la relation liant les deux grandeurs n'est plus vérifiée, on peut alors détecter un défaut.

V. CONCLUSION

Pour construire des modèles en vue du diagnostic de fonctionnement, la première étape est la compréhension du fonctionnement du système afin de dégager les liens entre les variables mesurées. Dans l'optique de mettre en évidence des relations de redondance linéaires entre les variables, nous avons utilisé le modèle ASM1 pour modéliser les deux biologies. Ne possédant pas les mesures de toutes les variables intervenant dans ce modèle, il a fallu d'abord le réduire en posant des hypothèses réalistes et communément acceptées. Dans le but de déterminer des relations les plus simples possibles, nous l'avons linéarisé. Nous avons ainsi déterminé la structure des relations entre les variables.

Après avoir trouvé la structure des relations entre les variables décrivant le fonctionnement des biologies, nous nous sommes ensuite intéressés à la détermination des relations de redondances entre les variables de la station, par régression linéaire multiple. Cette régression a permis, en ajoutant successivement les variables les plus explicatives, de construire des modèles du type boîte noire. Il faut cependant prendre en compte, dans la modélisation, la nécessité de localiser un défaut lorsqu'il est détecté. La dernière partie de cet article aborde ce problème avec la table d'occurrence (table VII) qui met en évidence les défauts localisables et non localisables.

REMERCIEMENTS

Les auteurs remercient sincèrement le Ministère de la Culture, de l'Enseignement Supérieur et de la Recherche du gouvernement luxembourgeois pour leur participation au financement de ces travaux.

RÉFÉRENCES

- [1] B. Chachuat. *Méthodologie d'optimisation dynamique et de commande optimale des petites stations d'épuration à boues activées*. Thèse de doctorat, Institut National Polytechnique de Lorraine, 2001.
- [2] K. V. Gernaey, M. C. M. van Loosdrecht, M. Henze, M. Lind, and S. B. Jorgensen. Activated sludge wastewater treatment plant modelling and simulation : state of the art. *Environmental Modelling & Software*, 19(9), pages 763–783, 2004.
- [3] C. Gomez-Quintero. *Modélisation et estimation robuste pour un procédé boues activées en alternance de phases*. Thèse de doctorat, Université Toulouse III, 2002.
- [4] W. Gujer, M. Henze, T. Mino, and M. van Loosdrecht. Activated Sludge Model No. 3. *Water Science and Technology*, 39(1), pages 183–193, 1999.

