Soft-computing et observateurs appliqués à des systèmes environnementaux

Frédéric Lafont

Université du Sud-Toulon-Var
Laboratoire des Sciences de l'Information et des Systèmes - UMR CNRS 6168

Habilitation à Diriger des Recherches

spécialité « Automatique »

par

Frédéric Lafont

Soft-computing et observateurs appliqués à des systèmes environnementaux

soutenue le 05/12/2011 devant le jury composé de :

Ugo Boscain	Directeur de recherche à l'Ecole Polytechnique	(Rapporteur)
Boutaib Dahhou	Professeur à l'Université Paul Sabatier	(Rapporteur)
Hassan Hammouri	Professeur à l'Université Lyon 1	(Rapporteur)
Eric Busvelle	Professeur à l'Université de Bourgogne	(Examinateur)
Michel Gauch	Professeur à l'Université Paul Cézanne	(Examinateur)
Mustapha Ouladsine	Professeur à l'Université Paul Cézanne	(Examinateur)
Jean-Paul Gauthier	Professeur à l'Université du Sud-Toulon-Var	(Examinateur)

Remerciements

Avant tout, je tiens à remercier tous les membres du jury pour avoir accepté de consacrer une partie de leur temps à la lecture de ce rapport.

Je remercie évidemment plus particulièrement Jean-Paul Gauthier qui m'a fait l'honneur de me diriger pendant l'essentiel des travaux que je présente ici.

Je remercie tous les membres du Laboratoire des Sciences de l'Information et des Systèmes. C'est dans ce laboratoire que j'ai réalisé l'essentiel du travail présenté ici.

Je remercie plus spécifiquement, l'équipe Estimation-COmmande-DIagnostic (ESCODI) du laboratoire, Jean-françois Balmat, Jean Duplaix, Nathalie Pessel. Leurs compétences, leurs conseils et leur bonne humeur ont fait que j'ai pu mener à bien ces travaux.

Merci aussi à Gilles Enea qui a toujours eu de trés bons conseils.

La Garde, le 8 novembre 2011.

Table des matières

Préface

L'objectif de ce manuscrit est de présenter mes travaux de recherche et d'animations scientifiques en vue de l'obtention de l'Habilitation à Diriger des Recherches. J'ai décidé de décomposer ce mémoire en plusieurs parties.

Le *premier chapitre* expose la présentation générale de mon dossier.
Ensuite, j'ai décidé de décomposer et développer mes activités de recherche en trois parties. Le *deuxième chapitre* présente mes contributions sur le plan de la commande et de l'identification à base de techniques de Soft-computing.
Le *troisième chapitre* met en évidence une étude industrielle développée en collaboration avec la DCNS.
Le *quatrième chapitre* propose deux études sur des observateurs : observateur grand gain adaptatif et observateur à entrées inconnues.

Ce manuscrit finit par une conclusion et une annexe avec certaines publications qui me paraissent intéressantes.

Présentation générale

1

Sommaire

Ce chapitre introductif est destiné à faire une présentation générale de mes activités de recherche, pédagogique, d'encadrement et enfin de responsabilités administratives.

1.1 Curriculum Vitae

Nom patronymique : LAFONT
Prénom : Frédéric
Date de naissance : 06 Mars 1969
Grade : MCF 6^{ième} échelon
Section CNU : 61
Tél. : 04 94 14 95 03
mèl : lafont@univ-tln.fr
Adresse professionnelle : Université du Sud-Toulon-Var, Laboratoire des Sciences de l'Information et des Systèmes (LSIS), UMR CNRS 6168, BP 20132, 83957 La Garde Cedex
Adresse personnelle : 12 lot les clémentines, 83220 Le Pradet

J'ai préparé ma thèse au Laboratoire d'Automatique et d'Informatique Appliquée de Toulon (LAIAT). Le travail, que j'ai effectué sous la direction du Professeur Patrick Abellard, a porté sur deux domaines : la nature du bruit sur des images radiographiques et la définition des réseaux de Petri à flux de données stochastiques. J'ai soutenu ma **Thèse de doctorat de l'Université de Toulon et du Var** le *03 décembre 1997* devant le jury composé de :

Arques Pierre-Yves	Professeur	Président
Delarque Alain	Professeur	Rapporteur
Ferrand Daniel	Professeur	Rapporteur
Conil jean-Louis	Industriel	Examinateur
Duplaix Jean	Maître de conférences	Examinateur
Abellard patrick	Professeur	Directeur

La première partie de la thèse présente les résultats concernant la modélisation du rachis à partir d'images stéréoradiographiques. Le principe de base de cette étude repose sur l'analyse du bruit contenu sur une image radiographique, puis sur la réduction de ce bruit, afin d'extraire des contours pour modéliser les vertèbres en vue d'une CAO.
Les algorithmes utilisés imposant des temps de traitements importants, nous avons étudié leur parallélisation par une modélisation de type Réseaux de Petri à Flux de Données. Dans cette seconde partie, nous nous sommes intéressés à l'ordonnancement de ces réseaux. A cet effet, l'existence de la variation du chemin critique est apparue en fonction du processeur traitant l'information. Nous avons donc défini une extension des Réseaux de Petri à Flux de Données : les Réseaux de Petri à Flux de Données Stochastiques. De plus, une méthode d'ordonnancement de ces Réseaux de Petri à Flux de Données Stochastiques, basée sur le maximum de vraisemblance, a été développée pour une implémentation des calculs sur une architecture parallèle de type multiprocesseur à flux de données.

Le financement a été assuré par une Bourse Doctorale Régionale PACA (1994-1996), puis un poste d'Attaché Temporaire d'Enseignement et de Recherche à l'IUT de Toulon au département Services et Réseaux de communication (Sept. 1996- Sept. 1997) et enfin par un poste de Contractuel

sur un poste de Prag dans ce même département (Sept. 1997 - Sept. 1998).

Depuis le 1er septembre 1998, je suis **Maître de Conférences** à l'IUT de Toulon au département Génie Industriel et Maintenance de l'Université du Sud Toulon-Var. J'effectue actuellement mon travail de recherche au sein du Laboratoire des Sciences de l'Information et des Systèmes, LSIS, UMR-CNRS 6168.
Je suis titulaire de la **Prime d'Encadrement Doctoral et de Recherche** depuis 1er octobre 2008.
La suite de ce chapitre présente très brièvement mes activités de recherche (page 6), d'enseignement et d'encadrements (page 10) et administratives et coopérations industrielles (page 11).

L'essentiel du dossier consiste en une présentation plus approfondie de mes activités de recherche qui sont détaillées aux chapitres 2, 3 et 4, suivie d'une sélection de sept articles représentatifs : Lafont et Balmat (2002), Trabelsi et al. (2007), Pessel et al. (2009), Balmat et al. (2009), , Balmat et al. (2011), Lafont et al. (2011) et Methnani et al. (2011).

1.2 Recherche

J'ai fais le choix de décomposer mes travaux de recherche en 3 parties :
1 La modélisation et la commande à base de soft-computing
2 Un module d'aide à la décision pour l'évaluation des risques
3 Les observateurs grand gain adaptatif et à entrées inconnues

1.2.1 La modélisation et la commande à base de soft-computing

Pendant les deux dernières décennies, un grand effort a été consacré pour le climat interne des serres agricoles. Une serre agricole est un système complexe dans lequel interviennent plusieurs échanges énergétiques et fonctions biologiques assurant le développement des cultures.
La commande d'une serre agricole a pour but de créer un microclimat favorable à une culture donnée. En effet, chaque culture a besoin de conditions climatologiques particulières (température, hygrométrie, ...). L'élaboration de ces contrôleurs nécessite une connaissance a priori du système serre. Depuis longtemps, le laboratoire LSIS s'est intéressé à ce problème et des techniques classiques de régulation telles que la commande adaptative multivariable, la commande optimale, la commande booléenne ont été développées. Les interactions entre les variables internes et externes, et la complexité des phénomènes (multivariable, non linéaire, non stationnaire) sont telles qu'il est souvent difficile d'implémenter les techniques conventionelles de régulation. De plus, ces méthodes, qui supposent des simplifications, sont souvent très sensibles aux perturbations qui n'apparaissent pas dans le modèle.
Pour résoudre ces problèmes, nous avons proposé une approche différente en utilisant des techniques basées sur la connaissance experte. Nous nous sommes naturellement intéressés au contrôle flou. En effet, la théorie de la logique floue, développée par Zadeh, apparait bien adaptée à la complexité du modèle de processus. L'avantage principal de ces techniques est

qu'il n'est pas nécessaire de définir un modèle de processus, ce qui facilite grandement l'implémentation du régulateur. Nous avons développé un régulateur flou « optimisé » en décentralisant et hiérarchisant les entrées pour réduire le nombre de règles. Les résultats de ce régulateur sont comparés ceux d'un régulateur flou basique (Lafont et Balmat 2002).

Dans un deuxième temps, afin d'améliorer les résultats obtenus par différentes commandes, nous avons décidé de travailler sur l'identification et la modélisation à base de soft-computing. Dans la littérature, les techniques pour la modélisation mathématique des processus sont classifiées en trois principales catégories : la modélisation par modèle de connaissance, la modélisation par modèle de représentation et la modélisation qui combine ces deux types. La première est basée sur les lois physiques régissant le processus et la deuxième est basée sur l'analyse des données d'entrées-sorties du processus.
D'autres théories de modélisation sont apparues telles que celles qui utilisent les concepts des réseaux de neurones et la logique floue. Ces deux théories permettent la description de la dynamique des systèmes complexes (non linéaires, de grande dimension, ...) de manière satisfaisante et peuvent être employées aussi bien dans un cadre déterministe que stochastique. Parmi les méthodes pour l'identification des modèles flous, nous trouvons celles qui peuvent générer automatiquement des règles floues à partir des données réelles et qui optimisent les paramètres de ces règles floues par différentes techniques telles que les réseaux de neurones, les algorithmes génétiques, les moindres carrés récursifs, ...
Les modèles flous de type Takagi-Sugeno permettent d'obtenir des sous-modèles linéaires dans tout l'espace d'entrée-sortie du processus, d'où la possibilité d'appliquer les théories de l'automatique classique pour développer une loi de commande qui satisfait les objectifs visés.
Nous avons proposé une identification à base de clusters flous où les prémisses des règles sont adaptés par l'algorithme récursif des moindres carrés avec facteur d'oubli (Trabelsi et al. 2007). Nous appliquons cet algorithme au système serre agricole.

Dans une dernière étude et sur un plan plus théorique, nous avons défini une méthodologie pour une structure multi-modèle. La structure de chaque modèle peut varier. La méthodologie rassemble l'analyse en composantes principales, les réseaux de neurones et la logique floue. En fonction des résultats de l'analyse en composantes principales, plusieurs structures de modèles sont retenues. Ces modèles sont identifiés à partir de réseaux de neurones. Enfin, un classifieur flou permet soit de commuter entre les différents modèles, soit de fusionner ceux-ci. Cette méthodologie a été appliquée au système serre agricole (Pessel et al. 2009).

1.2.2 Un module d'aide à la décision pour l'évaluation des risques et menaces maritimes

Cette étude est l'objet d'un travail de Recherche et Développement (R&D) avec la Direction des Constructions Navales Systèmes (DCNS) sur une période de deux ans. Le travail effectué concernait l'étude de faisa-

bilité d'un module d'aide à la décision pour l'évaluation des risques et menaces maritimes. Dans ce contexte, cette étude s'est intéressée plus particulièrement à l'apport des techniques neuronales et floues.
Les opérations de sauvegarde maritime comprennent de nombreux aspects :

- recherche et sauvegarde maritime,
- surveillance de la navigation maritime,
- surveillance des pêches,
- surveillance des pollutions marines.

Dans un premier temps, l'étude s'est focalisée sur un thème plus spécifique tel que les risques liés à la pollution. Cependant, l'approche proposée est suffisamment adaptable et évolutive pour permettre de traiter d'autres aspects de l'évaluation des risques.

La réalisation du module d'aide à la décision s'appuie sur un cadre formel basé sur une méthodologie générale permettant :

- d'identifier les risques pour une situation et pour chaque navire,
- de quantifier le niveau de ces risques,
- de définir un indicateur de risque pour chaque navire.

Ce module d'évaluation des risques et menaces a été développé en utilisant les techniques de la logique floue et des réseaux de neurones. Ces deux approches, qui peuvent être complémentaires, sont basées sur la modélisation de l'expertise. Pour l'approche floue on s'intéresse aux capacités d'explication des experts, alors que pour l'approche neuronale c'est au savoir-faire que l'on s'attache (Balmat et al. 2009), (Balmat et al. 2011).

1.2.3 Les observateurs grand gain adaptatif et à entrées inconnues

Ce travail traite du problème de structure d'observateur de systèmes dynamiques non linéaires avec une application à un système de traitement des eaux usées.
L'intérêt de développer des observateurs ou « capteurs logiciels » pour le traitement des boues activées pour le suivi en ligne porte principalement sur deux points :

- bien que des capteurs pour la mesure de variables chimiques et biologiques existent, certaines mesures sont toujours peu fiables et bruitées,
- les coûts d'implémentation et de maintenance des capteurs existants sont élevés.

Beaucoup d'articles sont relatifs à la synthèse d'observateurs non linéaires pour des procédés (bio)chimiques. Nous avons travaillé sur un observateur grand gain adaptatif. Cet observateur est, à la fois, grand gain, et aussi basé sur le filtre de Kalman étendu. Dans le cas de grandes variations, c'est un observateur grand gain (HG) qui garantit la convergence théorique globale avec une vitesse arbitraire, sous certaines hypothèses d'observabilité. Pour des petites erreurs d'estimation initiales,

l'algorithme se comporte comme un filtre de Kalman étendu classique (EKF) pour être plus ou moins optimal par rapport au bruit. La transition du mode HG au mode EKF s'effectue via une procédure d'adaptation basée sur l'« innovation » (niveau d'information relatif aux observations récentes).
L'EKF est largement utilisé et donne plutôt de bons résultats en pratique. L'inconvénient pour l'algorithme EKF est qu'il ne fonctionne que pour des conditions initiales bonnes. Au contraire, l'algorithme HG-EKF converge indépendamment des conditions initiales mais il est plus sensible vis à vis du bruit. Alors, l'idée est de « switcher » entre l'algorithme EKF et HG-EKF. Si l'erreur d'estimation du HG-EKF devient suffisamment faible alors l'EKF est utilisé. La commutation entre ces deux modes est donnée par le paramètre grand gain θ qui varie entre 1 et θ_{max}. L'adaptation est faite en utilisant une équation différentielle contrôlée par l'« innovation ». Habituellement, cette méthode est appliquée au préalable avec un changement de coordonnées dans le but de mettre le système sous forme canonique d'observabilité. Dans notre cas, ce changement de coordonnées est assez compliqué. Pour remédier à cela, nous préférons écrire notre observateur dans les coordonnées naturelles. La contrepartie de ce choix est que l'équation de Riccati du filtre de Kalman n'est pas sous une forme standard.
De plus, ici, dans le but de simplifier les calculs, nous utilisons deux observateurs en cascade (réduit et complet) : un premier observateur du type défini ci-dessus est appliqué à un modèle simplifié pour fournir une estimation intermédiaire de l'état, cette estimation étant utilisée elle-même comme la sortie du système non simplifié. En effet, pour l'observateur complet avec les 3 sorties mesurées, les calculs sont très lourds, même en travaillant en coordonnées naturelles (Lafont et al. 2011).

Dans une autre étude, nous proposons une méthodologie générale pour identifier et reconstruire des défauts capteur sur un procédé dynamique. Cette méthodologie est inspirée de la théorie d'identification générale : en effet, cette théorie d'identification apporte aussi une base de travail générale pour les problèmes d'« observabilité à entrées inconnues ». Beaucoup de problèmes de détection de défauts peuvent être formulés comme des problèmes d'observabilité ; les défauts, éventuellement additifs, doivent être considérés comme entrées inconnues. Nous appliquons cette méthode de détection de défauts capteur sur un système de traitement des eaux usées qui est un cas académique idéal car, premièrement, dans le cas 3-5 (3 capteurs, 5 états), la théorie s'applique génériquement et, deuxièmement, n'importe quel système est naturellement sous la « forme canonique d'observabilité » (lorsque les sorties appartiennent à l'état) requise pour appliquer l'observateur grand gain de base (Methnani et al. 2011).

1.3 Enseignement et encadrements

1.3.1 Enseignement

Depuis ma nomination à l'IUT de Toulon au département Génie Industriel et Maintenance, j'ai été chargé du cours, TD et TP d'électricité et d'électronique de puissance. Depuis trois ans, j'enseigne le cours et les TDs d'électricité, les TDs et les TPs d'informatique (programmation Labview) et les cours et les TDs d'Analyse vibratoire.
J'ai aussi enseigné pendant plusieurs années en Licence Professionnelle Maintenance et Management Environnemental le cours et les TDs de supervision.
J'ai également enseigné le cours de logique floue en Master Professionnel Ingénierie Marine pendant plusieurs années.

1.3.2 Encadrements

J'ai co-encadré, avec le Pr. Gilles Enea, la thèse de doctorat de Amine Trabelsi : « Contribution à l'analyse et à la commande avancée des systèmes multivariables : application à une serre agricole ». Cette thèse (co-tutelle franco tunisienne) a été soutenue le 28 octobre 2006 à l'Ecole Nationale d'Ingénieurs de Sfax devant le jury composé de

Ahmed Toumi	Professeur	Président
Mansour Souissi	Professeur	Rapporteur
Boutaieb Dahhou	Professeur	Rapporteur
Mohamed Chtourou	Professeur	Examinateur
Gilles Enea	Professeur	Directeur
Mohamed Kamoun	Professeur	Examinateur
Frédéric Lafont	Maître de conférences	Co-directeur

J'ai co-encadré, avec le Pr. Gilles Enea, la thèse de doctorat de Julio Cesar Ramos Fernandez : « Intégration de techniques floues pour la modélisation, l'identification et le contrôle des systèmes non-linéaires ». Cette thèse (co-tutelle franco mexicaine) a été soutenue le 2 février 2008 à l'Université Autonome de l'état d'Hidalgo devant le jury composé de

Joseph Aguilar Martin	Professeur	Président
Elvia Ruth Palacios Hernandez	Docteur	Rapporteur
Julio Waissman Vilanova	Docteur	Rapporteur
Gilles Enea	Professeur	Directeur
Virgilio Lopez Morales	Docteur	Examinateur
Frédéric Lafont	Maître de conférences	Co-directeur

Je co-encadre actuellement, avec le Pr. Jean-Paul Gauthier, Salowa Methnani qui prépare une thèse (co-tutelle franco tunisienne) concernant les détections de défauts capteurs et actionneurs à base d'observateurs à entrées inconnues. Cette thèse devrait être soutenue à la fin de l'année civile 2011.

J'ai également dirigé deux DEA (ou Master Recherche) sur les sujets « Etude de méthodes de variation du facteur d'oubli dans un algorithme de modélisation par cluster flou adaptatif » et « Modèle adaptatif flou pour le diagnostic ».

1.4 Administration, animation scientifique et coopération industrielle

1.4.1 Administration

Dans le département Génie Industriel et Maintenance, j'ai été amené à avoir différentes fonctions telles que :
- Responsable des emplois du temps du département Génie Industriel et Maintenance de l'I.U.T (2002-2004),
- Membre élu au conseil de département Génie Industriel et Maintenance depuis 2000,
- Responsable des notes et correspondant Apogée depuis 2006,
- Membre élu au Conseil d'Aide à la Recherche et au Transfert de Technologie de l'IUT de Toulon.

Sur le plan recherche, j'ai fait partie de la commission de spécialistes de rang B en 61^{ième} section à l'Université du Sud-Toulon-Var de 2001 à 2007.

J'ai été également membre suppléant de la Commission de Spécialistes de rang B en 60-62^{ième} section de 2004 jusqu'à 2007.

J'ai participé au comité de sélection en tant que rang B en 61^{ième} section à l'Université Paul Cézanne (Marseille) en 2010 et au comité de sélection en tant que rang B en 61^{ième} section à Université du Sud Toulon Var en 2011.

1.4.2 Animation scientifique et rayonnement

De 1999 à 2007, j'ai participé à plusieurs coopérations internationales telles que : un projet franco-tunisien CMCU, un projet franco-marocain CMIFM et un projet franco-mexicain ECOS/NORD. Ces trois projets portés sur la modélisation et la commande de systèmes multivariables : application à une serre agricole.

J'ai été invité à faire une présentation orale au troisième atelier d'actualisation en soft computing organisé à l'Université Autonome de l'Etat d'Hidalgo, Mexique. La présentation s'intitulait « A fuzzy parametric approach for the sensor fault detection ».

J'ai été sollicité pour être reviewer d'articles de la revue « Fuzzy Sets & Systems » de l'éditeur Elsevier et de la revue « Journal of Dynamical & Control Systems ».

J'ai organisé les Journées de doctorants de l'équipe EStimation-COmmande-DIagnostic (ESCODI) du laboratoire (19 mai 2005) : Chaque année, les doctorants de 1ère année de l'équipe présentent l'avancement de

leurs travaux dans une configuration « colloque ». En 2005, il a été décidé d'organiser cette journée à Toulon. J'ai été l'organisateur et le responsable logistique de cette journée (environ 50 personnes).

1.4.3 Coopération industrielle

En 2007 et 2008, l'équipe ESCODI du LSIS de l'Université du Sud-Toulon-Var a développé un projet Recherche & Développement pour le groupe DCNS de Toulon. Cette étude a porté sur l'apport de la logique floue et des réseaux connexionnistes sur un module « aide à la décision » dans le milieu maritime.

Modélisation et commande à base de soft-computing

2

Sommaire

Dans ce chapitre, nous présentons différents travaux à base de soft-computing : logique floue, réseaux de neurones, ... Nous nous sommes intérressés aussi bien à des problèmes d'identification que de commande sur des systèmes multivariables, non linéaires et non stationnaires.

Le contrôle du climat d'une serre agricole, dans un but d'améliorer la culture tout en réduisant les coûts de production, est devenu un sujet important pour les exploitants. Depuis de nombreuses années, le laboratoire LSIS s'est intéressé à ce problème. En effet, le laboratoire dispose d'une serre expérimentale (figure 2.1) qui permet d'acquérir les mesures de température interne et externe, d'hygrométrie interne et externe, du rayonnement global et de vitesse de vent. L'objectif d'une régulation est de maintenir une température et une hygrométrie à une certaine valeur (consigne) pour optimiser la production à l'aide de différents actionneurs (système de chauffage, de brumisation et d'ouvrant).

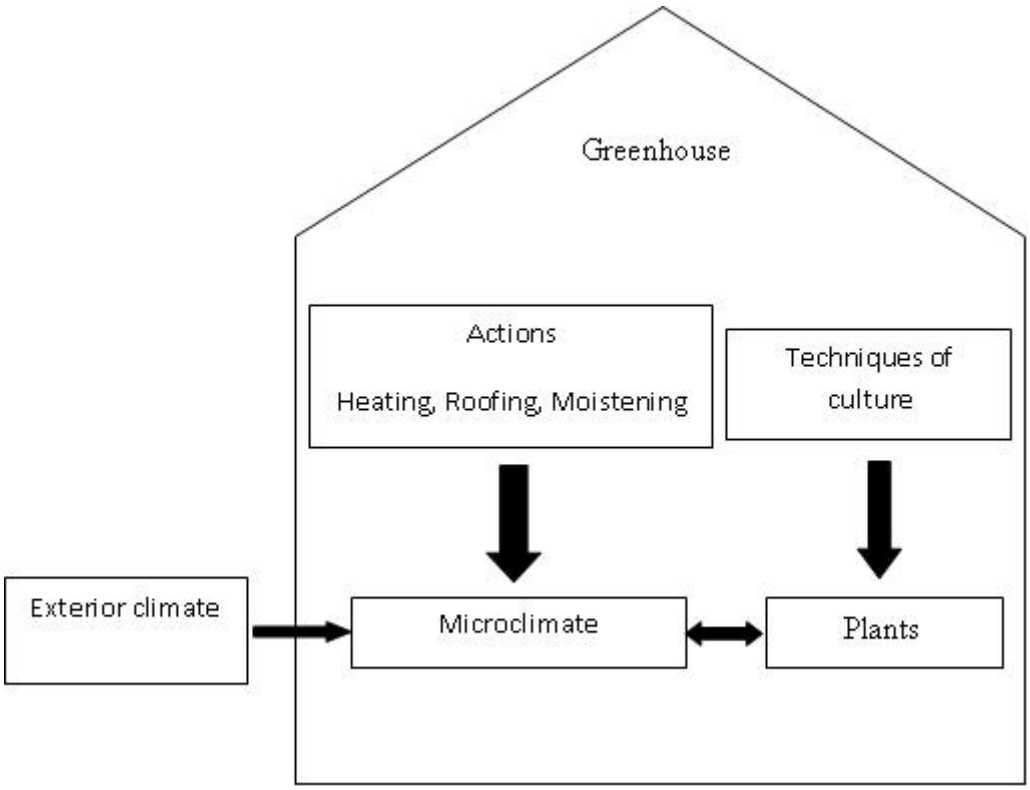


FIGURE 2.1 – *Système serre*

Des techniques de régulation conventionnelle (contrôle adaptatif multivariable, contrôle logique booléen, ...) ont été développées. Les interactions entre les données internes et externes, et la complexité du système (multivariable, non linéaire et non stationnaire) sont telles que ces techniques sont très difficiles à implémenter.
De plus, ces méthodes introduisent des simplifications et elles sont souvent très sensibles aux perturbations qui ne sont pas représentées dans le modèle.

Pour résoudre ces problèmes, nous avons envisagé de travailler sur des méthodes à base de la connaissance experte telle que la logique floue.

2.1 Commande floue

Cette section représente les résultats des travaux publiés par Lafont et Balmat (2002). L'avantage principal de cette technique est qu'il n'est pas nécessaire d'avoir un modèle pour développer une loi de commande.

Néanmoins, comme il est impossible de tester deux commandes in-situ simultanément, nous avons développé un modèle de simulation pour pouvoir comparer deux commandes sous les mêmes conditions météorologiques.

Un modèle physique basé sur l'équilibre des masses est défini comme suit :

$$\begin{array}{l} \frac{dT_{ai}}{dt} = (\alpha_1 + \alpha_2 Ov)(T_{ae} - T_{ai}) + \alpha_3 Ch + \alpha_4 Rg - \alpha_5, \\ \frac{dX_{ai}}{dt} = (\beta_1 + \beta_2 Ov)(X_{ae} - X_{ai}) + (\beta_3 + \beta_4 Rg)\Delta X_{ai} - \beta_5, \\ X_{ai} = H_{ai} Xsat_{ai}, \\ X_{ae} = H_{ae} Xsat_{ae} \end{array} \tag{2.1}$$

où T_{ae} est la température externe ($^\circ C$), Ov la commande de l'ouvrant ($^\circ$), Ch la commande de chauffage (kW), Rg le rayonnement global (kW/m^2), X_{ae} la température absolue externe (g/kg), T_{ai} est la température interne ($^\circ C$), X_{ai} la température absolue interne (g/kg), ΔX_{ai} le déficit hydrique (g/kg), $Xsat$ l'hygrométrie à la saturation, α_i et β_i les paramètres du modèle avec α_5 et β_5 les perturbations non mesurées.

L'inconvénient de ce modèle est que les paramètres que l'on trouve ne sont valables que pour une courte durée de la journée. Nous avons décidé de trouver un modèle valide pour une journée.
Pour cela, nous avons utilisé une méthode floue itérative utilisant un ensemble flou multi-dimensionnel. L'objectif était d'obtenir des règles floues de type Takagi-Sugeno. Ces règles sont générées automatiquement à partir des données entrées-sorties mesurées. Cette partition floue croit jusqu'à ce qu'un critère soit atteint.
La structure de ce modèle est défini comme suit :

$$\begin{array}{l} T_{ai}(k+1) = a_1 Ov(k) + a_2 Ch(k) + a_3 T_{ae}(k) + a_4 H_{ae}(k) + a_5 Rg(k) \\ +a_6 Vv(k) + a_7 T_{ai}(k) + a_8 H_{ai}(k) + a_9 Br(k), \\ H_{ai}(k+1) = b_1 Ov(k) + b_2 Ch(k) + b_3 T_{ae}(k) + b_4 H_{ae}(k) + b_5 Rg(k) \\ +b_6 Vv(k) + b_7 T_{ai}(k) + b_8 H_{ai}(k) + b_9 Br(k) \end{array} \tag{2.2}$$

où les paramètres a_i et b_i sont les paramètres du modèles.

L'algorithme débute à une règle. Si le critère n'est pas obtenu, l'espace entrées-sorties est partitionnée en deux règles. Cette procédure est répétée tant que le critère n'est pas atteint. Pour notre application, une bonne précision du modèle est obtenue lorsque le nombre de règles est 25. Nous avons validé le modèle obtenu sur une journée avec une commande classique binaire on-off (figure 2.2 et 2.3).

Le diagramme fonctionnel du système avec le contrôleur flou est représenté par la figure 2.4.

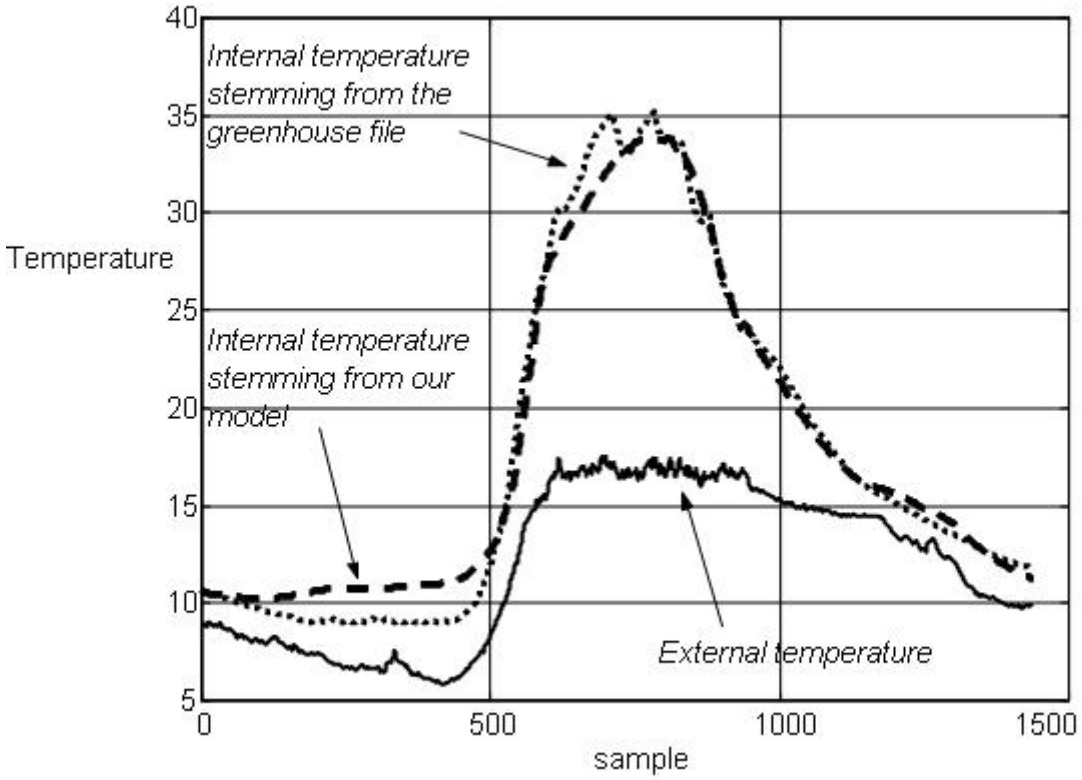


FIGURE 2.2 – *Modélisation de la température*

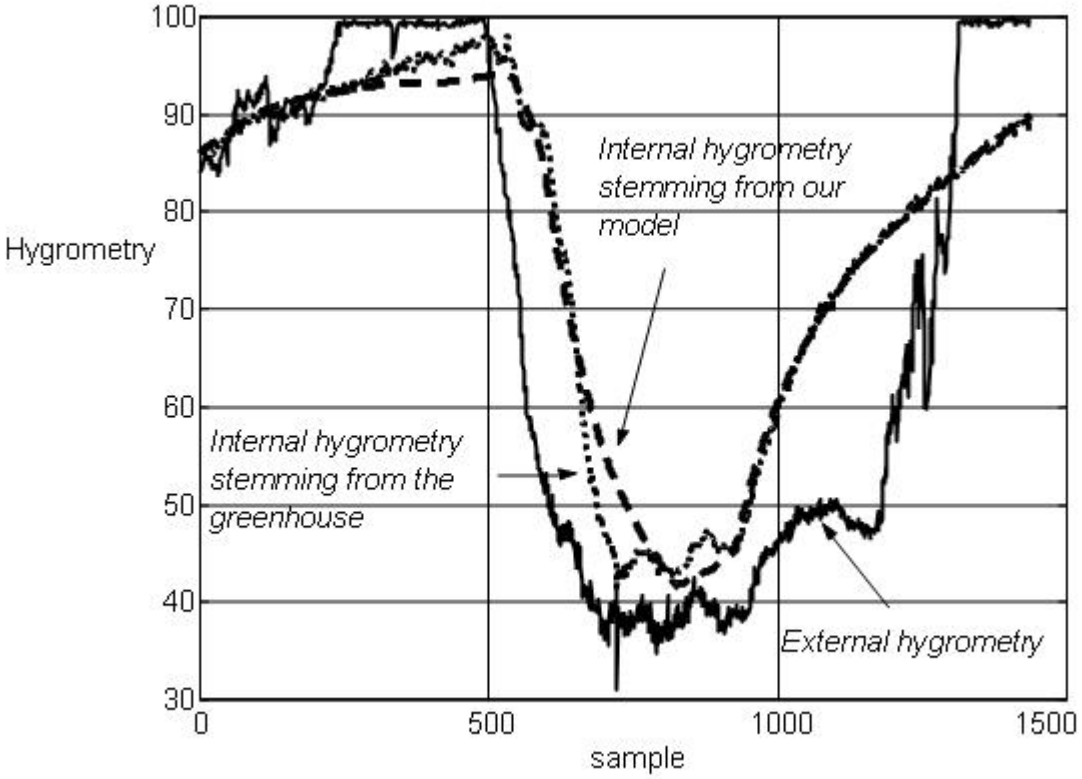


FIGURE 2.3 – *Modélisation de l'hygrométrie*

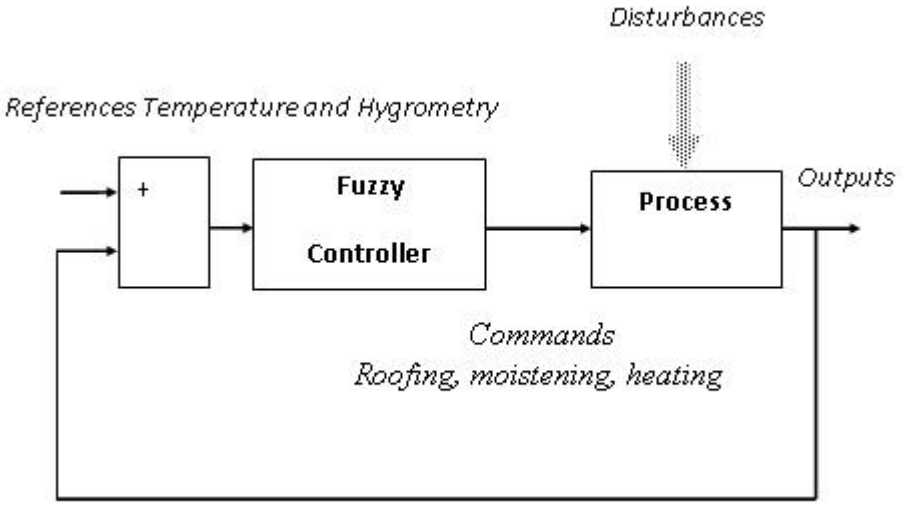


FIGURE 2.4 – *Boucle fermée avec le contrôleur flou*

	Contr. classique	Contr. basique	Contr. optimisé
Moy. erreur temp.	7.2	6.2	7.6
Moy. erreur hygro.	17.2	20.9	17.2
Chauffage (%)	3.4	2.8	0
Brumisation (%)	2.2	1.8	1.9
Ouvrant (%)	20	34	20

TABLE 2.1 – *Comparaisons des commandes*

Comme données, nous avons les sorties associées à la température et l'hygrométrie interne à la serre (mesures de capteurs), les consignes en température T_c et en hygrométrie H_C, les perturbations externes T_e et H_e. Une des difficultés, pour l'implémentation d'un système flou, est le choix et le nombre des entrées. Dans notre étude, nous avons défini un contrôleur flou à 4 entrées et 3 sorties. Nous avons utilisé les variations de la température et de l'hygrométrie (interne et externe) comparées aux consignes (figure 2.5).

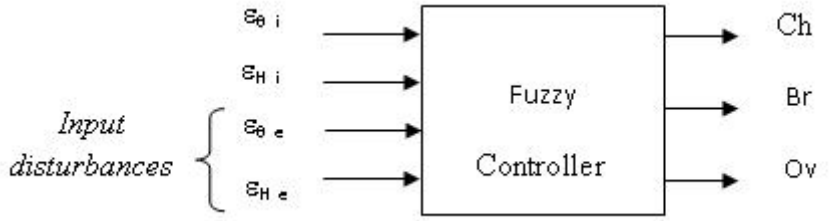


FIGURE 2.5 – *Contrôleur flou basique*

Nous avons donc défini les variations en température par $\epsilon T_i = T_c - T_i$ et $\epsilon T_e = T_c - T_e$ et les variations en hygrométrie $\epsilon H_i = H_c - H_i$ et $\epsilon H_e = H_c - H_e$. Les commandes de chauffage et de brumisation varient entre 0 et 1 (on-off), et pour l'ouvrant entre -1, 0 et 1 (ouvrir, aucune action, fermer).

La base de règles est composée de 81 règles. La méthode d'inférence est la méthode min-max et le centre de gravité est utilisé comme méthode de déffuzzification. Du fait du nombre de fonctions d'appartenance et donc des règles, nous avons décidé de développer un contrôleur flou optimisé en observant les différents couplages entre les différentes variables qui peuvent exister (figure 2.6).

A partir de l'étude de couplage, nous pouvons déterminé une nouvelle structure (figure 2.7) qui est plus facile à mettre au point car chaque contrôleur flou a un nombre d'entrées plus faible et donc une base de règles réduite par rapport au contrôleur flou basique.

Ce contrôleur optimisé permet de prendre en compte en entrée la variation des dérivées d'entrées (ce qui était impossible avec le contrôleur basique car il aurait fallu rajouter 4 entrées supplémentaires).

Nous pouvons donner un tableau comparatif des différents contrôleurs développés (Table 2.1).

L'étude du couplage des informations a permis de développer un contrôleur flou de complexité réduite avec une validité d'une journée (contrôleur basique valable quelques heures).

Inputs			Outputs	
Ch	Ov	Br	Ti	Hi
+	0	0	++	-
++	0	+	+++	0
0	+	+	-	+
0	0	+	0	++
0	++	0	---	-
0	0	++	0	+++

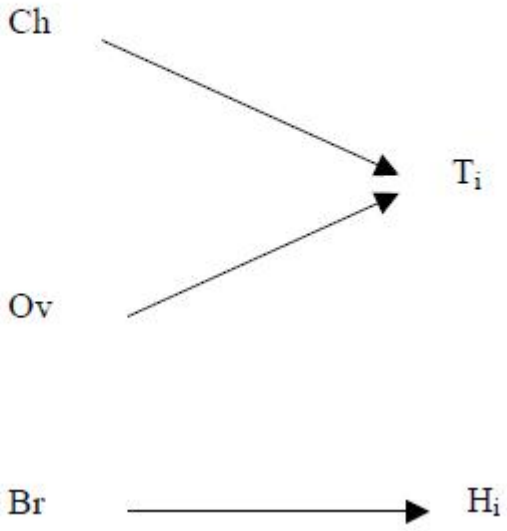


FIGURE 2.6 – *Couplage*

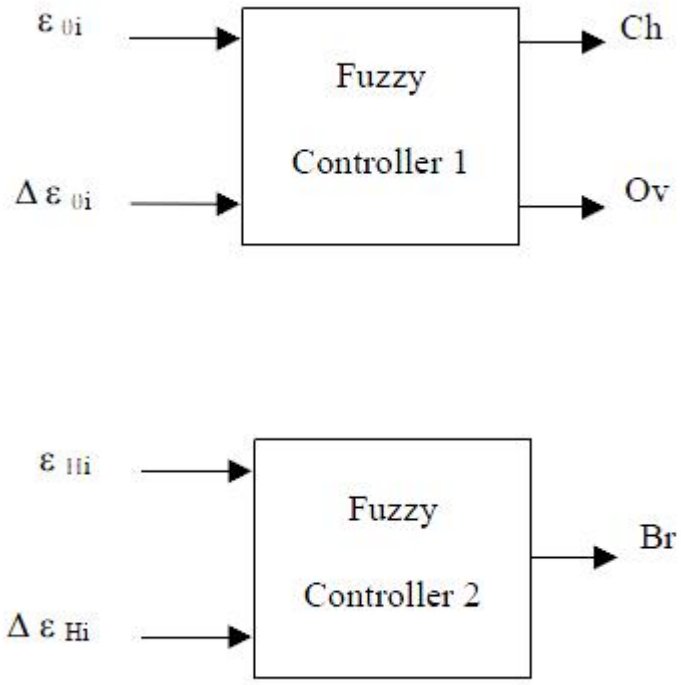


FIGURE 2.7 – *Contôleur flou optimisé*

2.2 Modélisation floue par clusters flous

Ce paragraphe représente les résultats des travaux publiés dans Trabelsi et al. (2007). Après avoir travaillé sur la commande et vu la complexité du système, nous nous sommes intéressés à la modélisation. Dans la littérature, les techniques pour la modélisation mathématique des processus sont classifiées en trois principales catégories : la modélisation par modèle de connaissance, la modélisation par modèle de représentation et la modélisation qui combine les deux modèles. La première est basée sur les lois physiques régissant le processus et la deuxième est basée sur l'analyse des données d'entrées-sorties du processus.

D'autres théories de modélisation sont apparues telles que celles qui utilisent les concepts des réseaux de neurones et la logique floue. Ces deux théories permettent la description de la dynamique des systèmes complexes (non linéaires, de grande dimension, ...) de manière satisfaisante.

Les modèles flous de type Takagi-Sugeno (TS) permettent d'obtenir des sous-modèles linéaires dans tout l'espace d'entrée-sortie du processus, d'où la possibilité d'appliquer les théories de l'automatique classique pour développer une loi de commande qui satisfera les objectifs visés.

L'objectif de l'analyse par cluster (groupe) est la classification d'objets en accord avec la ressemblance entre eux et l'organisation de données en groupes. Les techniques de clustering peuvent être appliquées aux données quantitatives, qualitatives ou à un mélange des deux. Les méthodes des clusters flous permettent l'appartenance d'objets aux divers groupes

de manière simultanée, avec des niveaux d'appartenance différents. Le concept de ressemblance parmi les objets peut être compris comme une ressemblance mathématique, définie au moyen d'une norme de mesure de la distance entre les vecteurs de données, ou comme une distance d'un vecteur de données à un objet prototype du groupe. Les prototypes ne sont pas connus habituellement par avance. Ils apparaissent simultanément à l'application des algorithmes de clustering avec la partition des données. Les prototypes peuvent être des vecteurs de la même dimension que les objets de données, mais ils peuvent aussi être définis comme des objets géométriques de « haut niveau », tels que des sous-espaces ou des fonctions linéaires et non linéaires. En se basant sur cette ressemblance, ces objets, représentés par les données d'entrées-sorties du processus à étudier, peuvent être regroupés de telle sorte qu'ils sont aussi similaires que possible à l'intérieur du cluster qu'ils forment et sont aussi différents que possible des objets des autres clusters. L'idée des clusters flous est schématisée sur la Figure 2.8 où les données sont regroupées en deux groupes autour de leurs prototypes v_1 et v_2 en utilisant la mesure de distance euclidienne.

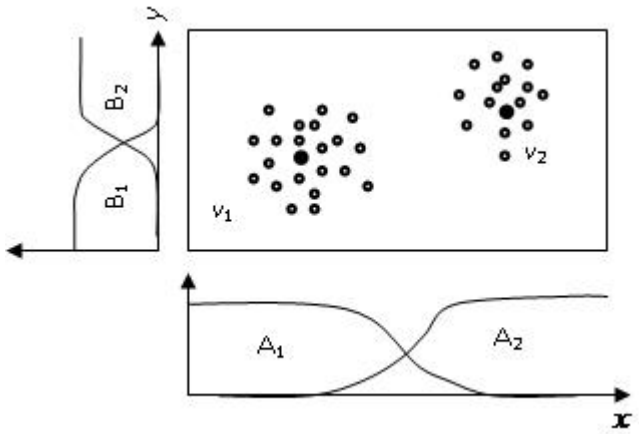


FIGURE 2.8 – *Principe des clusters flous*

Le partitionnement des données est exprimé dans une matrice de partition floue dont les éléments μ_{ij} sont les degrés d'appartenance des données $[x_i, y_i]$ à un cluster flou avec les prototypes v_j.

La matrice de partition floue et les objets prototypes sont obtenus par les différents algorithmes de clustering. En se basant sur cette idée, les règles peuvent êtres extraites par la projection des clusters sur les variables du modèle.

La forme des fonctions d'appartenance obtenues par la projection des clusters dépend de la distribution des données et le concept de similitude de ces données à un prototype donné nous laisse une multitude de choix de mesure de distance. Cette mesure de distance peut être adaptative telle que celle utilisée par Gustafson et Kessel en 1979 pour l'appliquer à l'identification des modèles flous de Takagi-Sugeno pour une large classe de systèmes non linéaires dynamiques. L'idée de base de cette méthode est illustrée par la Figure 2.9.

La forme hyperellipsoïdale des clusters est décrite par la structure

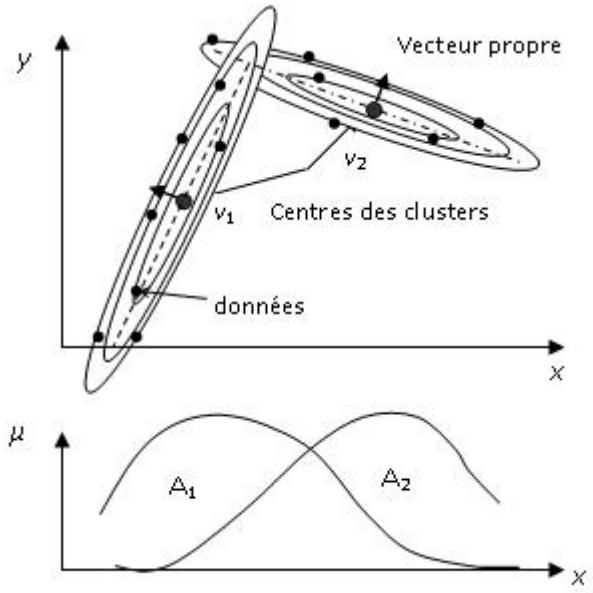


FIGURE 2.9 – *Clusters de Gustafson-Kessel*

propre de la matrice de covariance de ces clusters. En effet, Les vecteurs propres de cette matrice déterminent la direction des axes des hyperellipsoïdes et la longueur de ces axes est donnée par les valeurs propres de cette matrice. Les fonctions d'appartenances des ensembles flous A_1 et A_2 sont générées par la projection des clusters sur la variable d'entrée x. Le nombre de clusters c'est-à-dire le nombre de règles doit être spécifié avant l'étape de clustering. Ce nombre est lié aux types de non linéarités que le système est susceptible de présenter. Plus le nombre de règles dans le modèle est élevé, plus l'approximation sera bonne, mais aussi plus le nombre de paramètres à estimer ainsi que leurs variances seront élevés.

Cet algorithme requiert un jeu de données Z, un nombre de clusters N_c défini par l'utilisateur, le degré de pondération m et la tolérance tol pour l'arrêt de l'algorithme. La procédure peut être décomposé en différentes étapes :

– **Construction du régresseur**
La matrice de régression et le vecteur de sortie sont construits à partir des séquences de mesure :

$$X(k) = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(k) \\ \vdots \\ x(N-1) \end{bmatrix} \tag{2.3}$$

$$y(k) = \begin{bmatrix} y(2) \\ y(3) \\ \vdots \\ y(k+1) \\ \vdots \\ y(N) \end{bmatrix} \tag{2.4}$$

$$Z^T = [X\ y] \tag{2.5}$$

N est le nombre d'échantillons. Les données, fournies pour l'opération de clustering, sont concaténées dans la matrice Z.

– Partionnement des données

L'algorithme de Gustafson-Kessel permet, connaissant Z et le nombre de clusters N_c, de déterminer la matrice de partition floue $U = [\mu_{ik}]_{N_c \times N}$ avec $\mu_{ik} \in [0,1]$, la matrice de prototypes $V = [\nu_1, \ldots, \nu_{N_c}]$ et les matrices de covariance des clusters $F = [F_1, \ldots, F_{N_c}]$ (F_i sont des matrices définies positives dans $\mathbb{R}^{p \times p}$).

Une fois le triplet (U, V, F) déterminé, les paramètres des prémisses des règles (ν_i et σ_i) et les paramètres des conséquences (A_i et B_i) sont calculés.

– Détermination des paramètres des prémisses

Dans ce travail, nous avons utilisé les gaussiennes comme fonctions d'appartenances pour représenter les Ω_{ij} :

$$\Omega_{ij}\left(x_j(k)\right) = exp\left(-\frac{1}{2}\frac{\left(x_j - \nu_{ij}\right)^2}{\sigma_{ij}^2}\right) \tag{2.6}$$

L'équation 2.6 peut être écrite sous la forme compacte suivante :

$$\mu_j(x(k)) = exp\left(-\frac{1}{2}\left(x(k) - \nu_i\right)^T \left(F_i\right)^{-1}\left(x(k) - \nu_i\right)\right) \tag{2.7}$$

avec $\nu_i = \left[\nu_{1i}, \ldots, \nu_{pi}\right]$ le vecteur des centres et F_i la matrice des variances :

$$F_i = \begin{bmatrix} \sigma_{1i}^2 & 0 & \ldots & 0 \\ 0 & \sigma_{2i}^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{pi}^2 \end{bmatrix} \tag{2.8}$$

Ainsi l'algorithme de Gustafson-Kessel permet de déterminer le vecteur des centres qui sera égal au vecteur des prototypes V et la matrice des variances qui sera égale à la matrice des covariances des clusters F.

– Détermination des paramètres de conséquences des règles

Les paramètres des conséquences des règles sont estimés séparément par l'algorithme des moindres carrés pondérés en minimisant le critère suivant :

$$min_{\theta_i} \frac{1}{N} (y - X\theta_i)^T Q_i (y - X\theta_i) \tag{2.9}$$

Soit $\phi_i (x(k), \nu_i, \sigma_i)$ la fonction de validité pour la fonction d'appartenance qui est choisie comme gaussienne, définie par son centre ν_i et son écart type σ_i et donnée par :

$$\phi_i (x(k), \nu_i, \sigma_i) = \frac{\mu_i x(k)}{\sum_{i=1}^{K} \mu_i x(k)} \tag{2.10}$$

Q_i est une matrice contenant les valeurs de la fonction de validité $\phi_i (x(k), \nu_i, \sigma_i)$ pour le $i^{\text{ième}}$ modèle local pour chaque échantillon k.

$$Q_i = \begin{bmatrix} \phi_i (x(1), \nu_i, \sigma_i) & 0 & \dots & 0 \\ 0 & \phi_i (x(2), \nu_i, \sigma_i) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_i (x(N), \nu_i, \sigma_i) \end{bmatrix} \tag{2.11}$$

Le vecteur des paramètres de conséquence des règles est donné par :

$$\theta_i = \left[X^T Q_i X\right]^{-1} X^T Q_i y \tag{2.12}$$

L'extension que nous avons proposé de faire est l'adaptation on-line des paramètres des modèles linéaires locaux.

– **Adaptation on-line des paramètres des modèles linéaires locaux**

La raison principale pour laquelle nous proposons de faire l'adaptation on-line des paramètres des modèles linéaires locaux est la variation du comportement du processus dans le temps c'est-à-dire la non stationnarité des paramètres des modèles locaux.

Un modèle flou du type Takagi-Sugeno possède des paramètres non linéaires qui représentent la prémisse de la règle (centres et écarts-type des fonctions d'appartenance) et des paramètres linéaires qui représentent les paramètres de conséquences de la règle.

Généralement, les modèles de TS obtenus par la méthode des clusters flous ont des paramètres de conséquences constants, c'est à dire la conséquence d'une règle floue est donnée par :

$$y_i (k+1) = A_i y(k) + B_i u(k) \tag{2.13}$$

En utilisant la méthode du centre de gravité, la sortie du modèle global de TS est donnée par :

$$y(k+1) = \frac{\sum_{i=1}^{N_r} \mu_i (x(k))\, y_i (k+1)}{\sum_{i=1}^{N_r} \mu_i (x(k))} \tag{2.14}$$

Le système non linéaire est composé par un ensemble de modèles linéaires locaux à paramètres constants. Généralement, les systèmes dynamiques et non linéaires possèdent des paramètres évoluant au cours du temps, d'où la nécessité d'une adaptation paramétrique. En effet, les paramètres de chaque modèle local sont adaptés à chaque période d'échantillonnage par l'algorithme récursif des moindres carrés ordinaires avec facteur d'oubli. Dans la phase en ligne, les paramètres de la prémisse de la règle sont gardés fixes et seulement les paramètres de conséquences de la règle sont adaptés par l'algorithme d'adaptation paramétrique récursif, c'est à dire à chaque pas d'échantillonnage k, nous obtenons le modèle de TS donné par l'équation suivante :

$$y_i(k+1) = A_i(k)\, y(k) + B_i(k)\, u(k) \tag{2.15}$$

Il est important d'incorporer dans cet algorithme d'adaptation une procédure qui assure sa robustesse par rapport aux variations éventuelles de la dynamique du processus à commander. Une approche communément utilisée pour doter ces algorithmes d'une certaine robustesse consiste à éviter que le gain d'adaptation devienne nul.

Pour cela, nous avons choisi un algorithme avec facteur d'oubli. L'adaptation paramétrique est réalisée pour chaque modèle local, c'est à dire pour chaque règle, par une version récursive de l'algorithme des moindres carrés pondérés avec facteur d'oubli $\lambda(k)$:

$$\theta_i(k) = \theta_i(k-1) + \delta_i(k)\left(y(k) - x(k)^T\right)\theta_i(k-1) \tag{2.16}$$

$$\delta_i(k) = \frac{P_i(k-1)\, x(k)}{\frac{\lambda(k)}{\phi_i(x(k),\nu_i,\sigma_i)} + x(k)^T P_i(k-1)\, x(k)} \tag{2.17}$$

$$P_i(k) = \frac{1}{\lambda(k)}\left[I - \delta_i(k)\, x(k)^T\right] P_i(k-1) \tag{2.18}$$

Dans l'équation 2.16, le vecteur de paramètres θ_i à l'instant $(k=1)$ est égal à celui estimé dans l'équation 2.12. Il est adapté en ajoutant un vecteur de correction à $\theta_i(k-1)$. Dans les équations 2.17 et 2.18, la fonction ϕ_i permet une pondération sur les données actuelles via les fonctions de validité des règles. Le facteur d'oubli $\lambda(k)$ présent dans la matrice de gain $P_i(k)$ permet d'améliorer la capacité de ce gain d'adaptation et d'assurer un meilleur suivi des paramètres variant dans le temps. En effet, le facteur d'oubli empêche que le gain ne devienne trop petit de telle sorte que toute nouvelle donnée (valeur mesurée de l'entrée et de la sortie) dans le vecteur d'observation continue à avoir un effet sur la qualité de l'identification.

2.3 Structure multi-modèle

Une approche de modélisation dans le domaine des systèmes multi-entrées et multi-sorties (MIMO), non linéaires, non stationnaires et fortement perturbés est considérée dans cette étude. Pour cette classe de système, il est parfois très difficile de trouver un modèle général. Nous nous sommes fixés comme objectif d'obtenir un modèle permettant de comparer différentes lois de contrôles ou pour détecter des défauts capteurs.

Nous avons développé une approche multi-modèles (Pessel et al. 2009). Le diagramme de synthèse de notre méthode est représenté par la figure 2.10

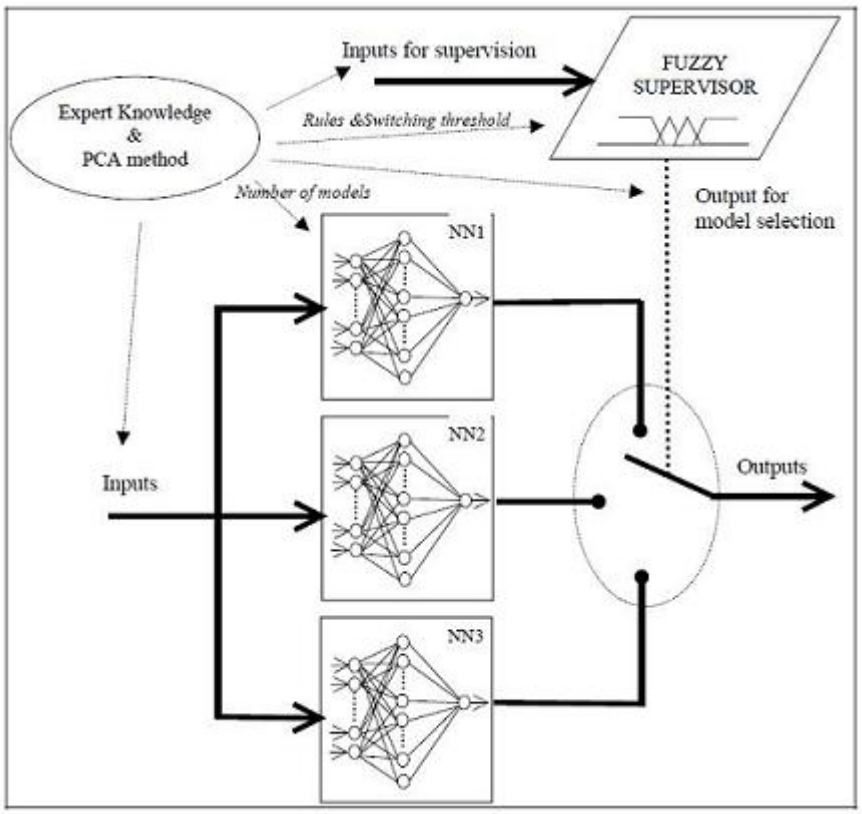


FIGURE 2.10 – *Synthèse de la méthode proposée*

L'application se fait sur le système serre agricole. Un point fondamental est de définir une stratégie pour l'obtention des sous-modèles. La connaissance experte du système nous permet de définir 3 modes de fonctionnement dans une journée : le jour, la nuit et le lever du soleil. L'étude de l'analyse en composantes principales nous permet de déterminer la structure de chaque sous-modèle. Nous pouvons voir les structures de ces sous-modèles par rapport au modèle global sur la figure 2.11

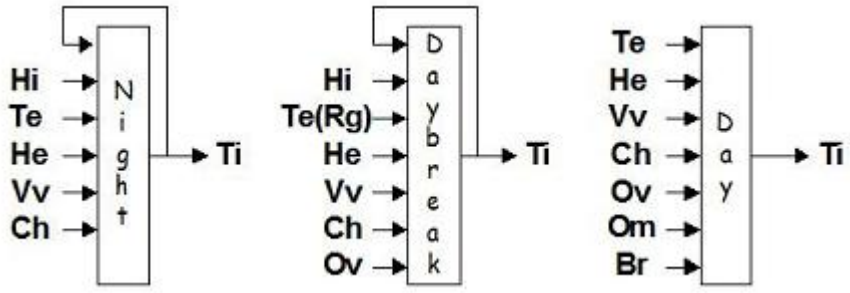


FIGURE 2.11 – *Modèle global et sous-modèles*

Chaque sous-modèle est identifié à l'aide des reséaux de neurones. Le superviseur flou est un classifieur flou hiérarchisé, i.e. qu'il est décomposé

		Modèle global	Commutation	Fusion
10 Mars	Mean	0.72	0.60	0.60
Appren.	Variance	0.49	0.29	0.29
11 Mars	Mean	1.58	0.74	0.74
Valid.	Variance	1.52	0.39	0.38
12 Mars	Mean	1.45	0.62	0.62
Valid.	Variance	1.44	0.34	0.34

TABLE 2.2 – *Comparaisons entre modèle global et multi-modèles*

en deux blocs flous de deux entrées et la sortie du premier est une entrée du second (figure 2.12).

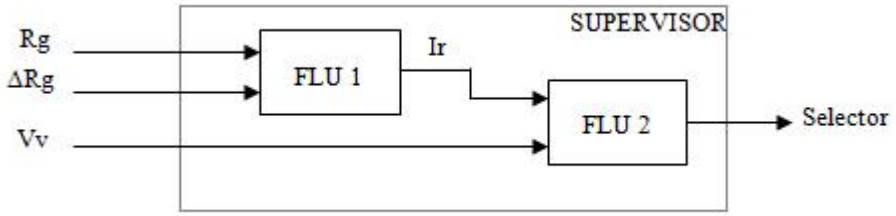


FIGURE 2.12 – *Classifieur flou hierarchique*

Nous avons comparé les résultats trouvés entre le modèle global et la structure multi-modèle en commutant ou en fusionnant les sous-modèles (Table 2.2).

Cette étude a porté sur la modélisation de systèmes complexes. La classe considérée peut être à la fois multivariable, non linéaire, non stationnaire et pour des systèmes fortement perturbés. Ces processus ne sont pas facilement modélisables. Nous avons proposé une structure variable multi-modèles à l'aide des techniques de soft-computing.

3 Un module d'aide à la décision pour l'évaluation des risques et menaces maritimes

Sommaire

Dans ce chapitre, nous présentons une étude industrielle, demandée par la DCNS, qui porte sur les problèmes d'évaluation des risques et menaces maritimes.

3.1 Introduction

Le travail a été réalisé dans le cadre d'un projet DCNS Systèmes d'Information et de Sécurité sur l'étude de la logique floue et des réseaux de neurones appliquées à l'évaluation des risques et menaces pour la sauvegarde maritime. L'objectif est de définir automatiquement un facteur de risque, d'une situation maritime à un instant donné pour chaque navire, qui sera utilisé pour la conception d'un système d'aide à la décision. Cette étude traite plus particulièrement de problème de la lutte anti-pollution en haute mer.
La structure du système est basée sur une architecture hiérarchique et modulaire constituée essentiellement d'un ensemble de blocs flous à deux entrées. Ce système d'aide à la décision fournit un facteur de risque qui tient compte des aspects statiques (caractéristiques des navires et paramètres liés aux détentions) et dynamiques (météorologie, évolution de la trajectoire du navire) que nous avons déterminés à partir d'une étude approfondie des données pertinentes relatives au domaine maritime. Ce chapitre est un résumé des articles Balmat et al. (2009) et Balmat et al. (2011).

3.2 Conception du système

L'objectif de notre travail est de concevoir un système d'aide à la décision capable de fournir une évaluation du facteur de risque maritime individuel pour chaque navire. Pour réaliser ce système nous avons sélectionné les données d'entrée pertinentes, tout en effectuant une analyse experte du problème.

3.2.1 Choix des données pertinentes pour la définition du facteur de risque

Un Indicateur de Risque Individuel de navires pour la Sécurité en mer (IRIS) existe et est basée sur l'analyse de données des accidents listés par l'Organisation Maritime Internationale (IMO) depuis plusieurs années. C'est à partir de ces travaux que nous avons développé un système d'aide à la décision pour l'Evaluation des Risques Maritimes (ERM). La sauvegarde maritime pour la protection de l'environnement dépend de nombreux éléments et critères. Ceux-ci peuvent être liés aux caractéristiques, aux évènements historiques, à la trajectoire concernant le navire et à la météorologie. Ces données sont les entrées du système ; elles permettent de calculer le facteur de risque individuel pour chaque navire. Il existe un grand nombre de bases de données (Lloyd's Register, IMO, EQUASIS, Paris MOU) qui sont susceptibles de fournir les informations nécessaires en fonction des données que nous choisissons. Ainsi, nous avons retenu comme données pertinentes et adaptées au problème, les données qui concernent les caractéristiques du navire (le type, le pavillon, l'année de construction, le tonnage, simple ou double coque), les évènements historiques du navire (le nombre de jours d'arrêt (détentions), le nombre de changements de compagnies), la trajectoire (position et vitesse du navire, provenance et destination) et la météorologie (état de la mer, vitesse du vent, visibilité, nuit/jour).

3.2.2 Une approche floue

Les données sont plus ou moins imprécises, elles doivent être utilisées dans un raisonnement nuancé donc incertain. On retrouve des éléments d'incertitudes, ne serait-ce que dans la détermination de toutes les conditions liées à l'environnement ou à la connaissance précise de la cinématique des navires. Cette incertitude peut aussi découler de données inexactes ou incomplètes. Une des difficultés de l'évaluation du risque réside dans la prise en compte d'un grand nombre de paramètres. Il n'est pas évident d'exprimer précisément une valeur, en revanche il peut être plus simple de faire formuler par un expert des règles à l'aide de termes vagues. L'analyse par logique floue est une approche utile et efficace pour résoudre des problèmes liés à l'imprécision, à l'incertitude et à la subjectivité des données. Elle permet de contourner l'insuffisance d'informations et la variation des connaissances disponibles. En fait, la logique floue permet de traiter l'imprécision et l'incertitude inhérentes à la plupart des problèmes. C'est pour cette raison que nous avons choisi d'utiliser une approche floue pour la conception de notre système d'aide à la décision.

3.3 Architecture proposée

L'une des principales difficultés dans la conception d'un système d'aide à la décision est liée à la prise en compte d'un nombre important de variables d'entrée. Cela a pour principal inconvénient de complexifier les systèmes à base de connaissances. Pour y remédier et afin de définir un système évolutif, nous proposons une architecture modulaire et hiérarchique (figure 3.1).

Modulaire car selon que l'on se situe en haute mer ou au bord des côtes, le changement des règles de décision doit être facilement réalisable. Hiérarchique de façon à avoir des blocs flous avec deux entrées seulement pour une facilité de mise au point (problème du nombre de règles pour chaque bloc). Dans cette architecture, nous trouvons essentiellement des classifieurs flous à deux entrées associés à quelques blocs non flous (pour la prise en compte, par exemple, du type de navire). L'évaluation des risques maritimes est réalisée pour chaque navire, à partir de quatre facteurs de risque relatifs aux caractéristiques statiques, à la météorologie, à l'évolution de la vitesse du navire et à la position par rapport à un rail de navigation.

Pour la partie statique, nous avons utilisé les résultats d'expertise pour évaluer l'influence de certaines variables statiques sur le taux d'accident. A partir de ces statistiques, nous avons conçu une structure floue à partir de laquelle nous pouvons déterminer un facteur de risque statique.

Pour la partie dynamique, nous avons développé le système en prenant en compte d'une part la météorologie et d'autre part l'évolution cinématique du navire. Cette structure est constituée de blocs flous qui permettent d'évaluer le facteur de risque météo, le facteur de risque évolution vitesse et le facteur de risque sortie rail. Le facteur de risque global est déterminé à partir du bloc logique de décision.

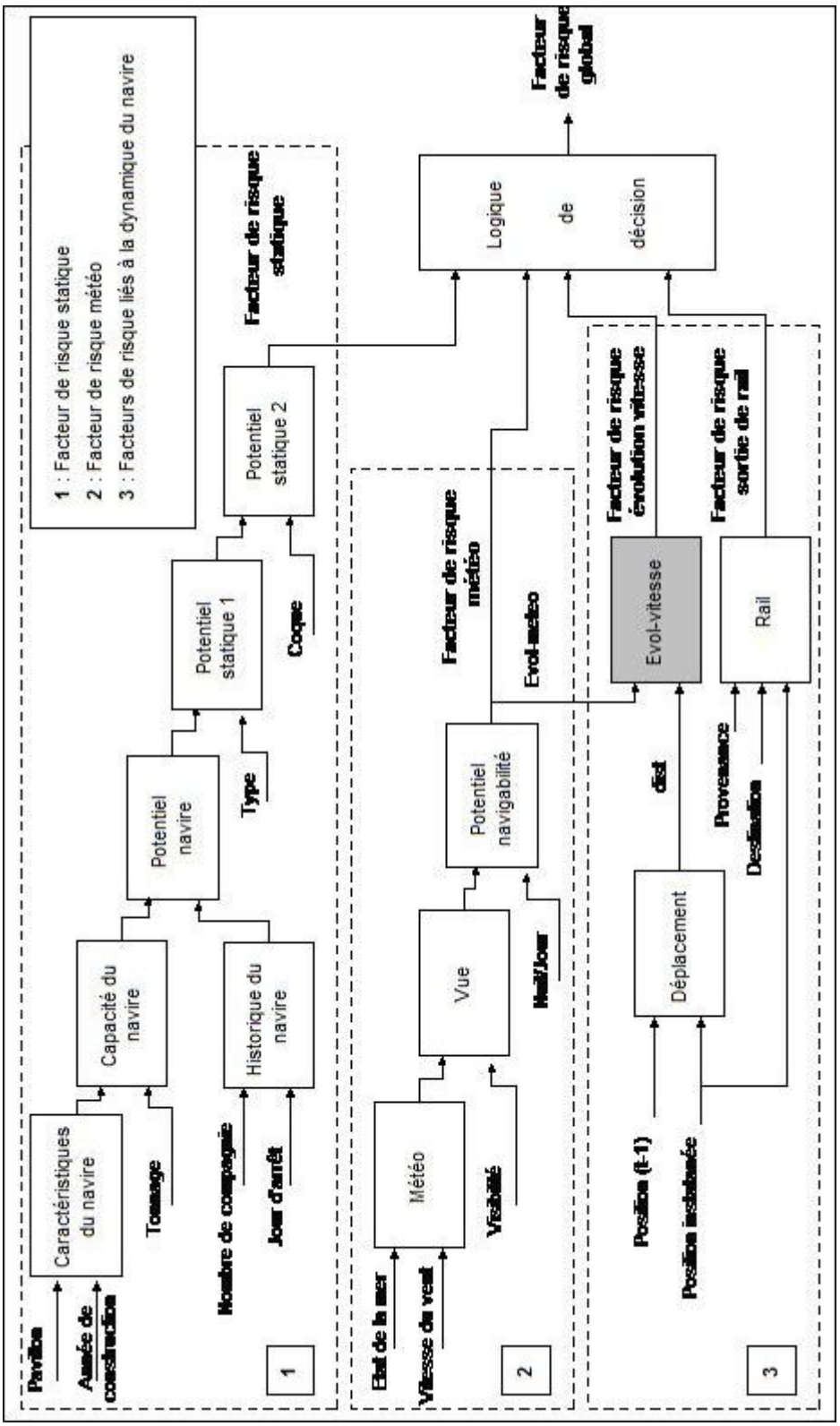


FIGURE 3.1 – *Architecture du système d'évaluation du risque maritime*

3.3.1 Facteur de risque statique

Les données d'entrées qui sont utilisées sont les suivantes : le pavillon, l'année de construction, le tonnage, le nombre de compagnies, le nombre de jour d'arrêt, le type de navire et le nombre de coques. Le facteur de risque statique est estimé à partir de quatre classifieurs flous à deux entrées associés à deux blocs de pondération. Le premier classifieur flou Caractéristiques du navire permet grâce à l'année de construction du navire (récent ou ancien) et à son pavillon (utilisation de l'Excess Factor de Paris MOU) de donner une première estimation sur le risque dû au navire. Cependant, deux caractéristiques du navire ne sont pas exploitées : le type du navire et le fait qu'il soit simple ou double coque. Ces deux paramètres ne sont pas des données floues. Le navire est soit un bateau transportant des passagers, soit un pétrolier, soit un container, soit un cargo soit un vraquier,.... D'après une étude sur la base de données de l'OMI, on a pu déterminer les valeurs des taux d'accident en fonction du type de navire. Une étude similaire a été réalisée en fonction du nombre de coques. Ces données non-floues sont prises en compte par deux blocs de pondération Potentiel statique 1 et Potentiel statique 2.

3.3.2 Facteur de risque météo

Le facteur de risque météo dépend des conditions météorologiques et de la période considérée dans la journée. Les informations concernant la météorologie (la vitesse du vent, l'état de la mer et la visibilité) sont données par Météo France. Dans le « Guide marine », il est précisé les différentes échelles existantes pour la météo marine : échelle Beaufort, échelle de l'état de la mer, échelle de visibilité. Ces échelles sont associées à des intervalles de valeurs (exemple pour l'échelle Beaufort : force 6 correspond à un vent frais dont la vitesse moyenne est comprise entre 39 et 49 km/h). A partir de ces valeurs, nous avons défini les intervalles des fonctions d'appartenance des blocs flous. Le facteur de risque météo est estimé grâce à deux classifieurs flous (Météo et Vue) associé à un bloc non flou (Potentiel Navigabilité) qui permet de pondérer la sortie du classifieur Vue en fonction de la période de la journée (entrée Nuit/Jour).

3.3.3 Facteurs de risque liés à la dynamique du navire

L'étude de la dynamique du navire permet de mettre en évidence deux comportements suspects caractérisés par : soit une distance parcourue entre deux instants successifs insuffisante (due à une dérive contrôlée ou à une avarie) ou trop importante, soit un changement de trajectoire (dû à une mauvaise mer ou à un acte de pollution). Ces deux comportements nous ont permis de définir deux facteurs de risque : le facteur de risque évolution vitesse et le facteur de risque sortie de rail. Ces facteurs sont estimés à partir de logiques de décision qui permettent d'évaluer en fonction de différents critères si le navire présente un comportement suspect. Pour ne pas être considéré comme suspect, un navire doit toujours se situer dans un cercle de rayon inférieur à un seuil fixé par rapport à sa trajectoire définie par sa provenance et sa destination. La valeur du seuil est

choisie en fonction de la tolérance d'écart de navigation du navire entre la trajectoire idéale (rail de navigation) et la trajectoire vraie.

3.3.4 Facteur de risque global

Le facteur de risque global est estimé à partir des quatre facteurs de risque présentés précédemment : le facteur de risque statique, le facteur de risque météo, le facteur de risque évolution vitesse et le facteur de risque sortie de rail. Un premier opérateur permet de réaliser une pondération entre le facteur de risque statique et le facteur de risque météo. Une combinaison logique entre la sortie de cet opérateur, le facteur de risque évolution vitesse et le facteur de risque sortie de rail permet d'évaluer le facteur de risque global.

3.4 Présentation du simulateur et résultats

Nous décrivons le simulateur développé qui nous permet de valider notre approche en simulant différents scénarii.

3.4.1 Description du simulateur

Une interface graphique a été développée sous Labview. Cette interface est composée de deux fenêtres : simulation et trajectoire. La première fenêtre permet de sélectionner, à partir de menus déroulants, un scénario constitué de données réelles. Ces données réelles sont associées à des fichiers contenant les différentes coordonnées (longitude/latitude) liées à la trajectoire et la météo de la zone de navigation du navire. Pour ces évaluations, nous avons utilisé des données de trafic maritime réelles issues du système naval d'information de commandement maritime NAOS de DCNS. Pour chaque navire, nous avons enregistré les données suivantes : date, heure, latitude, longitude, vitesse et cap. Pour les données météorologiques, elles sont enregistrées à partir de données de Météo France pour la journée correspondant à la journée de la trajectoire ou pour une journée type, en tenant compte des positions des navires et des zones météorologiques. Cette fenêtre permet de définir la tolérance par rapport au rail de navigation et de visualiser l'évolution des différents facteurs de risque au cours de la simulation. La deuxième fenêtre permet de visualiser la trajectoire idéale affichée par un trait plein, les positions successives du navire affichées au fil de la simulation, la destination indiquée dans une zone de texte, l'évolution de la météo (Mer, vent et visibilité) affichée dans des zones de texte et une fiche rappelant les informations générales du navire surveillé (issues de la Lloyd's Register). Le signalement du risque est réalisé avec deux indicateurs : la couleur associée à la position du navire et l'ouverture d'une fenêtre d'alarme. La symbolique des couleurs est la suivante : vert - navire non suspecté, rouge - navire potentiellement dangereux et orange - avis à confirmer. Une fenêtre d'alarme indique le ou les paramètre(s) (données d'entrée du simulateur) responsable(s) de cette alarme.

Heure	Latitude	Longitude	Vitesse (Noeuds)	Cap (°)
12 :14 :19	43°08′15″N	005°23′10″E	19,1	122,8
13 :14 :31	42°57′53″N	005°44′07″E	18,8	121,9
14 :14 :43	42°47′23″N	006°05′04″E	18,2	124,7
15 :14 :12	42°37′15″N	006°25′51″E	19,3	123,7
16 :14 :53	42°26′30″N	006°47′46″E	18,9	123,4
16 :46 :24	42°21′06″N	006°58′44″E	18,8	123,2
19 :33 :03	41°52′27″N	007°56′25″E	18,6	123,6

TABLE 3.1 – *Trajectoire du navire xxxxx le 01/10/2008*

3.4.2 Résultats

Dans un soucis de confidentialité, le nom du navire ainsi que ses numéros IMO et MMSI ne sont pas donnés. Pour mettre en oeuvre les simulations, nous avons pris en compte :

- les données statiques fournies par la Lloyd's Register (nombre de changements de compagnies, le nombre de détentions, simple coque ou double coque),
- les données météorologiques de Météo France sur une journée type (02/12/2008) en fonction de la position des navires,
- les données des trajectoires fournies par DCNS (Table 3.1),
- la visibilité (prise en compte de l'heure).

Le navire utilisé pour la simulation présente aucun changement de compagnies ni de jour de détention. De plus, il est constitué d'une double coque.

Les données météorologiques sont les suivantes : « *Situation générale le mardi 02 décembre 2008 à 00h UTC et évolution : Dépression 1004 hPa sur le nord du golfe de Gênes, se comble. Haute pressions se renforçant 1025 hPa sur la péninsule ibérique. Prévisions par zones valables jusqu'au mercredi 03 décembre à 06h UTC : LIGURE, Nord de CORSE Sud-Ouest 3 ou 4 virant bientôt Nord, puis revenant Sud-Ouest à l'ouest en fin de période. Mer peu agitée, localement agitée au début. Pluie et orages. Sud de CORSE Sud-Ouest 3 ou 4 virant Nord-Ouest cet après-midi puis fraîchissant 4 ou 5 en fin de période. Mer peu agitée à agitée. Pluie et orages.* »

La provenance des navires n'étant pas connue, nous avons pris comme point de départ du tracé de la trajectoire idéale, la première coordonnée du navire. Les trois premières positions du navire correspondent à une évolution normale et suivent la trajectoire idéale. Ces positions instantanées du navire sont affichées en vert (figure 3.2).

Les quatre positions suivantes ont des positions éloignées du rail de navigation défini pour la traversée Fos - Naples par rapport au seuil fixé. L'algorithme a détecté cette anomalie de trajectoire et affiche les quatre dernières positions en rouge. Parallèlement, un message d'alarme indique pour ces positions la cause de l'alarme.

De plus, une dégradation de la météo est apparue entre la sixième et la septième position du navire. En tenant compte de la vitesse maximale du navire, du temps séparant ces deux positions et de la légère dégradation de la météo, il apparaît impossible que le navire se trouve à la septième position. L'algorithme a détecté ce problème ; un deuxième message

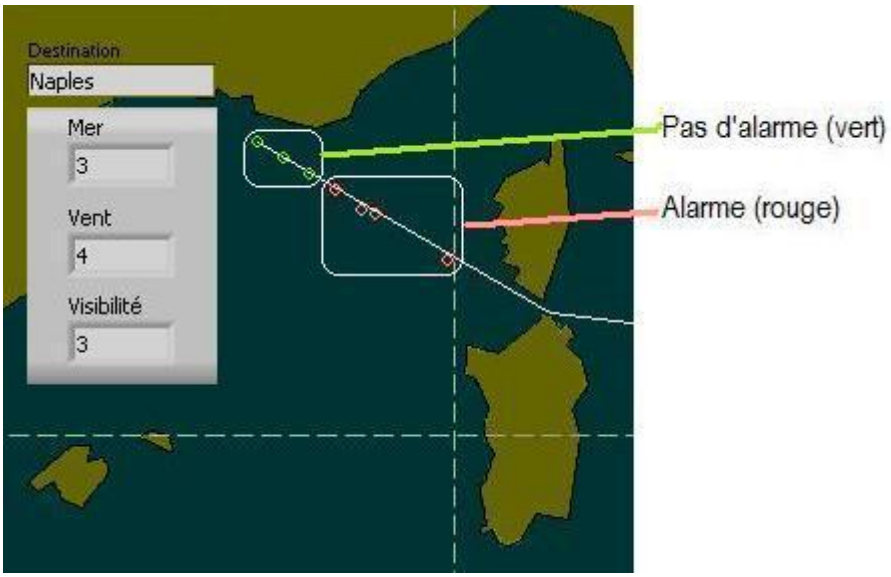


FIGURE 3.2 – *Résultats du scénario*

d'alarme associé à la dernière position s'affiche et indique : « Paramètre responsable - Trajectoire : évolution de la vitesse ».

Le simulateur développé a été livré à DCNS.

4 Les observateurs grand gain adaptatif et à entrées inconnues

Sommaire

Dans ce chapitre, nous présentons une extension des observateurs grand gain adaptatif pour les systèmes complexes. L'algorithme que nous proposons permet de développer l'observateur grand gain adaptatif tout en restant dans les coordonnées naturelles.
Une deuxième étude, présentée dans ce chapitre, concerne les observateurs à entrées inconnues pour le diagnostic de défaut.

4.1 Observateur grand-gain adaptatif

Cette étude a été publiée dans Lafont et al. (2011).

4.1.1 Système à l'étude et équations d'observateur

La forme canonique d'observabilité

Nous considérons un système non linéaire lisse de la forme :

$$\begin{array}{l} \frac{dx}{dt} = f(x,u), \\ y = h(x) = Cx, \end{array} \tag{4.1}$$

qui est transformé par un difféomorphisme ψ dans le système suivant :

$$\begin{array}{l} \frac{d\xi}{dt} = F(\xi,u) = A(t)\,\xi + b(\xi,u), \\ y = C\xi, \end{array} \tag{4.2}$$

où x, $\xi \in \mathbb{R}^n$ sont les vecteurs d'état, u est la variable de contrôle appartenant à un certain sous-ensemble limité à $\mathbb{R}^p$ et $y \in \mathbb{R}^{d_0}$ la sortie.

Des conditions d'observabilité sont nécessaires et suffisantes pour réaliser cette transformation. Ces conditions sont génériques lorsque le nombre de sorties est supérieur au nombre d'entrées.

Note : Nous avons choisi de considérer seulement une sortie linéaire, puisque cela correspond à notre cas pratique et permet une simplification des calculs. Cependant, le cas général est similaire.

Les matrices $A(t)$, C et le vecteur $b(\xi,u)$ ont les formes suivantes :

$$\begin{array}{l} A(t) = \begin{pmatrix} 0 & a_2(t) & 0 & \cdots & 0 \\ 0 & 0 & a_3(t) & \ddots & \vdots \\ \vdots & \cdots & \ddots & \ddots & 0 \\ \vdots & \cdots & \cdots & 0 & a_k(t) \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}, \\ C = (a_1(t), 0, \cdots, 0) = (Id, 0, \cdots, 0), \end{array} \tag{4.3}$$

où Id est la matrice identité de dimension d_0.

$$b(\xi,u) = \begin{pmatrix} b_1(\xi_1,u) \\ b_2(\xi_1,\xi_2,u) \\ \vdots \\ b_n(\xi_1,\cdots,\xi_n,u) \end{pmatrix}. \tag{4.4}$$

Le vecteur d'état $\xi(t)$ est supposé avoir une structure « bloc » $\xi = \left(\xi_1' \; \xi_2' \; \cdots \; \xi_n'\right)'$, où $\xi_i \in \mathbb{R}^{d_i}$ avec $d_0 \geq d_1 \geq \cdots \geq d_k$. Les matrices $a_i(t)$ ont une dimension $d_{i-1} \times d_i$ et appartiennent à un sous-ensemble compact K_i de l'ensemble des matrices $d_{i-1} \times d_i$ de rang maximum d_i.
$f(x,u)$, $a_i(t)$, $b_i(\xi,u)$ sont supposés lisses par rapport à ξ, u et t. b_i dépend de ξ de façon '« bloc » triangulaire. Ces objets sont lisses et à support compact.

Tout au long du chapitre x (resp. ξ) est appelé la coordonnée **naturelle** (resp. la coordonnée **observable**).
Les conditions de structure garantissent évidemment l'observabilité « uniforme » et l'observabilité « uniforme infinitésimale ». Les conditions de support compact peuvent être artificiellement forcées à l'extérieur du domaine « pratique » où l'état est supposé rester.

Il découle de la théorie de l'observabilité que la forme canonique avec les hypothèses de régularité associées est pertinente dans la plupart des situations. Pour tout système satisfaisant des hypothèses d'observabilité forte, les coordonnées peuvent être changées pour des « coordonnées observables » dans lesquelles la forme canonique a lieu.

TF signifie l'application tangente à $F : x \to F(x)$, $\mathbb{R}^n \to \mathbb{R}^n$ i.e. sa matrice Jacobienne dans les coordonnées. En conséquence, T^2F signifie la double tangente, une forme symétrique bilinéaire, à valeurs dans $\mathbb{R}^n$, et pour tout $u \in \mathbb{R}^n$, nous définissons la matrice $D^2F(x)\{u\}$ par $T^2F(u,v) = D^2F(x)\{u\} \cdot v$.

Nous notons L_b, la norme de la matrice Jacobienne $Tb(\xi,u)$ de $b(\xi,u)$ (i.e. $\|Tb(\xi,u)\| \leq L_b$). Puisque $b(\xi,u)$ est à support compact et u étant bornée, b est Lipschitzien par rapport à ξ uniformément en u : $\|b(\xi,u) - b(\eta,u)\| \leq L_b \|\xi - \eta\|$.

Structure de l'observateur dans les coordonnées observables

Soit Q $(n \times n)$, R $(d_0 \times d_0)$ des matrices symétriques définies positives. Soit θ le paramètre grand gain, $\theta \geq 1$. Pour $\theta = 1$, l'observateur sera juste un EKF ordinaire.
Posons $\Delta = BD\left(1, \frac{1}{\theta}, \cdots, \frac{1}{\theta^{k-1}}\right)$ la matrice diagonale bloc avec les blocs diagonaux $Id_{d_0}, \frac{1}{\theta}Id_{d_1}, \cdots$. Posons $Q_\theta = \theta\Delta^{-1}Q\Delta^{-1}, R_\theta = \theta^{-1}R$. Les équations du système dans les coordonnées observables sont :

$$\begin{array}{l} \frac{d\xi}{dt} = T\psi\left(\psi^{-1}(\xi)\right) f\left(\psi^{-1}(\xi), u\right), \\ \frac{d\xi}{dt} = F(\xi, u). \end{array} \tag{4.5}$$

$$y = C\xi. \tag{4.6}$$

Les équations pour le HG-EKF dans les coordonnées observables sont :

$$\frac{d\hat{\xi}}{dt} = F(\hat{\xi}, u) + PC'R_\theta^{-1}(y - C\hat{\xi}), \tag{4.7}$$

$$\frac{dP}{dt} = TF(\hat{\xi}, u)\ P + P\ TF(\hat{\xi}, u)' + Q_\theta - PC'R_\theta^{-1}CP. \tag{4.8}$$

Dans les coordonnées naturelles, nous avons $\hat{x} = \psi^{-1}(\hat{\xi}) = \Phi(\hat{x})$, où $\hat{x}$ représente l'estimée de x. Les équations pour le HG-EKF deviennent :

$$\frac{d\hat{x}}{dt} = f(\hat{x}, u) + pC'(\hat{x}, u)\, R_\theta^{-1}\left(y - h(\hat{x})\right), \tag{4.9}$$

$$\begin{aligned}\frac{dp}{dt} &= Tf(\hat{x},u)p + pTf(\hat{x},u)^{'} + q_\theta(\hat{x}) - pC^{'}R_\theta^{-1}Cp \\ &+T\psi(\hat{x})^{-1}D^2\psi(\hat{x})\left\{pC^{'}R_\theta^{-1}\left(h(\hat{x})-y\right)\right\}p \\ &+pD^2\psi(\hat{x})\left\{pC^{'}R_\theta^{-1}\left(h(\hat{x})-y\right)\right\}^{'}\left(T\psi(\hat{x})^{-1}\right)^{'},\end{aligned} \tag{4.10}$$

où

$$q_\theta\left(\hat{x}\right) = \left(T\psi\left(\hat{x}\right)\right)^{-1} Q_\theta\left(\left(T\psi\left(\hat{x}\right)\right)^{-1}\right)^{'}. \tag{4.11}$$

La condition que $\left(T\psi\left(\hat{x}\right)\right)^{-1}$ existe et peut être comme une condition nécessaire et suffisante d'observabilité infinitésimale.

4.1.2 Innovation

La fonction In_d introduite ci-dessous est appelée l'innovation, elle représente la mesure de qualité de l'erreur d'estimation sur un petit intervalle de temps d. La stratégie est d'adapter le paramètre grand gain θ grâce à In_d. Etant donné les propriétés d'observabilité de notre système, si l'estimée $\hat{x}$ est loin de x alors θ augmentera pour passer en mode grand gain. Inversement, si $\hat{x}$ est proche de x, l'innovation sera petite et θ décroira jusqu'à la valeur 1 (mode filtre de Kalman). Pour cela, la variable θ sera déterminée par l'équation différentielle (4.15) ci-dessous.

Soit $G_o\left(\theta\right)$ défini comme suit :

$$G_o\left(\theta\right) = \begin{cases} \frac{1}{\Delta T}\theta^2 \text{ si } \theta \leq \theta_1, \\ \frac{1}{\Delta T}\left(\theta - 2\theta_1\right)^2 \text{ si } \theta > \theta_1, \end{cases} \tag{4.12}$$

où $\theta_1 = \frac{1}{2}\theta_{max}$ et ΔT assez petit et constant.

L'innovation $In_d\left(t\right)$, avec l'horizon d'oubli d, est :

$$In_d\left(t\right) = \int_{t-d}^{t} \left\|y\left(\tau\right) - \hat{y}\left(\tau\right)\right\|^2 d\tau, \tag{4.13}$$

où $\hat{y}\left(\tau\right)$ est la prédiction de l'état initial $\hat{x}\left(t-d\right)$.

Nous définissons

$$G\left(\theta, In_d\right) = \mu\left(In_d\right) G_o\left(\theta\right) + \left(1 - \mu\left(In_d\right)\right)\lambda\left(1-\theta\right), \tag{4.14}$$

pour un $\lambda > 0$ et avec $\mu\left(In_d\right)$ une fonction lisse égale à 1 si $In_d \geq \gamma_1$, à 0 si $In_d \leq \gamma_0$, avec $0 \leq \mu\left(In_d\right) \leq 1$ pour $\gamma_0 \leq In_d \leq \gamma_1$. Un autre choix admissible pour μ est une fonction sigmoïde, $\mu : \left]-\infty;+\infty\right[\rightarrow \left]0;1\right[, \mu(In_d) = \frac{1}{1+e^{-\beta\cdot\left(In_d - m\right)}}$.

L'équation du paramètre θ est :

$$\dot{\theta} = G\left(\theta, In_d\right). \tag{4.15}$$

Les paramètres β et m de la sigmoïde jouent le même rôle que les paramètres γ_0 et γ_1. La valeur 0 de la fonction sigmoïde correspond à un « déplacement vers le mode filtre de Kalman » et la valeur 1 correspond à un « déplacement vers le mode grand gain ». Le temps de transition est contrôlé par le paramètre β (plus β est grand, plus la transition est

courte). En pratique, les meilleurs résultats sont obtenus pour un temps de transition petit i.e. une grande valeur de β.

Finallement l'observateur adaptatif dans les coordonnées originales est donné par l'ensemble des équations (4.9, 4.10, 4.13, 4.15).

Commentaire 1 : Nous pouvons résumer la méthodologie de la manière suivante :
1) Une simple équation du filtre de Kalman étendu, dépendant seulement de θ, réalise les 2 modes : $\theta = 1$ correspond au filtre de Kalman étendu ordinaire, θ grand à grand gain.
2) La garantie de convergence de l'erreur est obtenue seulement dans les coordonnées observables. Il est possible de surmonter cette difficulté en effectuant le changement de coordonnées à partir des équations (4.9, 4.10, 4.11) du filtre de Kalman étendu en coordonnées naturelles.
3) Les dynamiques du paramètre θ sont contrôlées par l'« innovation » calculée avec une fenêtre glissante. Une petite innovation signifie que l'erreur d'estimation est proche de zéro, d'où, la valeur de θ vers le mode EKF. Inversement, une grande innovation signifie une grande erreur d'estimation, d'où un « déplacement » vers un mode HG. Cela est calculé par l'équation (4.15).
4) Il est bien connu que la matrice de Riccati P est liée à la matrice d'observabilité du système linéarisé autour de la trajectoire estimée. Cela reflète l'« innovation » relative au système linéarisé. Cette « innovation linéarisée » n'est pas apparemment suffisante pour nos objectifs.
5) Garantie de convergence : pour toutes caractéristiques de bruit (Q, R) (dépendant du bruit du mode EKF), les paramètres (θ, m, d, β) peuvent être choisis de façon à ce que la convergence exponentielle arbitraire globale peut être réalisée : soit $\|\epsilon\| \leq e^{-\alpha(t-T^*)} \times \|\epsilon_0\|$, ($\alpha$ arbitraire, T^* arbitraire).
Commentaire 2 : Grâce à (4.13) le système observateur (4.9, 4.10, 4.13, 4.15) n'est pas un système d'équations différentielles ordinaires (ODE). Cependant l'existence et l'unicité des solutions sont garanties.

4.1.3 L'intérêt des coordonnées naturelles et de l'observateur en cascade

Il est évident que le changement de variables $\psi(x)$ n'est pas facile à appliquer. C'est la raison pour laquelle nous avons choisi de travailler dans les coordonnées naturelles. Dans ces équations naturelles pour notre observateur, il est assez difficile de calculer la Jacobienne inverse $D\psi(\hat{x})^{-1}$. Pour notre application ci-dessous, c'est le cas des équations du modèle complet. C'est pourquoi nous avons choisi (dans les coordonnées naturelles) la stratégie suivante.
Nous appliquons d'abord notre observateur au modèle simplifié (cinq états, trois sorties). Nous utilisons l'estimée de ce premier observateur comme l'entrée de l'observateur complet. De cette façon, le calcul des deux matrices Jacobiennes inverses est facilité (figure 4.1).

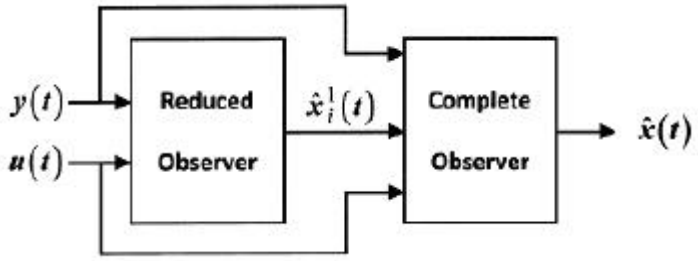


FIGURE 4.1 – *Observateur en cascade*

4.1.4 Application

Le processus considéré est une installation de traitement des eaux usées de petite taille composée d'un bassin unique d'aération équipé d'aérateurs de surface qui fournissent l'oxygène et permettent de mixer les eaux usées influentes avec la biomasse (figure 4.2).
Ici, nous nous intéressons à l'estimation en-ligne de la qualité de l'effluent.

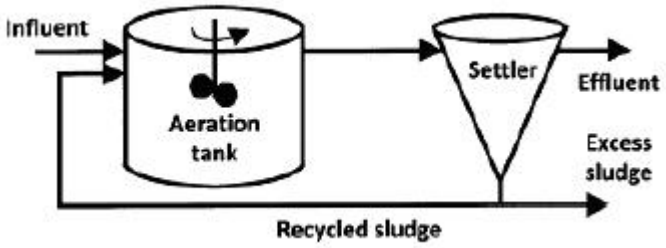


FIGURE 4.2 – *Installation du traitement des eaux usées*

Une directive de l'Union Européenne a fixé les concentrations polluantes maximales permises dans l'effluent pour les stations d'épuration de petites tailles : la demande biochimique en oxygène sur une période de cinq jours $DBO_5 < 25\ mg/l$, la demande chimique en oxygène $DCO < 125\ mg/l$ et les matières en suspension $MES < 35\ mg/l$. Ces trois quantités sont définies ci-dessous avec des termes de l'état du modèle.
Le modèle utilisé est basé sur le modèle « Activated Sludge Model N°1 » (ASM 1). Alors notre modèle de biodégradation a 12 variables d'état (Table 4.1) : actuellement, nous considérons seulement la biodégradation, les variables d'état décrivant l'alcalinité totale ne sont pas prises en compte.

Les trois exigences de qualité caractérisant les effluents sont définies par :

$$\begin{aligned} DBO_5 &= 0.25(S_S + X_S + (1 - f_p)(X_{BH} + X_{BA})), \\ DCO &= S_S + S_I + X_S + X_I + X_{BH} + X_{BA} + X_P, \\ MES &= 0.75(X_S + X_I + X_{BH} + X_{BA} + X_P). \end{aligned} \tag{4.16}$$

Remarque 1 : Les valeurs des paramètres stoechiométriques et cinétiques considérés sont listées dans les Tables 4.2 et 4.3. L'ensemble complet des équations et les conditions influentes peuvent être trouvées sur le site web de « the International Water Association task group on benchmarking of control strategies » pour les systèmes de traitement des eaux usées.

Définition	Notation
1. Soluble inert organic matter $(g\,COD.m^{-3})$	S_I
2. Readily biodegradable substrate $(g\,COD.m^{-3})$	S_S
3. Particulate inert organic matter $(g\,COD.m^{-3})$	X_I
4. Slowly biodegradable substrate $(g\,COD.m^{-3})$	X_S
5. Active heterotrophic biomass $(g\,COD.m^{-3})$	$X_{B,H}$
6. Active autotrophic biomass $(g\,COD.m^{-3})$	$X_{B,A}$
7. Particulate products arising from biomass decay $(g\,COD.m^{-3})$	X_P
8. Oxygen $(g\,COD.m^{-3})$	S_O
9. Nitrate and nitrite nitrogen $(g\,N.m^{-3})$	S_{NO}
10. $NH_4^+ + NH_3$ nitrogen $(g\,N.m^{-3})$	S_{NH}
11. Soluble biodegradable organic nitrogen $(g\,N.m^{-3})$	S_{ND}
12. Particulate biodegradable organic nitrogen $(g\,N.m^{-3})$	X_{ND}

TABLE 4.1 – *Liste des variables*

Paramètre	Unité	Valeur
Y_A	$g\ cell\ COD\ formed\ (g\ N\ oxidized)^{-1}$	0.24
Y_H	$g\ cell\ COD\ formed\ (g\ COD\ oxidized)^{-1}$	0.67
f_p	*dimensionless*	0.08
i_{XB}	$g\ N\ (g\ COD)^{-1}$ *in biomass*	0.08
i_{XP}	$g\ N\ (g\ COD)^{-1}$ *in particulate products*	0.06

TABLE 4.2 – *Paramètres Stoechiométriques*

Paramètre	Unité	Valeur
μ_H	d^{-1}	4.0
K_S	$g\ COD\ m^{-3}$	10.0
$K_{O,H}$	$g\ COD\ m^{-3}$	0.2
K_{NO}	$g\ NO_3 - N\ m^{-3}$	0.5
b_H	d^{-1}	0.3
$\eta_{NO,g}$	*dimensionless*	0.8
$\eta_{NO,h}$	*dimensionless*	0.8
k_h	$(g\ cell\ COD\ d)^{-1}$	3.0
K_X	$(g\ cell\ COD)^{-1}$	0.1
μ_A	d^{-1}	0.5
$K_{NH,A}$	$g\ NH_3 - N\ m^{-3}$	1.0
b_A	d^{-1}	0.05
$K_{O,A}$	$g\ COD\ m^{-3}$	0.4
k_a	$m^3\ (g\ COD\ d)^{-1}$	0.05

TABLE 4.3 – *Paramètres cinétiques*

Remarque 2 : Quant au modèle simplifié, nous pouvons fournir **de manière explicite toutes les formules et les valeurs** des constantes et des fonctions cinétiques.

Les hypothèses sur le modèle sont les suivantes :
- le réacteur est bien mixé,
- la séparation des phases liquide et solide est parfaite et aucune réaction apparaît dans le clarificateur (settler),
- la somme de tous les débits du settler est égale au débit influent dans le settler.

Le modèle est de la forme $\dot{x} = f(x, u)$, où la commande u représente l'état u_b des turbines et la valeur Q^{in} le débit influent. L'entrée u_b dans (2.5) est une séquence binaire commutant entre 0 et 1 et représentant l'état des turbines (off/on) qui aèrent le système.
Les coordonnées naturelles sont les concentrations des espèces, i.e. tous les composants x_i du vecteur d'état sont les concentrations listées en Table 3.1. Chaque équation est basée sur l'équilibre des masses, incluant la dégradation cinétique. Les composants f_i des dynamiques sont définis comme suit :

- pour les composants solubles (i= 1, 2, 9, 10, 11)

$$f_i(x) = \frac{Q^{in}}{V}\left(x_i^{in} - x_i\right) + r_i(x) \tag{4.17}$$

- pour les composants particulaires (i= 3, 4, 5, 6, 7, 12)

$$f_i(x) = \frac{1}{V}\left[Q^{in}\left(x_i^{in} - x_i\right) + Q^{rs}\frac{Q^{in} - Q^w}{Q^{rs} + Q^w}x_i\right] + r_i(x) \tag{4.18}$$

- pour la concentration d'oxygène dissous (i= 8)

$$f_8(x) = \frac{Q^{in}}{V}\left(x_8^{in} - x_8\right) + r_8(x) + u_b k_L a\left(S_O^{max} - S_O\right) \tag{4.19}$$

où $r_i(x)$, $i = 1, \cdots, 12$ sont des fonctions non linéaires. Elles représentent les taux de réaction dépendants des taux cinétiques de dégradation des composants.

$$\begin{aligned}
r_1 &= 0\\
r_2 &= -\tfrac{1}{Y_H}\rho_1 - \tfrac{1}{Y_H}\rho_2 + \rho_7\\
r_3 &= 0\\
r_4 &= \left(1 - f_p\right)\rho_4 + \left(1 - f_p\right)\rho_5 - \rho_7\\
r_5 &= \rho_1 + \rho_2 - \rho_4\\
r_6 &= \rho_3 - \rho_5\\
r_7 &= f_p\rho_4 + f_p\rho_5\\
r_8 &= -\tfrac{1-Y_H}{Y_H}\rho_1 - \tfrac{4.57-Y_A}{Y_A}\rho_3\\
r_9 &= -\tfrac{1-Y_H}{2.86Y_H}\rho_2 + \tfrac{1}{Y_A}\rho_3\\
r_{10} &= -i_{XB}\rho_1 - i_{XB}\rho_2 - \left(i_{XB} + \tfrac{1}{Y_A}\right)\rho_3 + \rho_6\\
r_{11} &= -\rho_6 + \rho_8\\
r_{12} &= \left(i_{XB} - f_p i_{XP}\right)\rho_4 + \left(i_{XB} - f_p i_{XP}\right)\rho_5 - \rho_8
\end{aligned} \tag{4.20}$$

$$\begin{aligned}
\rho_1 &= \mu_H\left(\tfrac{S_S}{K_S+S_S}\right)\left(\tfrac{S_O}{K_{O,H}+S_O}\right)X_{B,H}\\
\rho_2 &= \mu_H\left(\tfrac{S_S}{K_S+S_S}\right)\left(\tfrac{K_{O,H}}{K_{O,H}+S_O}\right)\left(\tfrac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_{NO,g}X_{B,H}\\
\rho_3 &= \mu_A\left(\tfrac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\tfrac{S_O}{K_{O,A}+S_O}\right)X_{B,A}\\
\rho_4 &= b_H X_{B,H}\\
\rho_5 &= b_A X_{B,A}\\
\rho_6 &= k_a S_{ND} X_{B,H}\\
\rho_7 &= k_h\tfrac{X_S/X_{B,H}}{K_X+X_S/X_{B,H}}\left[\left(\tfrac{S_O}{K_{O,H}+S_O}\right) + \eta_{NO,h}\left(\tfrac{K_{O,H}}{K_{O,H}+S_O}\right)\left(\tfrac{S_{NO}}{K_{NO}+S_{NO}}\right)\right]X_{B,H}\\
\rho_8 &= \rho_7\tfrac{X_{ND}}{X_S}
\end{aligned} \tag{4.21}$$

Remarque 3 : Les variables S_I, X_I et X_P, décrites avec les équations correspondant à $i = 1,\ 3,\ 7$, n'apparaissent pas dans les autres équations. De ce fait, ces variables ne sont pas observables, et nous ne pouvons pas faire mieux qu'une **simple prédiction** pour celles-ci. Alors, la dimension de l'espace d'état est $n = 9$.
La constante $k_L a$ est le coefficient d'oxygène $\left(k_L a = 10\ h^{-1}\right)$ et S_O^{max} la concentration de saturation de l'oxygène dissous $\left(S_O^{max} = 8\ mgl^{-1}\right)$.
Le volume du bassin d'aération est $V = 6000\ m^3$. Le settler est un

réservoir cylindrique où les boues sont renvoyées au bassin d'aération $\left(Q^{rs} = 18446\ m^3 jour^{-1}\right)$ ou retirées du système $\left(Q^{w} = 385\ m^3 jour^{-1}\right)$ pour incinération ou décharge.

Nous supposons raisonnable de disposer de trois mesures : S_O, S_{NO} et S_{NH} situées dans le bassin d'aération. Bien que le système de traitement des eaux usées avec ces trois mesures soit observable, nous utilisons d'abord un modèle simplifié de dimension plus petite développé.

Le modèle réduit

Il est procédé comme suit :

- Il regroupe les variables S_S et X_S dans une seule X_{DCO} (DCO pour « demande chimique en oxygène »), $X_{DCO} = S_S + X_S$.
- Il est bien connu que les dynamiques de X_{BH}, X_{BA} et X_{ND} sont lentes par rapport aux autres. Alors, elles sont supposées être constantes. De ce fait, les variables $\alpha_i,\ i = 1,..,8$ définies ci-dessous sont des constantes. Il est aussi communément accepté que les ratios $\frac{X_{ND}}{X_S}$, $\frac{X_{DCO}}{S_S}$, $\frac{X_{DCO}}{X_S}$ varient doucement. Comme conséquence, les variables α_9, K_{DCO} et K_{ND} sont aussi supposées comme des constantes.

En retirant les trois variables non observables X_P, X_I et S_I, nous aboutissons au modèle simplifié avec 5 variables d'état $S_O, S_{NO}, S_{NH}, X_{DCO}, S_{ND}$ et avec trois grandeurs observables S_O, S_{NO}, S_{NH}. Toutes ces simplifications aboutissent au modèle réduit suivant :

$$\begin{aligned}\dot{S_O} &= \tfrac{Q^{in}}{V}\left(S_O^{in} - S_O\right) + \alpha_1 \tfrac{X_{DCO}}{K_{DCO}+X_{DCO}} \\ &\cdot \tfrac{S_O}{K_{O,H}+S_O} + \widetilde{r_1}\left(y\right) + u_b k_L a\left(S_O^{max} - S_O\right)\end{aligned} \tag{4.22}$$

$$\begin{aligned}\dot{S_{NO}} &= \tfrac{Q^{in}}{V}\left(S_{NO}^{in} - S_{NO}\right) + \alpha_3 \tfrac{X_{DCO}}{K_{DCO}+X_{DCO}} \\ &\cdot \tfrac{K_{O,H}}{K_{O,H}+S_O} \tfrac{S_{NO}}{K_{NO}+S_{NO}} + \widetilde{r_2}\left(y\right)\end{aligned} \tag{4.23}$$

$$\begin{aligned}\dot{S_{NH}} &= \tfrac{Q^{in}}{V}\left(S_{NH}^{in} - S_{NH}\right) + \alpha_5 \tfrac{X_{DCO}}{K_{DCO}+X_{DCO}} \\ &\cdot \left(\tfrac{S_O}{K_{O,H}+S_O} + \eta_{NO,g} \tfrac{K_{O,H}}{K_{O,H}+S_O} \tfrac{S_{NO}}{K_{NO}+S_{NO}}\right) \\ &+ \widetilde{r_3}\left(y\right) + \alpha_6 S_{ND}\end{aligned} \tag{4.24}$$

$$\begin{aligned}\dot{X_{DCO}} &= \tfrac{Q^{in}}{V}\left(X_{DCO}^{in} - \tfrac{K_S}{K_{DCO}} X_{DCO}\right) \\ &+ \alpha_7 \tfrac{X_{DCO}}{K_{DCO}+X_{DCO}} \left(\tfrac{S_O}{K_{O,H}+S_O}\right. \\ &\left. + \eta_{NO,g} \tfrac{K_{O,H}}{K_{O,H}+S_O} \tfrac{S_{NO}}{K_{NO}+S_{NO}}\right) + \alpha_8\end{aligned} \tag{4.25}$$

$$\begin{aligned}\dot{S_{ND}} &= \tfrac{Q^{in}}{V}\left(S_{ND}^{in} - S_{ND}\right) - \alpha_6 S_{ND} + \alpha_9 \\ &\cdot \tfrac{X_{DCO}}{K_{ND}+X_{DCO}} \left(\tfrac{S_O}{K_{O,H}+S_O} + \eta_{NO,h} \tfrac{K_{O,H}}{K_{O,H}+S_O}\right. \\ &\left. \cdot \tfrac{S_{NO}}{K_{NO}+S_{NO}}\right)\end{aligned} \tag{4.26}$$

Les paramètres α_1, α_2, α_3, α_4, α_5, α_6, α_7, α_8, α_9, K_{ND} et K_{DCO} sont définis comme suit et sont donnés dans la Table 4.4. Les valeurs des concentrations influentes sont listées en Table 4.5.

Coefficient	Valeur
α_1	- 5892
α_2	- 875
α_3	- 1648
α_4	191
α_5	- 957
α_6	150
α_7	- 17855
α_8	830
α_9	561
K_{DCO}	574
K_{ND}	296

TABLE 4.4 – *Coefficients constants*

Concentration	Valeur
X^{in}_{DCO}	271.82 $g\ COD\ m^{-3}$
S^{in}_{NO}	0 $g\ COD\ m^{-3}$
S^{in}_{NH}	31.56 $g\ COD\ m^{-3}$
S^{in}_{ND}	6.95 $g\ COD\ m^{-3}$
S^{in}_{O}	0 $g\ COD\ m^{-3}$

TABLE 4.5 – *Concentrations influentes*

$$\begin{aligned}
\alpha_1 &= -\tfrac{1-Y_H}{Y_H}\mu_H X_{B,H} \\
\alpha_2 &= -4.57\tfrac{\mu_A}{Y_A} X_{B,A} \\
\alpha_3 &= -\tfrac{1-Y_H}{2.86 Y_H}\mu_H X_{B,H}\eta_{NO,g} \\
\alpha_4 &= \tfrac{\mu_A}{Y_A} X_{B,A} \\
\alpha_5 &= -i_{XB}\mu_H X_{B,H} \\
\alpha_6 &= k_a X_{B,H} \\
\alpha_7 &= -\tfrac{1}{Y_H}\mu_H X_{B,H} \\
\alpha_8 &= \left(1-f_p\right)\left(b_H X_{B,H} + b_A X_{B,A}\right) \\
\alpha_9 &= k_h \tfrac{X_{ND}}{X_S} X_{B,H}
\end{aligned} \tag{4.27}$$

$$\begin{aligned}
K_{DCO} &= K_S \tfrac{X_{DCO}}{S_S} \\
K_{ND} &= K_X \tfrac{X_{DCO}}{X_S} X_{B,H}
\end{aligned} \tag{4.28}$$

$$\begin{aligned}
\widetilde{r_1}\left(y\right) &= \alpha_2 \tfrac{S_{NH}}{K_{NH,A}+S_{NH}} \tfrac{S_O}{K_{O,A}+S_O} \\
\widetilde{r_2}\left(y\right) &= \alpha_4 \tfrac{S_{NH}}{K_{NH,A}+S_{NH}} \tfrac{S_O}{K_{O,A}+S_O} \\
\widetilde{r_3}\left(y\right) &= -\alpha_4 \tfrac{S_{NH}}{K_{NH,A}+S_{NH}} \tfrac{S_O}{K_{O,A}+S_O}
\end{aligned} \tag{4.29}$$

Analyse d'observabilité Pour le modèle simplifié (4.22 - 4.26), avec les sorties S_O, S_{NO} et S_{NH}, le domaine « pratique » est $(\mathbb{R}_+)^5$, l'orthant positif dans $(\mathbb{R})^5$.
Les points où S_O et S_{NO} sont tous les deux à zéro peuvent apparaître en pratique, cela est appelé le « travail anaérobique ». Ce type de fonctionnement reste néanmoins temporaire et n'est pas trés fréquent car la période off de l'aérateur est limitée par la contrainte opérative $t^{off}_{max} = 120\ min$.

Paramètre	observateur réduit	observateur complet
θ_{max}	20	10
β	$1664\frac{\pi}{e}$	$1664\frac{\pi}{e}$
m	2	40
ΔT	0.01	0.01
λ	200	200
d	0.1	0.1

TABLE 4.6 – *Paramètres pour l'adaptation*

Dans ce cas, la variable X_{DCO} n'a pas d'influence sur les sorties et donc le système n'est pas observable, la meilleure stratégie possible est une simple prédiction.
Sur $\mathfrak{D} \subset (\mathbb{R}_+)^5$, $\mathfrak{D} = \{S_O \neq 0 \text{ ou } S_{NO} \neq 0\}$, le fait que la matrice $a_2(t)$ soit de rang deux sur ce sous domaine $\mathfrak{D}$ permet de vérifier que le système est observable uniformément et observable uniformément infinitésimalement.

Changement de variables Le changement de variables Ψ qui lie les coordonnées naturelles aux coordonnées de l'observateur est classique. Il consiste juste à poser :

$$\widetilde{X_{DCO}} = \frac{X_{DCO}}{K_{DCO} + X_{DCO}}. \tag{4.30}$$

Observateur pour le modèle complet

L'analyse d'observabilité du système complet avec l'estimée provenant de l'observateur réduit est « triviale ». Nous aboutissons aux conclusions similaires d'observabilité uniforme et d'observabilité infinitésimale uniforme.
Dans ce cas l'état est de dimension 9 et la sortie a une dimension 6. Actuellement, les variables X_S et S_S regroupées dans X_{DCO}, peuvent être trouvées à partir du modèle réduit. Cela provient du fait que nous supposons que K_{DCO} est une constante (4.28) :

$$\frac{X_{DCO}}{S_S} = 1 + \frac{X_S}{S_S} = \frac{K_{DCO}}{K_S}. \tag{4.31}$$

Le changement de variables ψ des coordonnées naturelles aux coordonnées observables est le suivant :
$x = (S_O\ S_{NO}\ S_{NH}\ S_S\ X_S\ S_{ND}\ X_{BH}\ X_{BA}\ X_{ND})'$ est changé pour
$\xi = (S_O\ S_{NO}\ S_{NH}\ S_S\ X_S\ S_{ND}\ r_8\ r_9\ r_{11})'$.

Choix des paramètres pour l'innovation

Le choix des paramètres (θ_{max}, β, m, ΔT, λ, d) pour l'adaptation de l'innovation pour notre application est présenté dans la Table 4.6.

Conditions pour une simulation réaliste

Concentrations entrantes Dans un but de faire des simulations réalistes, les concentrations influentes, données dans la Table 4.5, ne peuvent pas être considérées comme constantes. Nous avons modélisé les variations de ces concentrations avec un bruit additif. En pratique, étant donnée la longueur des canalisations (plusieurs kilomètres), ces perturbations sont plutôt lentes. Cependant, nous avons volontairement choisi des dynamiques rapides pour ces bruits. Un exemple de ces variations est donné en figure 4.3.

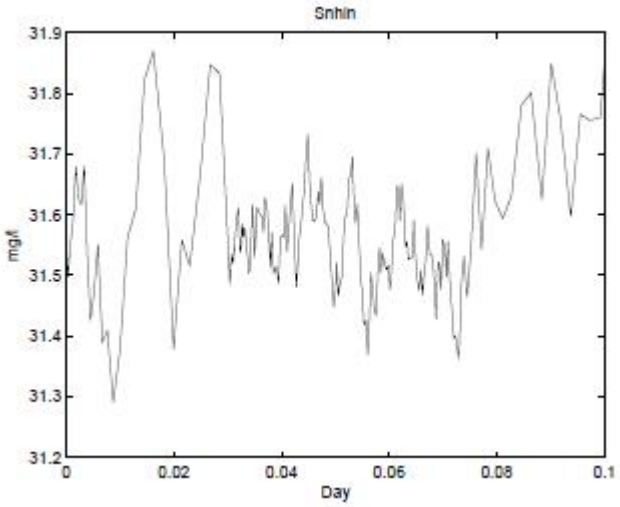


FIGURE 4.3 – *Variation de* S_{NH}^{in}

Désadaptation des taux cinétiques et des coefficients stoechiométriques Ces paramètres ne sont pas bien connus en pratique et peuvent être sujets à de larges variations. Nous avons considéré simultanément, pour chaque taux de réaction, une désadaptation périodique d'amplitude 20%. De plus, ces adaptations sont complètement asynchrones. Nous considérons 3 pèriodes sur les 14 jours avec un déphasage uniformément réparti sur les 8 taux de réaction.

Ces conditions de simulation sont problablement plus mauvaises que celles qui peuvent apparaitre dans la pratique.

4.1.5 Résultats

Comme accepté habituellement, toutes les simulations sont données avec les sorties perturbées par un bruit réaliste additif de type Orstein-Uhlenbeck. La commande alternative u_b a été choisie comme en pratique : « on » pendant 15 minutes et « off » pendant 5 minutes. Notre fichier de simulation (temps sec) couvre 14 jours et la valeur du débit entrant Q^{in} provient du fichier du benchmark.
Pour évaluer les performances de notre observateur, nous comparons un observateur Luenberger, un EKF ordinaire et notre grand gain adaptatif.

Variable	Luenberger		EKF		HG-EKF	
	m	σ	m	σ	m	σ
S_{ND}	-0.02	0.07	-0.02	0.07	-0.01	0.08
S_S	-0.03	0.11	-0.03	0.11	-0.02	0.10
X_S	0.30	7.85	0.28	7.86	-0.21	7.59
X_{BH}	96.17	81.39	96.68	81.29	41.64	85.52
X_{BA}	-8.29	4.50	-7.20	4.41	-5.89	3.14
X_{ND}	0.02	0.61	0.02	0.61	-0.02	0.60

TABLE 4.7 – *Comparaions entre Luenberger, EKF et adaptatif HG-EKF*

Variable	Luenberger		EKF		HG-EKF	
	m	σ	m	σ	m	σ
S_I	-0.03	0.12	-0.03	0.12	-0.03	0.12
X_I	3.93	1.89	3.94	1.88	3.76	1.84
X_P	28.48	8.42	28.51	8.40	20.29	7.82

TABLE 4.8 – *Comparaions entre Luenberger, EKF et adaptatif HG-EKF*

Aucune comparaison n'est montrée avec un ordinaire HG-EKF (non adaptatif) : dans ce cas les résultats sont plûtot mauvais, l'observateur étant trés sensible au bruit.
Les moyennes et les écart-types de l'erreur d'estimation sont calculés sur les 14 jours. Cependant, les figures présentées montrent seulement les 3 premiers jours, où l'effet des conditions initiales inconnues est significative.

La table 4.7 montre clairement une amélioration pour l'adaptatif HG-EKF.

Reconstruction des variables X_I, S_I, X_P

Comme nous l'avons dit ces variables non observables sont reconstruites par simple prédiction. Les résultats sont montrés en Table 4.8.

Qualité de l'effluent

Pour valider la méthode et estimer les sorties de l'effluent, nous simulons le clarificateur complet. Ce modèle simule les profils des composés solides à travers le settler. Les comparaisons de ces trois exigences de qualité avec leurs estimées sont présentées en Table 4.9.

La figure 4.4 montre les variables de sortie DBO_5, DCO, MES et leurs estimées pendant 3 jours. L'effet du grand gain sur la réponse est trés clair. L'erreur converge rapidement vers 0, ce qui n'est pas le cas du EKF où une erreur significative existe.

Variable	Luenberger		EKF		HG-EKF	
	m	σ	m	σ	m	σ
DBO_5	0.04	0.04	0.04	0.04	0.01	0.04
DCO	0.22	0.15	0.22	0.15	0.09	0.15
MES	0.18	0.13	0.19	0.13	0.09	0.10

TABLE 4.9 – *Qualité de l'effluent*

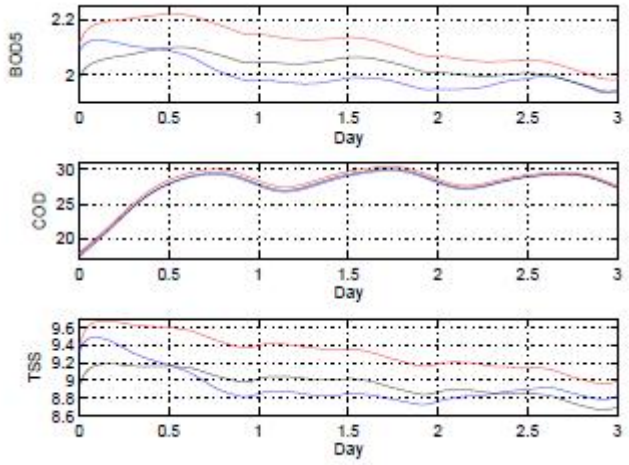


FIGURE 4.4 – *Qualité de l'effluent (noir : modèle - rouge : EKF - bleu : Adaptatif HG-EKF)*

La méthode proposée ici pour la reconstruction de l'état d'un système de traitement des eaux usées semble être une réelle amélioration par rapport aux méthodes classiques. Elle a deux intérêts techniques : premièrement, l'implémentation du HG-EKF adaptatif n'est pas trés compliquée grâce à l'utilisation des coordonnées naturelles qui simplifie fortement les calculs. Deuxièmement, l'utilisation d'un observateur en cascade apporte des calculs raisonnables pour le modèle complet.

4.2 Observateur à entrées inconnues pour le diagnostic

Cette étude a été publiée dans Methnani et al. (2011). Nous proposons une méthodologie générale pour identifier et reconstruire des défauts capteurs sur un procédé dynamique. La méthodologie est inspirée de la théorie d'identification générale : cette théorie d'identification fournit un cadre général pour le problème « d'observabilité avec entrées inconnues ». En effet, de nombreux problèmes de détection de défauts peuvent être formulés comme des problèmes d'observabilité, les défauts (éventuellement additifs) étant juste considérés comme entrées inconnues. L'estimation d'états et la détection et l'isolation de défauts (FDI) constituent le sujet de l'étude.

L'objectif principal d'une FDI est de détecter le défaut s'il apparaît en générant une alarme, mais aussi en identifiant la nature et la localisation du défaut. Un défaut est un mauvais fonctionnement d'un actionneur ou d'un capteur, ou plus généralement de variables d'état internes du système. Ces mauvais fonctionnements surviennent à cause de certaines circonstances anormales. S'ils ne sont pas détectés, des déviations non-admissibles peuvent causer de gros dégats.

Les méthodes de FDI à base d'observateurs suscitent un grand intérêt dans la communauté. Dans cette catégorie à base de modèles, des résidus sont calculés comme la différence entre les sorties mesurées et les sorties estimées par le modèle. A l'aide de ces résidus, une décision est prise pour signaler un défaut ou non. Une difficulté est de faire un observateur

robuste par rapport aux perturbations qui ne doivent pas être considérées comme des défauts.

Dans cette étude où les systèmes nonlinéaires (lisses) continus en représentation d'état sont considérés, nous proposons une méthodologie systématique dédiée à la reconstruction de défauts. Via cette méthode, il est possible de détecter un défaut dérive ou intermittent d'un capteur. Nous faisons l'hypothèse raisonnable que plusieurs défauts ne peuvent pas apparaitre simultanément, i.e. nous nous résumons au problème d'observabilité avec une seule fonction d'entrée inconnue. La théorie est tirée de la théorie d'identfication et cela conduit naturellement à l'utilisation d'observateurs grand gain.

4.2.1 Généralités

Le concept d'« observabilité à entrées inconnues » peut être vu comme simplement une reformulation du concept d'identification. Dans le contexte de FDI, une simple entrée inconnue correspond à un défaut simple.

Le concept de généricité considéré est par rapport à la topologie de Whitney. Dans la plupart des cas, les problèmes sont situés sur un sous-ensemble compact de l'espace d'état, il suffit en pratique de considérer la topologie métrique C^∞ : une fonction est proche de zéro si ces valeurs avec les valeurs de ses dérivés sont suffisamment petites.

4.2.2 Définitions et systèmes considérés

Les systèmes considérés sont des systèmes (C^ω or C^∞) de la forme :

$$\Sigma \begin{cases} \frac{dx}{dt} = f(x, \varphi(t)) \\ y = h(x, \varphi(t)) \end{cases} \tag{4.32}$$

Où l'état $x = x(t)$ est de dimension n, $x(0) = x_0$. L'observation y est à valeurs dans $\mathbb{R}^{d_y}$ et f, h sont respectivement un champ de vecteur lisse (paramétré) et une fonction lisse. La fonction φ (l'entrée inconnue) est une fonction du temps (dans le contexte d'identificabilité, c'est une fonction inconnue de l'état). Pour simplifier, chaque trajectoire est supposée être définie sur l'intervalle $\left[0, T_{x_0,\varphi}\right[$ dépendant des conditions initiales et de la fonction inconnue φ, mais contenant un intervalle de temps fixe $I = [0, i]$.

L'objectif est d'estimer à la fois la variable d'état x et la fonction inconnue φ: $\mathbb{R}^+ \to \mathbb{R}$. L'inconnue φ sera notée par d.

Soit $\Omega = X \times L^\infty[I]$, où $L^\infty[I]$ est l'ensemble des fonctions mesurables bornées à valeurs dans $\mathbb{R}$ définies sur I, et par $L^\infty[\mathbb{R}^{d_y}]$ l'ensemble des fonctions bornées mesurables de I vers $\mathbb{R}^{d_y}$.

Ensuite, nous pouvons définir l'application entrées/sorties P_Σ qui applique l'état initial x_0 et la fonction d'entrée $\widehat{\varphi}$ vers la fonction de sortie

y :

$$P_\Sigma : \begin{array}{c} \Omega \to L^\infty[\mathbb{R}^{d_y}] \\ (x_0, \widehat{\varphi}(.)) \to y(.) \end{array} \tag{4.33}$$

4.2.3 Le cas générique 3-5

Le cas 3-sorties 5-états est le cas générique le plus simple. Il a la bonne propriété d'être naturellement sous la forme canonique usuelle dès que les sorties sont des états, ce qui est souvent le cas.

Nous prenons un système de la forme :

$$Y = (y_1, y_2, y_3) = (x_1, x_2, x_3),\ x = (x_1, \ldots, x_5),\ \dot{x}(t) = f(x)$$

Nous souhaitons réaliser la reconstruction de défauts pour un défaut additif $d(t)$ sur la première sortie, i.e. en fait, $y_1(t) = x_1(t) + d(t)$. Posons $z_1(t) = x_1(t) + d(t),\ z_2(t) = x_2(t), \ldots,\ z_5(t) = x_5(t)$, le système peut être réécrit comme :

$$\begin{array}{l} y_1(t) = z_1(t), y_2(t) = z_2(t), y_3(t) = z_3(t), \\ \dot{z}_1(t) = f_1(z_1(t) - d(t), z_2(t), \ldots, z_5(t)) + \dot{d} \\ \dot{z}_i(t) = f_i(z_1(t) - d(t), z_2(t), \ldots, z_5(t)), i = 2, \cdots 5. \end{array} \tag{4.34}$$

où :

$$\dot{z} = g(z, d, \dot{d}) \tag{4.35}$$

La stratégie la plus basique

Une façon simple de procéder est de supposer que $\dot{d} = 0$. Nous obtenons une équation 6-états de la forme :

$$\begin{array}{r} \dot{z}(t) = g(z_1(t), z_2(t), \ldots, z_5(t), d) \\ \dot{d} = 0 \end{array} \tag{4.36}$$

où, en posant $Z = (z, d)$,

$$\begin{array}{r} \dot{Z} = G(Z) \\ y = (Z_1, Z_2, Z_3) \end{array} \tag{4.37}$$

Un échelon sur d correspond exactement à un saut (peut-être grand) de l'état Z dans le modèle (4.37).
Dans ce cas, un observateur grand gain fait le travail de reconstruction : il a précisement la propriété de récupérer arbitrairement les changements rapides dans les conditions initiales.

Le système (4.37) est un système général 6-états 3-sorties qui est observable :

En effet, en général (pour un système générique), la matrice 3×3 formée par les lignes :

$$(\frac{\partial G_i}{\partial z_4}, \frac{\partial G_i}{\partial z_5}, \frac{\partial G_i}{\partial d}), i = 1, \ldots, 3 \tag{4.38}$$

est inversible, ce qui entraine par le théorème des fonctions implicites à paramètres (en gelant les variables z_1, z_2, z_3) que l'application $\tilde{G} = (G_1(z_4, z_5, d), G_2(z_4, z_5, d), G_3(z_4, z_5, d))$ a une inverse $\tilde{G}_1$.

Connaissant la sortie $Y(t) = (z_1(t), z_2(t), z_3(t))$ et sa dérivée, nous obtenons $(\dot{z}_1(t), \dot{z}_2(t), \dot{z}_3(t)) = \tilde{G}(z_4(t), z_5(t), d(t))$, qui peut être inversé pour chaque valeur de $z_1(t), z_2(t), z_3(t)$, et nous obtenons la connaissance de $z_4(t), z_5(t), d(t)$.

La stratégie générale

Une stratégie plus générale est d'utiliser un **modèle local** pour le défaut $d(t)$. Par exemple, un simple modèle local est $d^{(k)} = 0$. La question n'est pas que ce polynome modélise la fonction d globallement comme une fonction de t, mais seulement localement, sur des intervalles de temps raisonnables (raisonnables par rapport aux performances demandées pour la reconstruction entrée-état).

Ici, nous sommes dans la situation générale d'un système 6+k-états, 3-sorties. Le fait que le système original est infinitésimallement observable-inconnu implique que le système 6+k étendu peut être mis sous une certaine forme normale d'observabilité approprié.
Pour cette forme normale, l'utilisation des dérivées approximées permettrait la reconstruction de l'état.

4.2.4 Résultats

Nous avons testé cette méthode sur le modèle de station d'épuration. Nous avons pris le modèle réduit.

Défaut échelon

Au second jour, un défaut échelon est appliqué au capteur S_{NO} (figure 4.5). L'amplitude est égal à 2 mg/l (comparée à une valeur moyenne de 6 mg/l). Les 3 variables d'état S_{NO}, S_{NH} et S_O sont mesurées.

Défaut dérive et intermittent

Dans un but de valider complétement la méthode, il a été intéressant de considérer, après l'échelon, les types les plus classiques de dysfonctionnement : défaut dérive et défaut intermittent. Les résultats de simulations sont montrés respectivement sur les figures 4.6 et 4.7.

Sur ces 2 figures, on peut voir que la méthode reconstitue la forme et l'amplitude du défaut avec une bonne fidélité, mis à part le bruit des

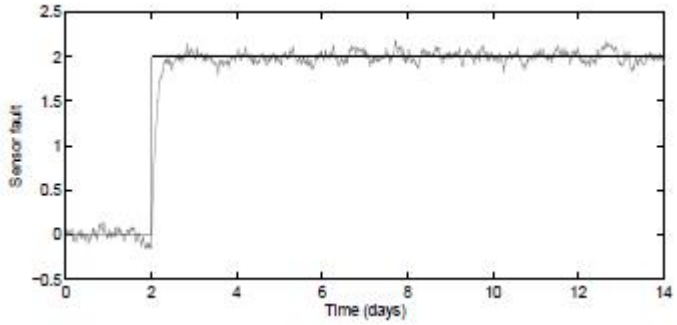


FIGURE 4.5 – *Défaut échelon*

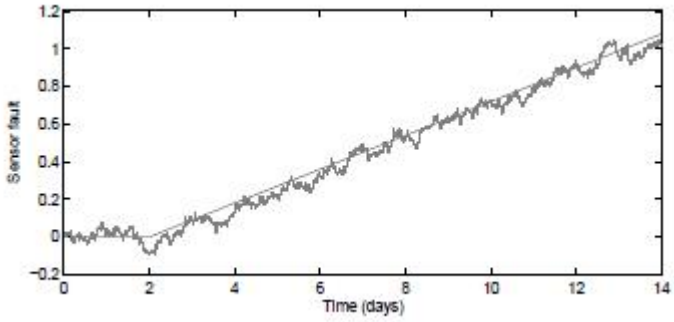


FIGURE 4.6 – *Défaut dérive*

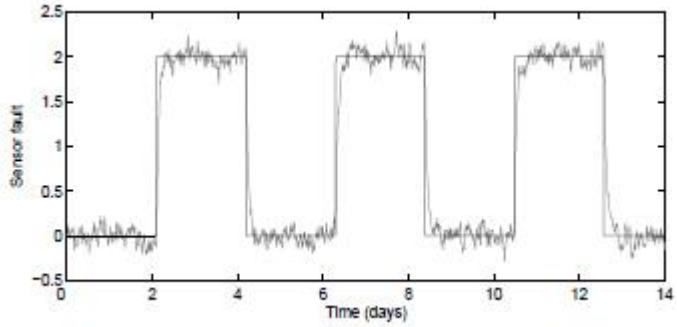


FIGURE 4.7 – *Défaut intermittent*

mesures.

Une approche pour identifier et reconstruire les défauts capteur pour une classe de systèmes non linéaires a été proposée. Elle est basée sur la théorie de l'observabilité des entrées inconnues. Le défaut capteur est considéré comme entrée inconnue.

Perspectives de recherche

Tout au long de ce mémoire, je me suis attaché à présenter les différentes activités auxquelles j'ai pu participé :

1. **Commande et modélisation à base de techniques de soft-computing**
2. **Etude R&D d'une collaboration industrielle**
3. **Recherche sur les observateurs grand gain et entrées inconnues**

Perspectives

A partir de l'observateur à entrées inconnues pour le diagnostic, nous allons développer un banc d'observateurs de façon à détecter aussi bien des défauts capteurs que des défauts actionneurs. Dans cette configuration, nous serons toujours limités à un défaut simultanément. Un article est en cours de rédaction.

Sur le plan théorique, nous allons travailler sur le cas 4 sorties-7 états de manière à essayer de détecter deux défauts simultanément.

De plus, nous allons utilisé la théorie de l'identification pour le diagnostic dans le cadre du pronostic.

Un projet industriel avec Opera Ergonomie s'intitulant Plateforme Autonome de Renseignement Embarquée (PARE) va démarrer début 2012. L'objectif du projet PARE est de concevoir et développer un système permettant de classifier les navires et les embarcations à partir d'un ensemble d'informations hétérogènes.

Journaux

[J1] J-F. Balmat, F. Lafont, R. Maifret, et N. Pessel. Maritime risk assessment (marisa), a fuzzy approach to define an individual ship risk factor. *Ocean Engineering.*, 36(15-16) :1279–1286, 2009.

[J2] J-F. Balmat, F. Lafont, R. Maifret, et N. Pessel. A decision-making system to maritime risk assessment. *Ocean Engineering.*, 38 :171–176, 2011.

[J3] F. Lafont et J-F. Balmat. Optimized fuzzy control of a greenhouse. *Fuzzy Sets and Systems.*, 128(1) :47–59, 2002.

[J4] F. Lafont et J-F. Balmat. Fuzzy logic to the identification and the command of the multidimensional systems. *International Journal of Computational Cognition.*, 2(3) :21–47, 2004.

[J5] F. Lafont, E. Busvelle, et J-P. Gauthier. An adaptive high-gain observer for wastewater treatment systems. *Journal of Process Control.*, 21(6) :893–900, 2011.

[J6] F. Lafont, N. Pessel, et J-F. Balmat. Evaluation of a new fuzzy approach in the recursive least squares algorithm for system identification. *Transactions on Systems, Signals & Devices - Issues on Systems, Analysis & Automatic Control.*, 3(1) :51–67, 2008a.

[J7] F. Lafont, N. Pessel, et J-F. Balmat. A multiple sensor fault detection method based on fuzzy parametric approach. *The Informatics in Control, Automation and Robotics - Lecture notes in electrical engineering.*, 24(1) :37–50, 2008b.

[J8] S. Methnani, J-P. Gauthier, et F. Lafont. Sensor fault reconstruction and observability for unknown inputs, with an application to wastewater treatments plants. *International Journal of Control.*, 84(4) :822–833, 2011.

[J9] N. Pessel, J. Duplaix, J-F. Balmat, et F. Lafont. A multi-structure modeling methodology. *Springer Verlag - Soft computing based modeling in intelligent systems - Studies in Computational Intelligence.*, 5 :93–114, 2009.

[J10] J-C. Ramos-Fernandez, V. Lopez-Morales, F. Lafont, G. Enea G., et J. Duplaix. A neurofuzzy structure modelling evapotranspiration in a greenhouse crop. *Ingeniería Investigación y Tecnología.*, XI(2) :127–139, 2010.

[J11] A. Trabelsi, F. Lafont, M. Kamoun, et G. Enea. Identification of nonlinear multivariable systems by adaptive fuzzy takagi-sugeno model. *International Journal of Computational Cognition.*, 2(3) :137–153, 2004.

[J12] A. Trabelsi, F. Lafont, M. Kamoun, et G. Enea. Fuzzy identification of a greenhouse. *Applied Soft Computing.*, 7(3) :1092–1101, 2007.

Conférences internationales

[C1] J-F. Balmat et F. Lafont. Multi-model architecture supervised by kohonen map. *Sciences of Electronic, Technology of Information and Telecommunication (SETIT 2003) - Tunisie.*, 2003.

[C2] J-F. Balmat, F. Lafont, et N. Pessel. Neural networks for system design to maritime risk assessment. *IEEE - 4th International Conference on Intelligent Information Technology Application (IITA 2010) - Chine.*, 2010.

[C3] J. Duplaix, N. Pessel, F. Lafont, et J-F. Balmat. Identification du microclimat d'une culture hors-sol sous serre : système hiérarchisé fortement perturbé. *CIFA'10 Nancy, Vol. Session invitée : Modélisation expérimentale pour les systèmes environnementaux - Nancy.*, 2010.

[C4] F. Lafont, J-F. Balmat, N. Pessel, et J. Duplaix. A software sensor for the wastewater treatment. *9th International Conference on Sciences and Techniques of Automatic control & computer engineering (STA 2008)) - Tunisie.*, 2008.

[C5] F. Lafont, J-F. Balmat, et M. Taurines. Fuzzy forgetting factor for system identification. *3th IEEE International Conference on Systems, Signals & Devices (SSD 2005) - Tunisie.*, 2005.

[C6] F. Lafont, N. Pessel, et J-F. Balmat. A fuzzy parametric approach for the model-based diagnosis. *4th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2007) - Angers.*, 2007.

[C7] S. Methnani, T. Damak, A. Toumi, F. Lafont, et J-P. Gauthier. Adaptive-high-gain observers with an application to wastewater treatment plants. *International Conference on Communications, Computing and Control Applications (CCCA 2011) - Tunisie.*, 2011.

[C8] N. Pessel, J-F. Balmat, F. Lafont, et J. Bonnal. An improved pca fault detection for the diagnosis. *9th WSEAS International Conference on Automatic Control, Modeling & Simulation (ACMOS 2007) - Turquie.*, 2007a.

[C9] N. Pessel, J-F. Balmat, F. Lafont, et J. Bonnal. Neuronal principal component analysis for the diagnosis of a non linear system. *15th Mediterranean Conference on Control and Automation (MED 2007) - Grèce.*, 2007b.

[C10] N. Pessel, J. Duplaix, J-F. Balmat, et F. Lafont. Data analysis for neuro-fuzzy model approach. *IEEE International Workshop on Soft Computing Applications (SOFA 2005) - Roumanie.*, 2005.

[C11] A. Trabelsi, F. Lafont, M. Kamoun, et G. Enea. Neural identification of a greenhouse. *International Conference on Sciences and Techniques of Automatic (STA 2004) - Tunisie.*, 2004.

Conférences nationales

[N1] J-F. Balmat et F. Lafont. Une approche neuronale multimodèle pour l'identification d'une serre agricole. *11ième Journées Neurosciences et Science pour l'Ingénieur (NSI) - La londe.*, 2002.

[N2] M. Bouchouicha, J. Duplaix, F. Bouchara, F. Lafont, J-F. Balmat, et G. Enea. Automatisation d'une serre : Plate-forme expérimentale multi-applications et multi-utilisateurs. *Sciences and Techniques of Automatic (STA 2002) - Tunisie.*, 2002a.

[N3] M. Bouchouicha, F. Lafont, et J-F. Balmat. Neural networks, fuzzy logic and genetic algorithms for greenhouse identification. *Conference on Electro-Technical and Automatic Control (JTEA 2002) - Tunisie.*, 2002b.

[N4] F. Lafont et J-F. Balmat. Modélisation floue itérative d'une serre agricole. *Rencontres francophones sur la logique floue et ses applications - Belgique.*, 2001.

[N5] F. Lafont, J-F. Balmat, J. Duplaix, N. Pessel, et J-P. Gauthier. Observateur grand gain adaptatif pour les systèmes de traitement des eaux usées. *Journées Identification et Modélisation Expérimentale - Douais.*, 2011.

[N6] F. Lafont, J-F. Balmat, N. Pessel, et R. Maifret. Une approche floue pour l'evaluation des risques maritimes. *STIC & Environnement 2009 - Calais.*, 2009.

[N7] A. Trabelsi, M. Chaabane, et F. Lafont. Commande neurronale par modèle inverse des systèmes non linéaires. *Journées en Génie Electrique et Informatique - Tunisie.*, 2003.

[N8] A. Trabelsi, M. Chaabane, et F. Lafont. Modélisation et commande avancée d'une serre agricole. *Journées Scientifiques et Technologiques sur la Conduite Automatique de Systèmes Agricoles - Tunisie.*, 2004.

[N9] A. Trabelsi, F. Lafont, M. Kamoun, et G. Enea. Modélisation d'une serre agricole par la méthode des clusters flous. *Journées en Génie Electrique et Informatique - Tunisie.*, 2005.

Rapports de recherche

[R1] J-F. Balmat, F. Lafont, et N. Pessel. Etude sur les réseaux connexionnistes et la logique floue appliqués à l'evaluation des risques et menaces pour la sauvegarde maritime. *(LSIS.RR.2007.006).*, 2007.

[R2] J-F. Balmat, F. Lafont, et N. Pessel. Etude sur les réseaux connexionnistes et la logique floue appliqués à l'evaluation des risques et menaces pour la sauvegarde maritime. *(LSIS.RR.2008.006).*, 2008.

Bibliographie annexe

5

5.1 Commande floue

ELSEVIER

Fuzzy Sets and Systems 128 (2002) 47–59

FUZZY sets and systems

www.elsevier.com/locate/fss

Optimized fuzzy control of a greenhouse

F. Lafont *, J.-F. Balmat

MS/SSD, Université de Toulon et du Var, B.P. 132-83957 La Garde Cedex, France

Abstract

Computer systems can be used to control the greenhouse climate in order to improve the culture development and to minimize the production costs.

We have a system which allows to acquire the measurements of internal and external temperature and hygrometry, global radiation, and wind velocity. It consists of a heating system, moistening ducts and a static ventilation to control the internal climate.

Since 1991, a classical controller on–off has been implemented in our experimental greenhouse, which enabled us to have a great number of data files. Knowing that the conventional techniques of regulation are difficult to implement in this type of system (multivariable, nonlinear, nonstationary) where the interdependence of temperature and hygrometry with the meteorological disturbances are strong, we were brought to study the fuzzy controllers.

This paper shows that it is possible to successfully control a greenhouse by using these techniques. During the probation period, we compare the various results obtained with these controllers.

Keywords: Fuzzy system models; Fuzzy control; Process control

1. Introduction

The control of greenhouse climate, in order to improve the development of a specific cultivation and to minimize the production costs, is becoming increasingly important for the growers [5]. For a few years, our laboratory has been interested in this problem, and the conventional techniques of regulation (multivariable adaptive control, optimal control, logical control) have been developed. Interactions between the internal and external variables, and the complexity of the phenomena (multivariable, nonlinear, nonstationary) are such that it is often difficult to implement the conventional techniques of regulation. Moreover, these methods induce choices to simplify assumptions, and they are often very sensitive to the disturbances which are not envisaged in the model.

To solve these problems, we propose a different approach by using techniques based on expert knowledge and by taking the study from the system as a unit. Thus, we naturally took an interest in fuzzy control. Indeed, the criteria which allows to use the fuzzy controllers, for the regulation of the systems, are associated with the complexity of the process model [4]. The essential advantage of these techniques is that it is not necessary to define a process model, which largely facilitates the implementation of this type of regulator.

* Corresponding author. Tel.: +33-4-94-14-20-78; fax: +33-4-94-14-21-68.

E-mail address: lafont@univ-tln.fr (F. Lafont).

0165-0114/02/$ - see front matter
PII: S0165-0114(01)00182-8

In this paper, we first describe the experimental greenhouse system (structure, sensors, etc.), then we propose our method of comparison of the regulators (basic and optimized fuzzy) which we studied in simulation. Finally, we give the results and the prospects considered.

2. Regulation of greenhouse

A good microclimate (temperature, hygrometry) is closely related to the external weather conditions and the greenhouse itself (structure, plants, etc.).

At the present time, few greenhouses are equipped with a sophisticated device of regulation. Many greenhouses are still controled manually and require the intervention of the grower.

However, there are installations with thermostatic systems which are characterized by independent regulations for each function (heating, moistening, etc.) and on–off commands (logical with low and high thresholds). Such a system of regulation is often insufficient because it does not take into account the relationships between the variables.

Some studies were developed using fuzzy logic [6] by considering the temperature and the hygrometry as inputs and the heating as output. In our study, we take into account more inputs (temperature, hygrometry, external disturbances) and more outputs (heating, moistening, roofing).

3. The experimental greenhouse

As shown in Fig. 1, the principal physical characteristics of our experimental greenhouse are:

- Greenhouse out of glass and with metal reinforcements (classical structure of greenhouse).
- Volume: 120 m^3.
- Surface: 40 m^2.
- Sensors with semiconductor AD 590 (temperature measurement of the air, internal and external).
- Sensors with capacitive effect HR (measurement of the relative hygrometry, internal and external).
- Electromagnetic anemometer (velocity measurement of the wind).

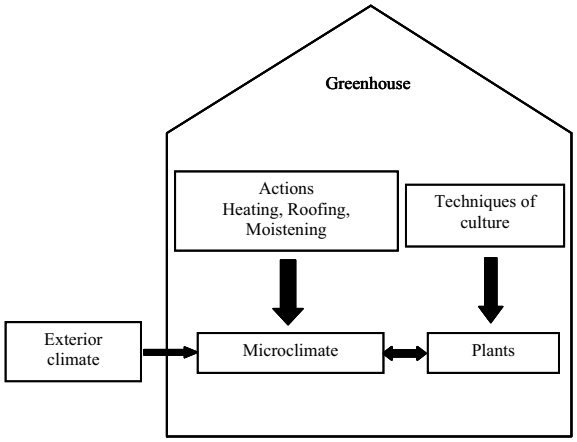


Fig. 1. Principle diagram of greenhouse.

- Solarimeter with thermopile (measurement of the global radiation).
- Three heating units of a total power of 5 kW.
- A static ventilation ranging from 0 to 32.5 angular degrees.
- Moistening ducts.

4. Simulation

In our work, the fuzzy controller was not implemented on site for several reasons. The ideal experimentation should allow to simultaneously compare two distinct controllers at the same time, in the same greenhouse. It would be necessary to share the greenhouse in two isolated parts with the same setup of sensors in each one. Our greenhouse does not allow us to perform this experiment. Moreover, the simulation allows to test rapidly the different controllers in all seasons. Thus, we have compared the different controllers which were developed in simulation, by using a known model.

4.1. Simulation model

4.1.1. Physical model

The physical model is based on the writing of equations which allow to describe the state of the system. We are interested in a dynamic model in order to elaborate a simulation tool to develop and to test the fuzzy controller. A lot of models have been studied [3,14,15]

and we have chosen the model developed by Viard Gaudin [16]. The energy balances constitute the rules of the evolution of model [12]. Thus, we obtain the general equations of the internal temperature and of the internal absolute hygrometry.

The differential equations (nonlinear) are the following:

$$\frac{\mathrm{d}T_{\mathrm{ai}=}}{\mathrm{d}t} = (\alpha_1 + \alpha_2 O_{\mathrm{v}})(T_{\mathrm{ae}} - T_{\mathrm{ai}}) + \alpha_3 \mathrm{Ch} + \alpha_4 \mathrm{Rg} - \alpha_5$$

$$\frac{\mathrm{d}X_{\mathrm{ai}}}{\mathrm{d}t} = (\beta_1 + \beta_2 O_{\mathrm{v}})(X_{\mathrm{ae}} - X_{\mathrm{ai}}) + (\beta_3 + \beta_4 \mathrm{Rg})\Delta X_{\mathrm{ai}} - \beta_5$$

$$X_{\mathrm{ai}} = H_{\mathrm{ai}} X_{\mathrm{sat}_{\mathrm{ai}}}$$

$$X_{\mathrm{ae}} = H_{\mathrm{ae}} X_{\mathrm{sat}_{\mathrm{ae}}}$$

with input variables include T_{ae} the external temperature (°C), O_{v} the roofing command (angular degree), Ch the heating command (kW), Rg the global radiation (kW/m^2), X_{ae} the external absolute hygrometry (g/kg), α_5 and β_5 the disturbances not measured. Output variables include T_{ai} the internal temperature (°C), X_{ai} the internal absolute hygrometry (g/kg). Intermediate variables include ΔX_{ai} the deficit of water (g/kg), X_{sat} the hygrometry at the saturation (g/kg). Model parameters include α_i and β_i which are general parameters of the model.

First, we have used this physical model to simulate the behavior of the greenhouse and we have used an identification model of greenhouse based on the square means. We have executed the simulation in using the meteorological data file during a period of 6 h (from 7 a.m to 1 p.m) and we have compared the classical command with the fuzzy commands [10]. However, this model is only valid for 6 h. For this reason, we have developed a fuzzy model which takes into account more variables and which is able to simulate a whole day.

4.1.2. Fuzzy modeling

We have used an iterative fuzzy modeling method using multi-dimensional fuzzy sets. The objective was to obtain fuzzy logic rules of Takagi–Sugeno type [1,2]. It was to automatically generate a fuzzy partition of the process-input space and of its associated Takagi–Sugeno set of rules from measured data, and, next, to make them grow so as to achieve an optimal partition and rule base with respect to the given performance criteria.

The chosen structure for this model is as follows: Regarding the internal temperature (T_{ai}):

$$T_{\mathrm{ai}\,k+1} = a_1 O_{\mathrm{v}\,k} + a_2 \mathrm{Ch}_k + a_3 T_{\mathrm{ae}\,k} + a_4 H_{\mathrm{ae}\,k} + a_5 \mathrm{Rg}_k + a_6 V_{\mathrm{v}\,k} + a_7 T_{\mathrm{ai}\,k} + a_8 H_{\mathrm{ai}\,k} + a_9 \mathrm{Br}_k.$$

Regarding the internal hygrometry (H_{ai}):

$$H_{\mathrm{ai}\,k+1} = b_1 O_{\mathrm{v}\,k} + b_2 \mathrm{Ch}_k + b_3 T_{\mathrm{ae}\,k} + b_4 H_{\mathrm{ae}\,k} + b_5 \mathrm{Rg}_k + b_6 V_{\mathrm{v}\,k} + b_7 T_{\mathrm{ai}\,k} + b_8 H_{\mathrm{ai}\,k} + b_9 \mathrm{Br}_k$$

with T_{ae} the external temperature, O_{v} the roofing command, V_{v} the wind velocity, Ch the heating command, Rg the global radiation, Br the moistening command, H_{ae} the external hygrometry, $a_1, a_2, \ldots, a_9$ are parameters of the model for the temperature and $b_1, b_2, \ldots, b_9$ are parameters of the model for the hygrometry. A good precision for the model is achieved thanks to 10 multi-dimensional fuzzy sub-sets (25 rules).

4.1.3. Model validation

On a greenhouse, we cannot measure the cross-validation. In order to validate the model in simulation, for a day, we have used the program of the classical command (on–off), implemented on our greenhouse, while taking the disturbances into account. (Figs. 2 and 3)

4.2. The fuzzy controller

A fuzzy controller (Fig. 4) is a system which works on numerical data and converts it into a symbolic form through a data base (fuzzification). A logic of decision-making (rule base) is implemented, thus it is possible to provide a symbolic answer which must be converted into a numerical data (defuzzification) [7].

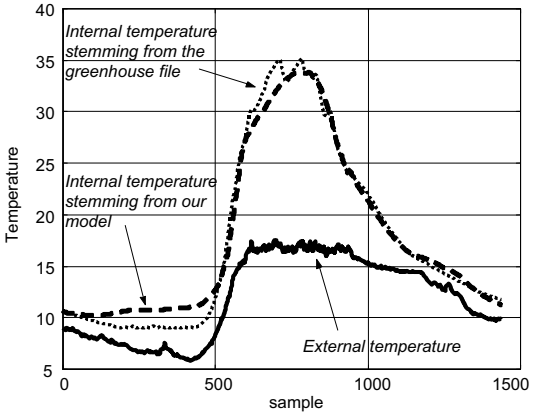


Fig. 2. Temperature curve.

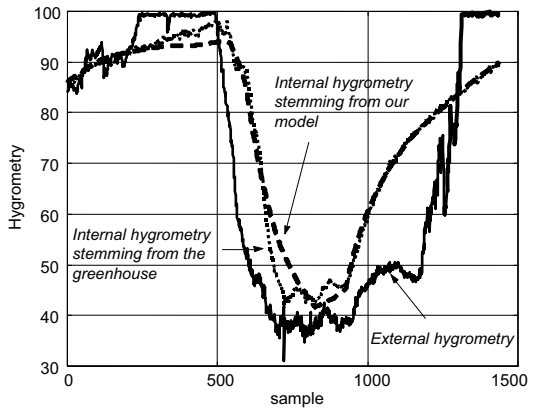


Fig. 3. Hygrometry curve.

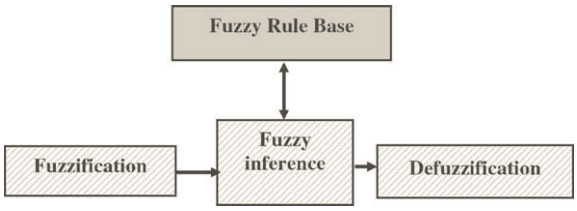


Fig. 4. Fuzzy controller.

The functional diagram of the system with the fuzzy controller is represented in Fig. 5.

Implementation of fuzzy control for the greenhouse. The realization of a fuzzy controller consists of the following stages:

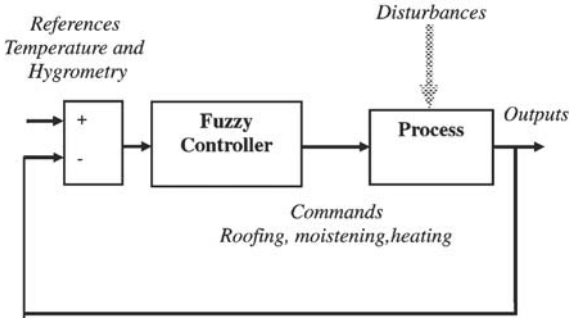


Fig. 5. Functional diagram of the system.

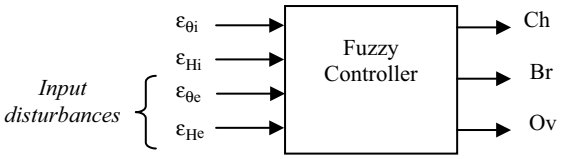


Fig. 6. Basic fuzzy controller with input and output variables.

- choice of the input and command variables,
- fuzzification: definition of the membership degrees for each variable,
- rule base creation,
- choice of the fuzzy inference type, which allows to activate the rules according to the input linguistic variables,
- defuzzification.

We have:

- outputs associated with the temperature (θ_i) and the hygrometry (H_i) inside the greenhouse (sensors measurements),
- references in temperature (θ_c) and hygrometry (H_c),
- disturbances due to the external temperature (θ_e) and the external hygrometry (H_e).

4.2.1. Basic fuzzy controller [10]

4.2.1.1. Choice of the input and command variables. One of the difficulties, for the implementation of a fuzzy system, is the choice and the number of input variables. In our study, we have defined a fuzzy controller with four inputs and three outputs. We have used the temperature and hygrometry variations (internal and external) compared to their references.

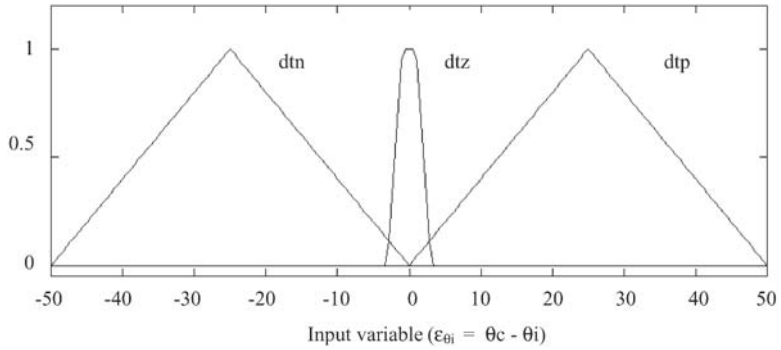


Fig. 7. The input ($\varepsilon\theta_{\mathbf{i}} = \theta_{\mathrm{c}} - \theta_{\mathrm{i}}$).

Thus, we have defined (Fig. 6):

- variations in temperature ($\varepsilon_{\theta_{\mathrm{i}}} = \theta_{\mathrm{c}} - \theta_{\mathrm{i}}$ and $\varepsilon_{\theta_{\mathrm{e}}} = \theta_{\mathrm{c}} - \theta_{\mathrm{e}}$) and in hygrometry ($\varepsilon_{H_{\mathrm{i}}} = H_{\mathrm{c}} - H_{\mathrm{i}}$ and $\varepsilon_{H_{\mathrm{e}}} = H_{\mathrm{c}} - H_{\mathrm{e}}$),
- commands of the heating (Ch), the moistening (Br) and the roofing (O_{v}).

These controller outputs vary between 0 and 1 for the heating and the moistening, and between -1 and 1 for the roofing. A threshold is defined for each command, so we have 0 or 1 for the heating and for the moistening (on–off) and -1, 0 or 1 for the roofing (open, no action or closed).

4.2.1.2. Fuzzification. Each input is defined with several membership functions. The domain of application varies according to the input. We give an example in Fig. 7.

$\varepsilon_{\theta_{\mathrm{i}}}$ belongs to the class dtn if $\theta_{\mathrm{c}} < \theta_{\mathrm{i}}$, $\varepsilon_{\theta_{\mathrm{i}}}$ belongs to the class dtp if $\theta_{\mathrm{c}} > \theta_{\mathrm{i}}$ and $\varepsilon_{\theta_{\mathrm{i}}}$ belongs to the class dtz if $\theta_{\mathrm{c}} \approx \theta_{\mathrm{i}}$.

The commands of the heating and the moistening have two classes (heating or not, moistening or not). On the other hand, the roofing output has three classes (closed, open, no action).

The choice of the membership functions allows to introduce a hysteresis around the input. The tolerance is $\pm 2^{\circ}$ for $\varepsilon_{\theta_{\mathrm{i}}}$ and $\pm 10\%$ for $\varepsilon_{H_{\mathrm{i}}}$.

4.2.1.3. Fuzzy rule base. By taking into account the number of inputs, the number of membership functions and the constraints associated with the greenhouse (for example, a moistening every 10 min, at the most, not to damage the cultures), the fuzzy base contains 81 rules. They were defined starting from an evaluation of the files of the greenhouse (many files since 1991) and from the expert knowledge of the grower.

Example. If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtz) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhz) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detp) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehn) then (Ch is PC)(Br is PB)(O_{v} is Z).

If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtz) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhn) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detn) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehn) then (Ch is PC)(Br is PB)(O_{v} is F).

If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtz) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhn) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detz) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehp) then (Ch is PC)(Br is PB)(O_{v} is O).

If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtz) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhp) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detn) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehn) then (Ch is PC)(Br is B)(O_{v} is F).

If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtn) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhz) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detn) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehn) then (Ch is PC)(Br is PB)(O_{v} is O).

If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtp) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhn) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detn) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehp) then (Ch is C)(Br is PB)(O_{v} is F).

If ($\varepsilon_{\theta_{\mathrm{i}}}$ is dtp) and ($\varepsilon_{H_{\mathrm{i}}}$ is dhp) and ($\varepsilon_{\theta_{\mathrm{e}}}$ is detp) and ($\varepsilon_{H_{\mathrm{e}}}$ is dehz) then (Ch is C)(Br is B)(O_{v} is F).

PC: no heating, C: heating; PB: no moistening, B: moistening; Z: no action, O: open, F: closed.

4.2.1.4. Choice of the fuzzy inference and the method of deffuzification. The selected method of inference [7] for our application is the method of Mamdani which is known as the max–min method:

- the min for the implication: for each rule, the system takes as a conclusion the smallest value of premises,
- the max for the aggregation: the maximum of all the minimum is taken for the same output characteristic.

The method of deffuzification is the classical method of center-of-gravity (COG).

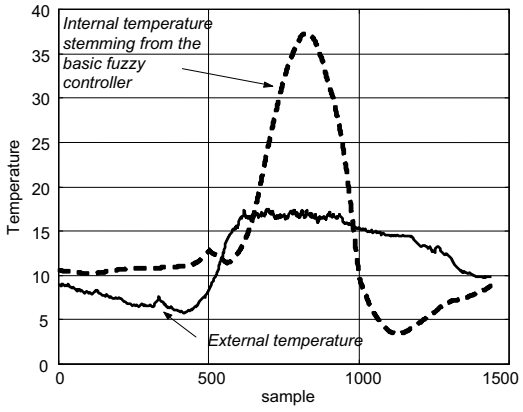


Fig. 8. Temperature curve for the basic fuzzy controller.

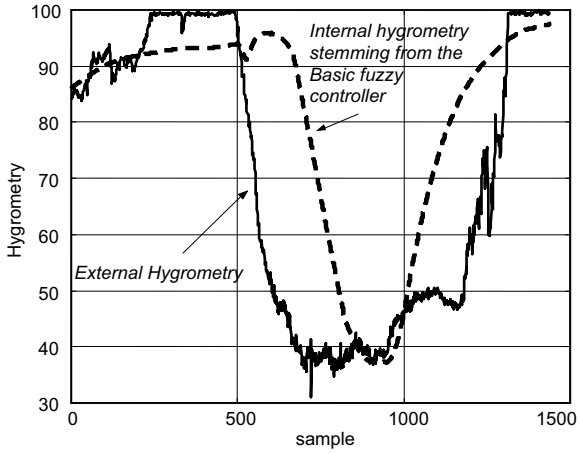


Fig. 9. Hygrometry curve for the basic fuzzy controller.

4.2.1.5. Results. The temperature reference is fixed to 11°, it is not possible to follow the reference when the external temperature is very important. (Fig. 8)

The hygrometry reference is fixed to 70%. (Fig. 9)

At time 1020 (min), the heating command is activated because the internal temperature has reached 9° effective. (Fig. 10)

The moistening is limited to one action every 10 min. (Fig. 11)

At time 500, the roofing command is activated because the internal temperature has reached 13° effective. (Fig. 12)

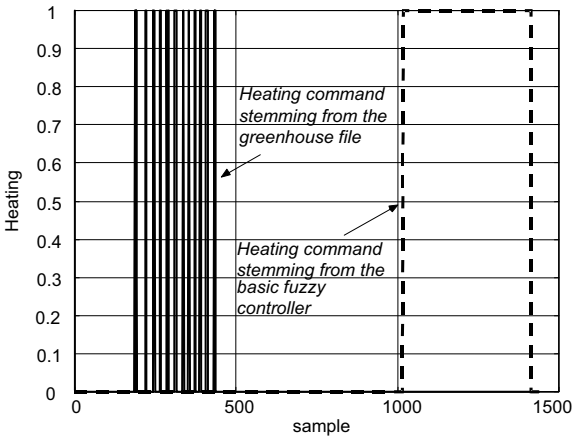


Fig. 10. Heating command for the basic fuzzy controller.

4.2.2. Optimized fuzzy controller

With the structure of basic fuzzy controller, we can see a problem for the end of the day. Indeed, we can verify, in Figs. 8 and 9, that the controller has a correct working until midday. But after, the commands are not good because we do not take into account the variation senses. Thus we note, for example, in the case of temperature, that the rules have only been defined for an increasing variation. In fact, the fuzzy controller does not have the derivative of each input. In our approach, the fuzzy controller is composed of 81 rules. If we increase the number of inputs by adding their derivatives, the fuzzy controller becomes unmanageable. So, we have imagined a fuzzy controller with a distinct structure [9,13]. We want to tackle multivariable problems, by taking profits of numerous studies on SISO fuzzy controllers. Over these last years, several methods, to reduce rules, have been proposed such as the sensory fusion approach or the change of the fuzzy controller structures (the decentralized or the hierarchical structure). In our work, we have decided to develop a decentralized control structure including two robust fuzzy controllers. For this, we have analyzed the couplings in our multivariable system. We have begun our study by a qualitative modeling of the process behavior. We have looked for the interdependence between inputs and outputs. We have found that the heating and the roofing act on the temperature and that the moistening acts on the hygrometry (Fig. 13).

So, the first fuzzy controller is made for the heating and the roofing with the temperature input; the second

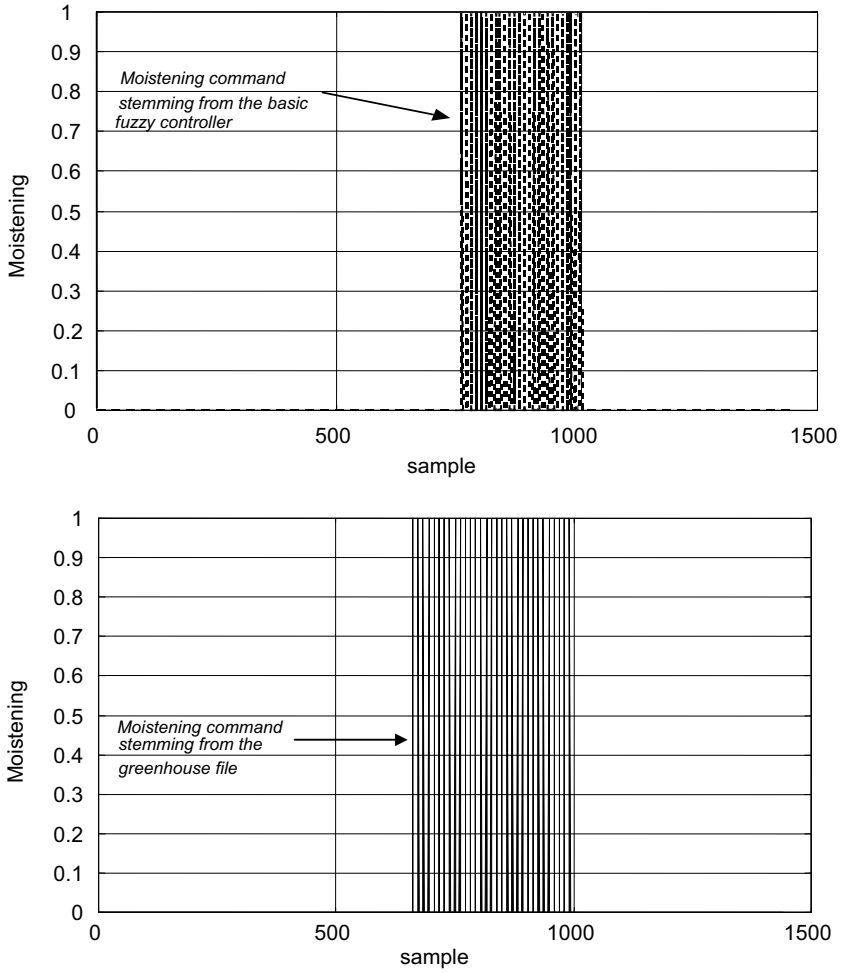


Fig. 11. Moistening command for the basic fuzzy controller.

for the moistening with the hygrometry input. Thus, we can introduce the derivative of each input (Fig. 14).

4.2.2.1. Choice of the input and command variables. We have defined a fuzzy controller (named fuzzy controller 1) with two inputs and two outputs and a fuzzy controller (named fuzzy controller 2) with two inputs and one output. We have used temperature and hygrometry variations (internal and external) compared to their references.

Thus, we have defined:

- variations in temperature ($\varepsilon_{\theta_i} = \theta_c - \theta_i$ and $\Delta\varepsilon_{\theta_i}$) and in hygrometry ($\varepsilon_{H_i} = H_c - H_i$ and $\Delta\varepsilon_{H_i}$),
- commands of the heating (Ch), the moistening (Br) and the roofing (O_v).

4.2.2.2. Fuzzification and choice of the fuzzy inference and the method of deffuzification. The fuzzification, the fuzzy inference method and the method of deffuzification are the same as the basic fuzzy controller.

4.2.2.3. Fuzzy rule base. With the same constraints as the basic fuzzy controller, the fuzzy base of controllers contains nine rules.

Fuzzy rule base of controller 1

If (ε_{θ_i} is dtp) and ($\Delta\varepsilon_{\theta_i}$ is ddtp) then (Ch is C)(O_v is F).
If (ε_{θ_i} is dtp) and ($\Delta\varepsilon_{\theta_i}$ is ddtz) then (Ch is C)(O_v is F).

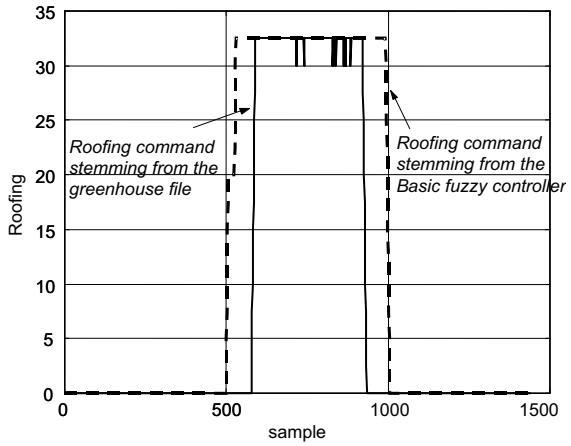


Fig. 12. Roofing command for the basic fuzzy controller.

If (ε_{θ_i} is dtp) and ($\Delta\varepsilon_{\theta_i}$ is ddtn) then (Ch is C)(O_v is F).
If (ε_{θ_i} is dtz) and ($\Delta\varepsilon_{\theta_i}$ is ddtp) then (Ch is PC)(O_v is Z).
If (ε_{θ_i} is dtz) and ($\Delta\varepsilon_{\theta_i}$ is ddtz) then (Ch is PC)(O_v is Z).
If (ε_{θ_i} is dtz) and ($\Delta\varepsilon_{\theta_i}$ is ddtn) then (Ch is PC)(O_v is F).
If (ε_{θ_i} is dtn) and ($\Delta\varepsilon_{\theta_i}$ is ddtp) then (Ch is PC)(O_v is O).
If (ε_{θ_i} is dtn) and ($\Delta\varepsilon_{\theta_i}$ is ddtz) then (Ch is PC)(O_v is Z).
If (ε_{θ_i} is dtn) and ($\Delta\varepsilon_{\theta_i}$ is ddtn) then (Ch is PC)(O_v is Z).

Fuzzy rule base of controller 2

If (ε_{H_i} is dhp) and ($\Delta\varepsilon_{H_i}$ is ddhp) then (Br is B).
If (ε_{H_i} is dhp) and ($\Delta\varepsilon_{H_i}$ is ddhz) then (Br is B).
If (ε_{H_i} is dhp) and ($\Delta\varepsilon_{H_i}$ is ddhn) then (Br is PB).
If (ε_{H_i} is dhz) and ($\Delta\varepsilon_{H_i}$ is ddhp) then (Br is PB).
If (ε_{H_i} is dhz) and ($\Delta\varepsilon_{H_i}$ is ddhz) then (Br is PB).

Inputs			Outputs	
Ch	Ov	Br	Ti	Hi
+	0	0	++	-
++	0	+	+++	0
0	+	+	-	+
0	0	+	0	++
0	++	0	---	-
0	0	++	0	+++

Keys of the symbols
+ increase of the variable
- decrease of the variable
0 variation of the variable null or quite null

The number of signs represents the variation magnitude of the variable

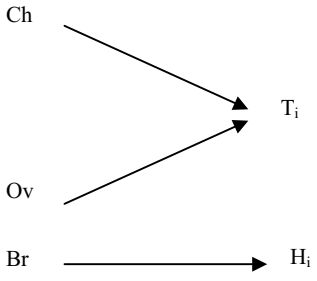


Fig. 13. Qualitative analysis results and simple causal graph.

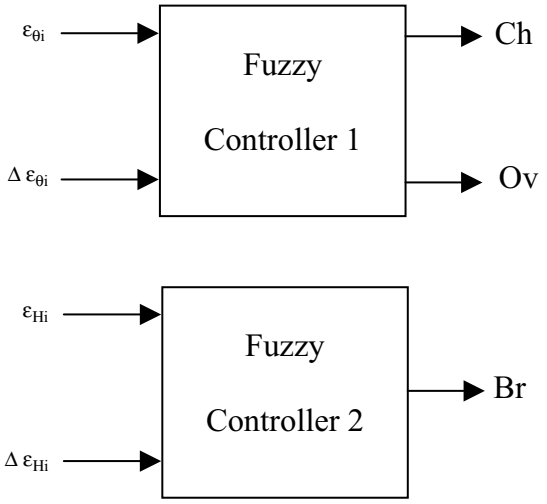


Fig. 14. Optimized fuzzy controller with input and output variables.

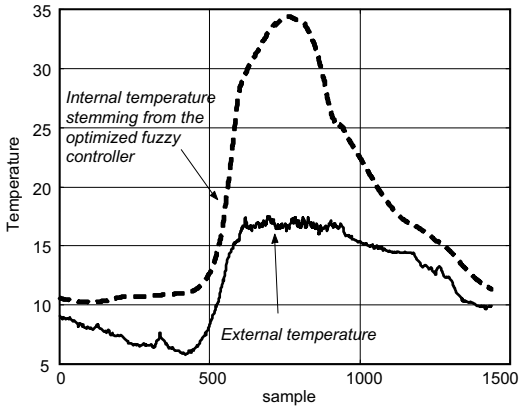


Fig. 15. Temperature curve for the optimized fuzzy controller.

If (ε_{H_i} is dhz) and ($\Delta\varepsilon_{H_i}$ is ddhn) then (Br is B).
If (ε_{H_i} is dhn) and ($\Delta\varepsilon_{H_i}$ is ddhp) then (Br is PB).
If (ε_{H_i} is dhn) and ($\Delta\varepsilon_{H_i}$ is ddhp) then (Br is PB).
If (ε_{H_i} is dhn) and ($\Delta\varepsilon_{H_i}$ is ddhz) then (Br is PB).
If (ε_{H_i} is dhn) and ($\Delta\varepsilon_{H_i}$ is ddhn) then (Br is PB).

4.2.2.4. Study of stability. We must verify that these fuzzy controllers are stable. We have used a method based on the extension of the classical Lyapunov synthesis method by constructing a Lyapunov function candidate V and then by determining the conditions required to make it a Lyapunov function of the closed-loop system [8,11]. We have chosen as function $V(x_1,x_2)=1/2(x_1^2+x_2^2)$.

We have $V(0,0)=0$, $V(x_1,x_2)>0$ and if $\dot{V}(x_1,x_2)\leqslant 0$ (with $x_2=\dot{x}_1$) then the controller is stable.

Example for the roofing. We assume that the reference θ_c and its derivatives are bounded and available to the controller. We choose $V=1/2(e^2+\dot{e}^2)$ where $e=\theta_c-x_1$, $x_1=\theta_i$, $x_2=\dot{x}_1$ then $\dot{V}=e\dot{e}+\dot{e}\ddot{e}=e\dot{e}+\dot{e}(\ddot{\theta}_c-\ddot{x}_1)=e\dot{e}+\dot{e}(\ddot{\theta}_c-\dot{x}_2)$ and denoting $w=\ddot{\theta}_c-\dot{x}_2$, we obtain $\dot{V}=e\dot{e}+\dot{e}w$. Hence, we require that $\dot{V}=e\dot{e}+\dot{e}w\leqslant 0$. So if e and $\dot{e}$ have opposite signs, then it is necessary that $w=0$. If e and $\dot{e}$ are both positive, then $w<-e$.
If e and $\dot{e}$ are negative, then $w>-e$.
If $e=0$ and $\dot{e}$ is positive, then $w>0$.
If $e=0$ and $\dot{e}$ is negative, then $w<0$.
$\forall e$ and $\dot{e}=0$, then $\forall w$ we have $\dot{V}=e\dot{e}+\dot{e}w\leqslant 0$.
We find the same rules which we have defined so, the fuzzy controller is stable.

4.2.2.5. Results. On the temperature curve (Fig. 15), of the samples 0 at 450, we note that the internal temperature is close to the reference. At time 450 min, the internal temperature deviates quickly from the reference until it reaches the maximum value (34°C) at midday (750 min). The evolution of internal temperature is the same at the evolution of external temperature with an amplification (no means of powerful cooling). At the end of the day, the gap between the internal temperature and the reference is reduced.

We can make the same remarks with the evolution of internal hygrometry (Fig. 16). For this one, the minimum value is 40% and the maximum is 92%.

Concerning the commands:

- The heating command is not activated because the internal temperature is greater than the reference (11°). (Fig. 17) (Table 1)
- At time 650 min, the moistening is activated because the internal hygrometry is inferior to 60% (reference (70%) — tolerance (10%)) (see Section 4.2.1.2.) until the time 950. (Fig. 18)
- The roofing is activated when the internal temperature is high (in comparison with reference) and

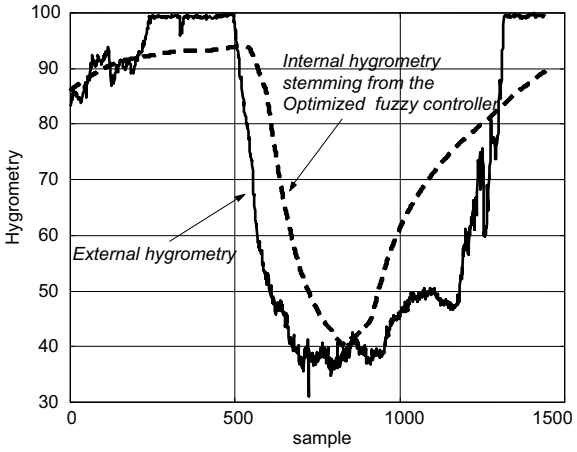


Fig. 16. Hygrometry curve for the optimized fuzzy controller.

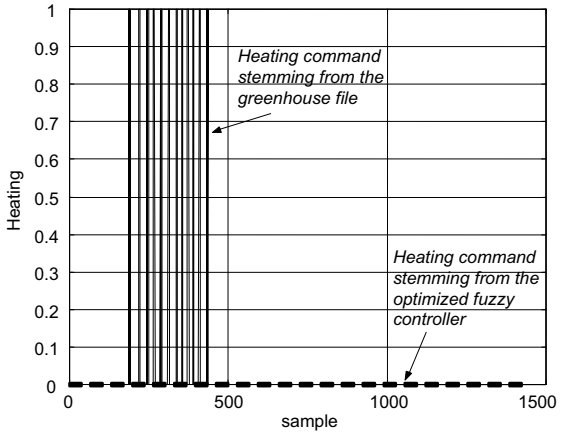


Fig. 17. Heating command for the optimized fuzzy controller.

it remains open (maximum opening: 32°5) for 5 h (samples 600 at 900). (Fig. 19)

Note: When the moistening is activated, the roofing is open. The goal is to reduce the internal temperature.

5. Comparisons

The basic controller allows to take into account external disturbances (Te, He) but its actual use poses a problem because the derivative variations are not available (Figs. 20 and 21). Indeed, on account of the inputs number, it is not reasonable to add these inputs. This controller is not perfect for a whole day (it is valid only for half a day: either in the morning, which represents a growth of the temperature and a fall of the hygrometry or in the evening, which represents a fall of the temperature and a growth of the hygrometry). On the other hand, the optimized fuzzy controller has a structure which is easier to implement (two inputs, nine rules for each fuzzy controller). It gives good results thanks to the introduction of variation senses of inputs, yet, it does not take into account external disturbances. The moistening commands are identical for the two controllers (Figs. 11 and 18). We see that the two roofing commands have the same form (Figs. 12 and 19). So, for these two commands, in this example, the energy consumption during the day of the simulation is identical.

Note: We have tested these controllers on other days, the results are similar to those presented in this paper.

We are going to work on an optimized fuzzy controller with all inputs (Ti, Hi, Te, He and their derivatives); this allows us to have the advantage of two fuzzy controllers presented in this paper.

6. Conclusions

This paper shows that a fuzzy controller can be applied successfully to control the greenhouse climate. Thus, we have developed two types of multivariable fuzzy controller (basic fuzzy controller and optimized

Table 1
Comparison of the controllers (N = number of samples)

	Classical controller	Basic fuzzy controller	Optimized fuzzy controller
Mean value of deviation in temperature $\sum \lvert\theta_i - \theta_c\rvert/N$	7.2	6.2	7.6
Mean value of deviation in hygrometry $\sum \lvert H_i - H_c\rvert/N$	17.2	20.9	17.2
Heating (%)	3.4	2.8	0
Moistening (%)	2.2	1.8	1.9
Roofing (% of opening on one day)	20	34	20

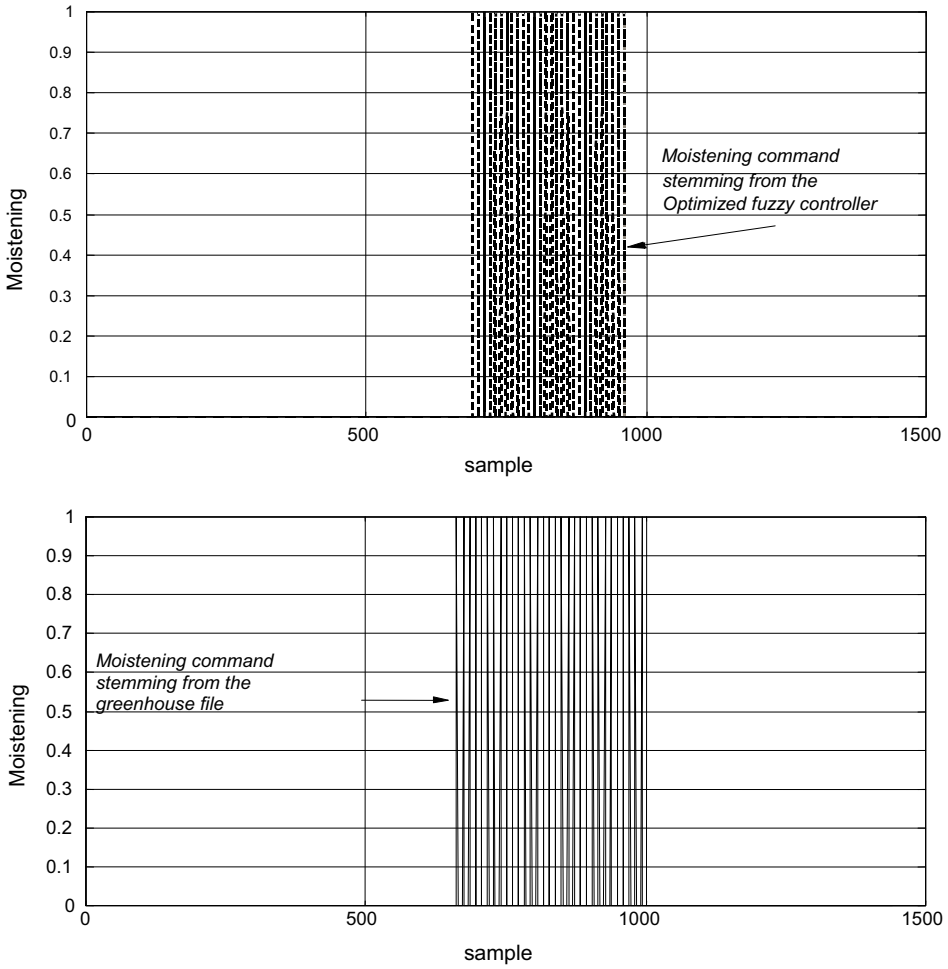


Fig. 18. Moistening command for the optimized fuzzy controller.

fuzzy controller) with a significant number of inputs and outputs. In this study, we have shown advantages and disadvantages of these controllers. In the first approach, we have developed and simulated a basic fuzzy controller, and its advantage is that we have taken disturbances as inputs. This solution, however, can only be valid for half a day. It would therefore be necessary to enter, as input variables, the variation senses of the internal temperature and of the internal hygrometry, these data being essential to obtain a good performance of the system during the entire day. In this case, the number of inputs becomes too great (which implies a difficult evaluation with an increase in the number of rules) and the development of a regulator of such a type turns out to be a real problem. In the second approach (the optimized fuzzy logic controller), the results obtained (Table 1) are practically the same ones as those of the classical controller with no heating command.

Advantages of the optimized fuzzy controller are:

- taking into consideration variation senses in order to run for a whole day;
- brutal variations are eliminated (Fig. 20 at time 920);
- ease of use of controllers because of the small number of rules;
- easiness for the analysis of closed loop stability.

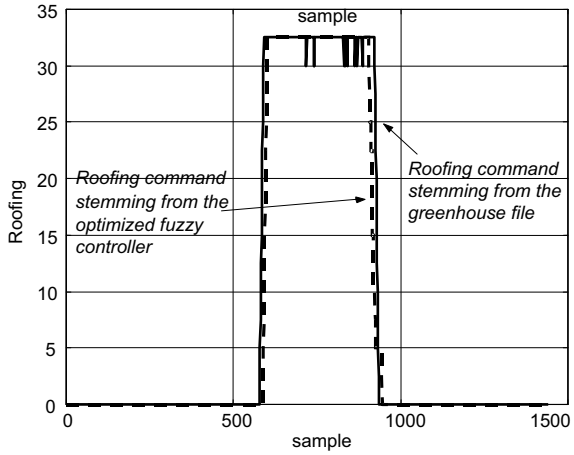


Fig. 19. Roofing command for the optimized fuzzy controller.

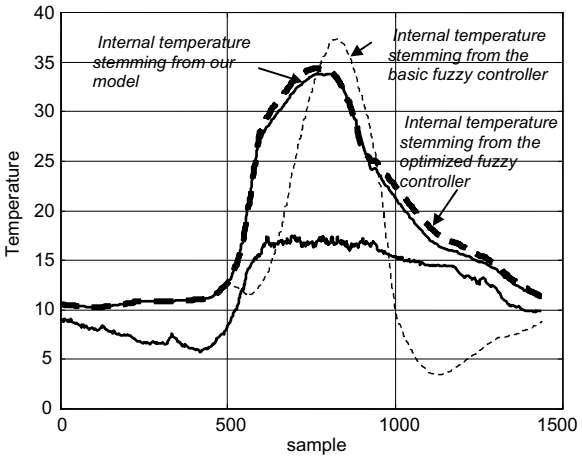


Fig. 20. Comparison of temperature curves.

The disadvantage is that this controller does not take into account disturbances (Te and He). Yet, we think that techniques of mathematical fusion, decentralized structure and hierarchical organization [9,13] will permit us to develop a complete fuzzy controller combining the advantages of the two controllers presented in this paper.

References

[1] P. Bortholet, D. Passaquay, S. Boverie, A. Titli, Iterative fuzzy modeling and control of nonlinear system using multidimensional fuzzy sets, Toulouse, LAAS, Report no 98373, September 1998.

[2] P. Bortholet, D. Passaquay, S. Boverie, A. Titli, Iterative fuzzy modeling and control of a waste-water treatment process, Toulouse, LAAS, Report no 99126, March 1999.

[3] G.P.A. Bot, Greenhouse climate: From physical processes to a dynamic model, Ph.D. Thesis, Agricultural University, Wagemingen, Netherlands, 1983.

[4] D. Dirankov, H. Hellendoom, M. Rcinfrank, An Introduction to Fuzzy Control, Springer, Berlin, 1993.

[5] J. Duplaix, Implantation et validation d'une identification en ligne pour serre agricole, DEA Informatique et automatique, University III, Marseille, 1991.

[6] E. Feki, R. M'hiri, M. Annabi, A. Ghorbel, Control of greenhouse climate, Proc. CESA'98, Computational Engineering in Systems Applications, Nabeul-Hammamet, Tunisia, April 1–4, 1998, pp. 663–666.

[7] R. Jager, Fuzzy logic in control, Thesis Technische Universiteit Delft, ISBN 90-9008318-9, 1995.

[8] A. Kandel, Y. Luo, Y-Q. Zhang, Stability analysis of fuzzy control systems, Fuzzy Sets and Systems 105 (1999) 33–48.

[9] V. Lacrose, A. Titli, Multivariable fuzzy control using rule-base reduction and self-tuning, Toulouse, LAAS, Report no 96112, April 1996.

[10] F. Lafont, J.-F. Balmat, Fuzzy control of the greenhouse climate, Int. Conf. Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications, ACIDCA'2000, vol. Systems Analysis and Automatic Control, Monastir, Tunisia, 22–24 March 2000, pp. 118–122.

[11] M. Margaliot, G. Langholz, Fuzzy Lyapunov-based approach to the design of fuzzy controllers, Fuzzy Sets and Systems 106 (1999) 49–59.

[12] L. Oueslati, Commande multivariable d'une serre agricole par minimisation d'un critère quadratique, Thesis, Laboratory MS/SSD, University of Toulon, June 1990.

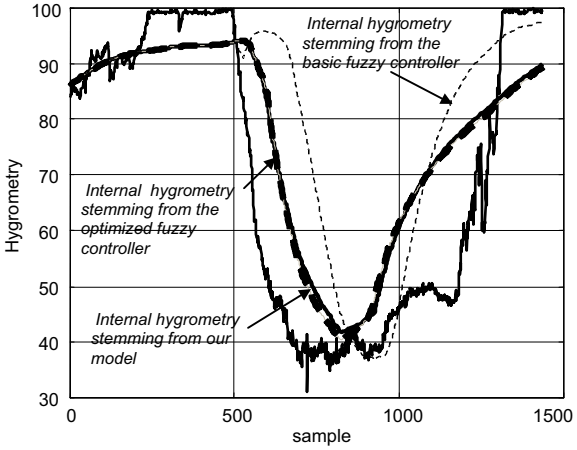


Fig. 21. Comparison of hygrometry curves.

[13] C. Raymond, S. Boverie, A. Titli, First evaluation of fuzzy MIMO control laws, Toulouse, LAAS, Report no 94554, June 1994.
[14] T. Takakura, K.A. Jordan, L.L. Boyd, Dynamic simulation of plant growth and environment in the greenhouse, Trans. ASAE 14 (1971) 964.
[15] A.J. Udink Ten Cate, Modeling and simulation in greenhouse climate control, Acta Horticulturae 174 (1985) 461–467.
[16] C. Viard Gaudin, Simulation et commande auto-adaptative d'une serre agricole, Thesis, University of Nantes, 1981.

5.2 Modélisation floue

Available online at www.sciencedirect.com
ScienceDirect
Applied Soft Computing 7 (2007) 1092–1101
Applied Soft Computing
www.elsevier.com/locate/asoc

Fuzzy identification of a greenhouse

Amine Trabelsi [a,*], Frederic Lafont [b], Mohamed Kamoun [a], Gilles Enea [b]

[a] Unité de Commande Automatique, Ecole Nationale d'ngénieurs de Sfax, B.P.W, 3038 Sfax, Tunisie, France
[b] Laboratoire LSIS, UMR CNRS 6168, Université du Sud-Toulon-Var, B.P 20132, 83957 La Garde Cedex 20, France

Received 10 June 2005; received in revised form 15 January 2006; accepted 12 June 2006
Available online 28 November 2006

Abstract

Nonlinear dynamic systems' modelling is difficult. The solutions proposed are generally based on the linearization of the process behaviour around the operating points. Other researches were carried out on this technique of linearization not only around the operating points, but also in all the input–output space allowing the obtaining of several local linear models. The major difficulty with this technique is the model transition. Fuzzy logic makes it possible to solve this problem thanks to its properties of universal approximator. Indeed, many techniques of modelling and identification based on fuzzy logic are often used for this type of systems. Among these techniques, we find those based on the fuzzy clustering technique. The proposed method uses in a first stage the fuzzy clustering technique to determine both the premises and the consequent parameters of the fuzzy Takagi–Sugeno rules. In a second stage these consequent parameters are adapted by using the recursive weighted least squares algorithm with a forgetting factor. We will try in this paper to apply this method to model the air temperature and humidity inside the greenhouse.

Keywords: Identification; MIMO systems; Fuzzy clustering; Adaptive fuzzy model; Greenhouse

1. Introduction

The agricultural greenhouses were used to protect the crop against the weather changes. With technical progress, the greenhouses have become a production means used to control the crop environment in order to obtain higher quality thus, making it possible to increase the economic benefit of the producer. Indeed, the producers aim is to minimize the production costs by reducing the consumption of water, fertilisers, CO_2 and energy.

Thus, the agricultural greenhouses objectives are:

- To obtain the highest productivity.
- To ensure a production quality which is in conformity with the commercial objectives by setting quality standards as for flowers: length and diameter of the floral stems, absence of deformation, colouring, etc.
- To control the calendars of production, i.e. to program the date of the beginning of the plant production and this can be achieved through the control of photosynthesis, breathing and the temperature cycles and alternations required by certain plants to be able to flower.
- To save energy. In fact, greenhouses are energy consuming, as they need to be heated, costs a lot to producers. In Europe, for example, the expenditure of heating represents between 10 and 30% of the running costs for the greenhouse crops. The reduction in the expenditure of energy should not be made at the expense of the productivity. But, an ''intelligent'' energy saving can be made with understanding well the heat transfers in the greenhouses and which also permits to be able to measure the heating installation correctly.

These four objectives can be achieved by developing, in a first stage a good prediction model of the inside air temperature and humidity, and in a second stage a control law to permit to these outputs to follow specific values depending on the plants nature. In this paper, we are interested only in the modelling phase.

This contribution deals with modelling a class of nonlinear dynamic processes by local linear models. The latter are Takagi–Sugeno fuzzy models [1]. The output of these fuzzy systems is calculated as an interpolation of locally valid linear models. On the one hand, this allows a linguistic interpretation of the fuzzy rules. On the other, classical linear control concepts can be applied to the local linear models [2,3]. The information process from the fuzzy models can be utilised by a great variety

* Corresponding author. Tel.: +33 216 98 25 18 98; fax: +33 216 74 21 28 16.
E-mail address: amintrabelsi@yahoo.fr (A. Trabelsi).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2006.06.009

of control methodologies. Indeed, the control performance strongly depends on the model accuracy. Hence, a great portion of the design effort has to be spent on modelling. Moreover, time-variant behaviour of the plant which is caused by disturbances or aging components should be considered in the process model. Therefore, an on-line adaptation of the process model is required. Here, the local linear models in the rule consequents of Takagi–Sugeno fuzzy models are updated. Assuming that the nonlinear structure of the process does not change significantly, the premises of the fuzzy rules are kept fixed and only the linear parameters in the consequents are locally updated by a recursive weighted least squares algorithm (RWLS).

The outline of this paper is as follows: first, a problem of MIMO systems modelling is introduced. In Section 3, the Takagi–Sugeno fuzzy model as well as a suitable off-line identification algorithm is presented. Section 4 describes the procedure of on-line adaptation of the fuzzy model including a forgetting factor. Section 5 shows the application of the proposed method on the greenhouse climate modelling. Section 6 concludes the paper.

2. Fuzzy process models

Modelling and identification are important steps in the design of control system. In fact, the establishment of a ''good'' model permits on the one hand to test a controller before its implementation in the real process and on the other hand to make possible to use it, as in an adaptive control scheme. Typical applications of these models are the simulation, the prediction or the control system design [4–6].

We consider a MIMO system with n_i inputs named u and n_o outputs named y. This system can be approximated by a set of discrete time fuzzy MISO models.

We consider also:

- Two polynomials $A(q^{-1})$ and $B(q^{-1})$ defined by:

$$\begin{aligned} A(q^{-1}) &= a_0 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{n_A} q^{-n_A} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_B} q^{-n_B} \end{aligned} \quad (1)$$

  q is a backward shift operator ($q^{-n}y(k) = y(k-n)$).
- Two integers m and n, $m \leq n$ which define a delayed sample of a discrete time signal as:

$$\{y(k)\}_m^n = [y(k-m), y(k-m-1), \ldots, y(k-n)]. \quad (2)$$

The MISO models are input–output Nonlinear Auto Regressive with eXogenous input (NARX) defined by:

$$y_l(k+1) = f_l(x_l(k)), \quad l = 1, 2, \ldots, n_o. \quad (3)$$

where the regression vector is given by:

$$x_l(k) = [\{y_1(k)\}_0^{n_{yl1}}, \{y_2(k)\}_0^{n_{yl2}}, \ldots, \{y_{n_0}(k)\}_0^{n_{yln_0}}, \{u_1(k)\}_{n_{dl1}}^{n_{ul1}}, \{u_2(k)\}_{n_{dl2}}^{n_{ul2}}, \ldots, \{u_{n_i}(k)\}_{n_{dln_i}}^{n_{uln_i}}]$$

n_y and n_u define the number of delayed outputs and inputs, respectively. n_d is the number of pure delays. n_y is a $n_o \times n_o$ matrix and n_u, n_d are $n_o \times n_i$ matrixes. f_l are unknown nonlinear functions.

For a nonlinear MIMO system, fuzzy Takagi–Sugeno (TS) models represent an efficient tool to model this kind of system [7].

3. Takagi–Sugeno type fuzzy models

The TS model has attracted the attention of many searchers. In fact, this model consists of if-then rules with fuzzy antecedents and mathematical functions in the consequent part [8]. The antecedents of fuzzy sets divide the input space into a number of fuzzy regions, while the consequent functions describe the system's behaviour in these regions [9,10].

MISO models are estimated of an independent manner, so, to simplify the notation, the output index l is omitted and we will be interested only in the multi-input, mono-output case.

The fuzzy rules are defined as:

$$\begin{aligned} R_i: \quad & \textbf{If } x(k) \textbf{ is } \Omega_i \\ & \textbf{then } y^i(k+1) = A_i y(k) + B_i u(k) + \alpha_i, \quad i = 1, 2, \ldots, K \end{aligned} \quad (5)$$

Ω_i are fuzzy variables, with trapezoidal, triangular, Gaussian or other membership functions representing a fuzzy subspace in which the implication R_i can be applied for reasoning. $A_i = [A_{i1}, \ldots, A_{in_o}]$, $B_i = [B_{i1}, \ldots, B_{in_i}]$ are vectors of polynomials, K is the rule's number and α_i is an offset coefficient. Here, we choose Gaussian membership functions because they are the more used in the establishment of a Takagi–Sugeno fuzzy model by fuzzy clustering technique. In fact, with this method, the membership functions are multidimensional. It is not desirable that the transitions between every cluster are linear (because in that case, the commutation logic is nearly Boolean). So, we use Gaussian membership functions which need only two parameters (centers and standard deviations).

The antecedent of (Eq. (10)) can be written:

$$\begin{aligned} R_i: \quad & \textbf{if } x_1(k) \textbf{ is } \Omega_{i1} \textbf{ and} \cdots \textbf{and } x_p(k) \textbf{ is } \Omega_{ip} \\ & \textbf{then } y^i(k+1) = A_i y(k) + B_i u(k) + \alpha_i \\ & i = 1, 2, \ldots, K \end{aligned} \quad (6)$$

where

$$p = \sum_{j=1}^{n_o} n_{y_j} + \sum_{j=1}^{n_i} n_{u_j} + 1. \quad (7)$$

The task of system identification is to determine both the nonlinear parameters of the antecedents and the linear parameters of the consequent of the rules.

In general, there are two ways to obtain this information. Human experts may be able to formulate their process knowledge in fuzzy rules. However, this method is often inefficient because human cannot detect all the details. Therefore, numerous approaches have been proposed [11] which compute nonlinear dynamic fuzzy models from input/output data measurement, e.g., local linear model tree method (LOLIMOT), tree construction algorithms [12], or neuro-fuzzy approaches [13].

The output of TS model is computed as:

$$y(k+1) = \frac{\sum_{i=1}^{K} \mu_i(x(k)) y^i(k+1)}{\sum_{i=1}^{K} \mu_i(x(k))} \tag{8}$$

or

$$y(k+1) = \sum_{i=1}^{K} y^i(k+1) \Phi_i(x, c_i, \sigma_i) \tag{9}$$

where $\Phi_i(x, c_i, \sigma_i)$ is the validity function for the Gaussian membership functions with centers c_i and standard deviations σ_i defined as:

$$\Phi_i(x, c_i, \sigma_i) = \frac{\mu_i(x(k))}{\sum_{i=1}^{K} \mu_i(x(k))} \tag{10}$$

$$\mu_i(x(k)) = \prod_{j=1}^{p} \exp\left(-\frac{1}{2} \frac{(x_j - c_{ij})^2}{\sigma_{ij}^2} \right) \tag{11}$$

$\mu_i(x(k))$ is the degree of fulfillment of the rule i.

The structure of the model, i.e. the matrixes n_y, n_u and n_d are determined by the user on the basis of system's prior knowledge and/or by comparison of different structures based on error criteria [8,12]. Once the structure is fixed, the n_o MISO parameters are estimated independently by fuzzy clustering [14].

The model identification procedure based on the proposed method consists of two distinct steps.

In the first step, called off-line identification of the fuzzy model, both nonlinear parameters of the Gaussian membership functions, namely the centers c_i and standard deviations σ_i, and the linear parameters of the local models are determined by fuzzy clustering method.

In the second step, called on-line adaptation of the fuzzy model, the consequence's parameters of fuzzy rules are adapted by a recursive weighted least squares method [3].

3.1. Off-line identification of the fuzzy model

This procedure is carried out into four steps:

- construction of the regression data,
- determination of the clusters corresponding to a set of local linear submodels,
- determination of the antecedent membership function from the cluster parameters,
- estimation of rule's consequence parameters.

3.1.1. Regression data

The available data samples are collected in a matrix Z formed by concatenating the regression matrix X and the output vector Y:

$$X = \begin{bmatrix} x(1) \\ x(k) \\ x(N-1) \end{bmatrix}, \qquad Y = \begin{bmatrix} y(2) \\ y(k+1) \\ y(N) \end{bmatrix}, \qquad Z^{\mathrm{T}} = [X \quad Y]. \tag{12}$$

N is the number of data samples.

3.1.2. Construction of the fuzzy clusters

There are various algorithms to construct the fuzzy clusters such as: the C-means algorithm [15], the Gath–Geva algorithm [16] and the Gustafson–Kessel algorithm [17] which will be used in our contribution.

Through clustering, the data set Z is partitioned into Nc clusters. In this paper, Nc is determined by testing many values according to an error criterion. The result is a fuzzy partition matrix $U = [\mu_{ik}]_{\mathrm{Nc} \times N}$, whose element $\mu_{ik} \in [0, 1]$ represents the degree of membership of the observation in cluster i, a prototype matrix $V = [v_1, \ldots, v_{\mathrm{Nc}}]$ and a set of cluster covariance matrixes $F = [F_1, \ldots, F_{\mathrm{Nc}}]$ ($\{F_i\}$ are definite positive matrixes).

Once the triplet (U, V, F) is determined, the parameters of the rules premises (c_i and σ_i) and the consequent parameters (A_i, B_i and α_i) are computed. For more details, see [18].

3.1.3. Determination of the antecedent membership functions from the cluster parameters

In this paper, Gaussian membership functions are used to represent the fuzzy sets Ω_{ij}:

$$\Omega_{ij}(x_j(k)) = \exp\left(-\frac{1}{2} \frac{(x_j - c_{ij})^2}{\sigma_{ij}^2} \right) \tag{13}$$

This choice leads to the following compact formula for (11):

$$\begin{aligned} \mu_i(x(k)) &= \Omega_i(x(k)) \\ &= \exp\left(-\frac{1}{2} (x(k) - c_i^x)^{\mathrm{T}} (F_i^{xx})^{-1} (x(k) - c_i^x) \right) \end{aligned} \tag{14}$$

where $c_i^x = [c_{1i}, \ldots, c_{pi}]$ denotes the center of the ith multivariate Gaussian and F_i^{xx} stands for a diagonal matrix that contains σ_{ij}^2 variances.

3.1.4. Estimation of the consequent parameters

The consequent parameters in each rule are estimated separately by the weighted least squares method by minimizing the following criterion [18]:

$$\min_{\theta_i} \frac{1}{N} (Y - X_e \theta_i)^{\mathrm{T}} Q_i (Y - X_e \theta_i) \tag{15}$$

where $X_e = [X \quad 1]$ is the regression matrix extended by a unitary column and Q_i is a matrix containing the values of the validity functions Φ_i of the ith local model for each data sample:

$$Q_i = \begin{bmatrix} \Phi_i(x(1), c_i \sigma_i) & 0 & \cdots & 0 \\ 0 & \Phi_i(x(2) c_i, \sigma_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_i(x(N), c_i, \sigma_i) \end{bmatrix} \tag{16}$$

The weighted least squares estimated of the consequent parameters ($\theta_i = A_i$, B_i, α_i) is given by:

$$\theta_i = [X_e^{\mathrm{T}} Q_i X_e]^{-1} X_e^{\mathrm{T}} Q_i Y \tag{17}$$

4. On-line adaptation of the fuzzy model

There are two reasons for applying on-line identification. First, a too simplistic (e.g. linear) model may be used, which is only capable of describing the process behaviour within a small operating regime. The need for on-line adaptation then emerges from the process nonlinearities that are not represented by the model. This strategy is employed in classical linear adaptive control [19]. However, the second reason for the requirement of on-line adaptation is time-variant behaviour of the process. This problem, addressed here, equally exists for both linear and nonlinear models.

A Takagi–Sugeno fuzzy model possesses nonlinear parameters which determine the rule premises (centres and standard deviations of the validity functions) and linear parameters which determine the rule consequents. For on-line adaptation, the following strategy is pursued. It is assumed that at least a rough model has been identified off-line with the Gustafson–Kessel fuzzy clustering algorithm.

In the on-line phase the rule premises are kept fixed and only the rule consequents are adapted. The advantage of this approach is the exploitation of mature, computationally effective and numerically robust linear recursive algorithms. A drawback lies in the requirement of an off-line model and in its limited structural flexibility. Since many time-variant processes such as the thermal processes do not significantly change their nonlinear structure over time but only their gains, time constants or zeros the latter drawback might not be very severe. In general however, if one does not have any prior knowledge of the process, it will not be known to what extent the nonlinear structure of the time-variant process changes. But even if the structural properties of the process change markedly, the proposed approach can be expected to perform better than a linear adaptive model because this is the least flexible case of the Takagi–Sugeno fuzzy model with only one rule.

Generally, the fuzzy TS models obtained by clustering are constant consequence parameters, i.e. a rule's consequence is written as:

$$y^i(k+1) = A_i y(k) + B_i u(k) + \alpha_i \tag{18}$$

But in our case, these parameters are updated. It means that at every moment k, one obtains a TS model:

$$y^i(k+1) = A_i(k) y(k) + B_i(k) u(k) + \alpha_i(k) \tag{19}$$

It is, for example, possible to utilise the following weighted recursive least squares algorithm with forgetting factor λ to estimate the parameters of each local linear model.

$$\theta_j(k) = \theta_j(k-1) + \delta_j(k)(y(k) - x^{\mathrm{T}}(k)(\theta_j(k-1)) \tag{20}$$

$$\delta_j(k) = \frac{P_j(k-1)x(k)}{x^{\mathrm{T}}(k)P_j(k-1)x(k) + \lambda/\Phi_j(x(k), c_j, \sigma_j)} \tag{21}$$

$$P_j(k) = \frac{1}{\lambda}[I - \delta_j(k)x^{\mathrm{T}}(k)]P_j(k-1). \tag{22}$$

In (20), the parameter vector θ_j is the same as for off-line identification in (17). It is updated by adding a correction vector to the old estimate $\theta_j(k-1)$. In (21) and (22), λ is a forgetting factor that implements forgetting of the old measurements, Φ_j is the weighting of the actual data with the rule activation and P_j is a matrix of the adaptation gain.

5. Greenhouse climate modelling

The main objective of greenhouse crop production is to increment the economic benefits of the farmer compared to traditional agriculture methods [20]. The implementation of an adequate automatic control system for controlling the climate of the greenhouse (temperature, humidity) can lead to an increased production and quality of the horticultural products, reducing pollution and energy consumption [21].

In recent years, there have been many researches on analysis and control of the environment inside greenhouses [2,22–28]. To do this control efficiently, it is necessary to elaborate models that adequately describe the system to be controlled.

Generally, there are three categories of models that could be used to simulate and predict the greenhouse environment.

The first category is based on the concept of energy and mass balance [29,30]. The drawback of this methodology is that these models are difficult to tune in practice, since they use a large number of parameters and physical variables which are time-variant and weather-dependent.

The second category is based on Soft computing and computational intelligence such as artificial neural networks and fuzzy clustering. These techniques are applicable when the expert knowledge is not available and only input–output data of the system is available.

Seginer et al. [31] and Boaventura Cunha [32] use black-box neural network models to model greenhouse climate. Linker et al. [33] pointed to problems associated with the poor extrapolation of such models (the need for extrapolation may arise from weather conditions and/or new control actions), which is due to the absence of prior knowledge.

Often, such prior knowledge exists. Consequently, Linker and Seginer [34] propose the third category of models called the hybrid model which combines physical and neural network model in series or in parallel configurations.

Sigrimis and Rerras [35] elaborate a linear model to greenhouse climate modelling as follows:

$$[y_{k+1}^{\mathrm{T}}] = \begin{bmatrix} y_k^{\mathrm{T}} & u_k^{\mathrm{T}} & w_k^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} A^{\mathrm{T}} \\ B^{\mathrm{T}} \\ C^{\mathrm{T}} \end{bmatrix} + [v_k^{\mathrm{T}}]. \tag{23}$$

where y is the output, u the input and w is the disturbance.

The compact form becomes:

$$Y_N = \Psi_N \theta + \varepsilon_N(\theta). \tag{24}$$

where $\Psi = \begin{bmatrix} y^{\mathrm{T}} & u^{\mathrm{T}} & w^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ is the matrix formed by measured data that grows in time, Y the matrix formed by output observations and ε the matrix formed by errors associated with the estimate θ for each observation y.

The identification problem then can be stated as follows: given a set of N input–output observations, it is desirable to

estimate the matrix θ of system parameters that will provide the best linear fit to the observed data.

In this paper, we try to elaborate a fuzzy model for greenhouse climate modelling. To justify the use of the nonlinear model, we make a comparison between our model and the Sigrimis linear model. In fact, in the beginning, we compare the model's outputs with those from greenhouse data file (cf. Figs. 10 and 11) then, with a Sigrimis linear model (cf. Figs. 12 and 13).

5.1. Greenhouse presentation

In the laboratory, we have an experimental agricultural greenhouse in order to search and to develop a control law to regulate the temperature and the humidity inside it.

To do this control, first a model must be elaborated and we choose an adaptive fuzzy model-based control to test an adaptive fuzzy controller (Fig. 1).

This greenhouse (Fig. 2) has a floor area of 80 m^2 and equipped with many sensors:

- sensors of internal and external temperature, data expressed in °C and named T_i and T_e,
- sensors of internal and external humidity, data expressed in % and named H_i and H_e,
- sensor of solar radiation, data expressed in W m^{-2} and named Ray,
- sensor of wind speed, data expressed in km h^{-1} and named V_v,

and various actuators:

- gas heating system with a heating power of 58 kW, binary command named C_h,
- roofing, expressed in percent (50% maximum) named Ouv,
- moistening, binary command named Bru,
- shadow/thermal screen, expressed in percent named Rid.

5.2. Sigrimis model

Sigrimis and Rerras [35] proposed also a recursive weighted least square algorithm to estimate the linear model (Eq. (24)):

(1) Initialize: θ and variance P (e.g., $\theta = 0$ and $P(k) = 10I$).

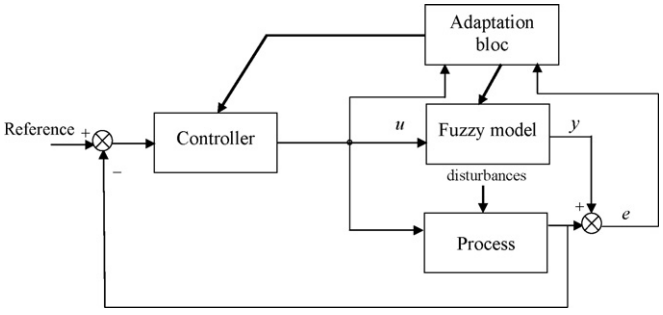


Fig. 1. Fuzzy model-based control with on-line adaptation.

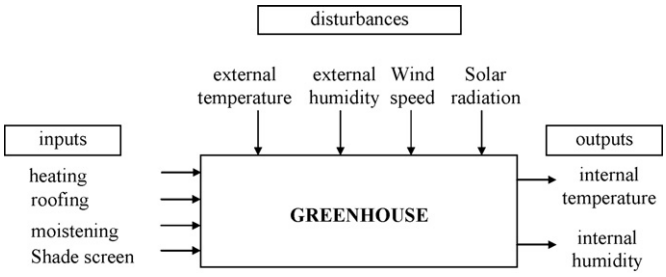


Fig. 2. Block diagram of the greenhouse.

(2) Gain:

$$L(k+1) = P(k)\Psi(k+1)\left[\frac{\delta}{\alpha} + \Psi^{T}(k+1)P(k)\Psi(k+1)\right]^{-1}. \quad (25)$$

(3) Estimate:

$$\theta(k+1) = \theta(k) + L(k+1)[y(k+1) - \Psi^{T}(k+1)\theta(k)]. \quad (26)$$

(4) Variance:

$$P(k+1) = \frac{[I - L(k+1)\Psi^{T}(k+1)]P(k)}{\delta}. \quad (27)$$

(5) Repeat: get new Ψ, $k \leftarrow k+1$ and repeat from step 2.

They recommended to use a forgetting factor $\delta = 0.99$ and to start with a stability factor $\alpha = 10$. If noise is minimal and measurements carry sufficient information for all parameters then lower values for α (i.e. $\alpha = 5$ or $\alpha = 2$) might be used.

5.3. Greenhouse fuzzy modelling

There are some researches concerning the identification of a greenhouse by fuzzy logic such as the method elaborated by [5] and called ''Iterative fuzzy modelling''.

Due to the large amount of collected data (perturbations, commands) and the fact that the observed data reflect a nonlinear nature, a MIMO model based on fuzzy logic is being

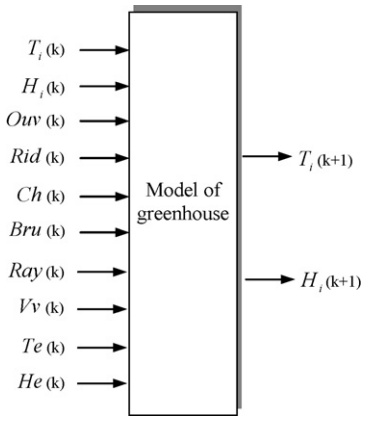


Fig. 3. Structure of model (for training).

developed to obtain a reliable prediction of the greenhouse climate based on fuzzy clustering method. This method was explained in Section 3.1. We take the data of a whole day for the training according to the model structure shown in Fig. 3.

The sampling time is equal to 3 min.

The rules are linear conclusions of system's inputs, for example, for the rule j:

$$\begin{aligned} T_i^j(k+1) &= a_{j1T}T_i(k) + a_{j2T}H_i(k) + b_{j1T}\mathrm{Ouv}(k) \\ &\quad + b_{j2T}\mathrm{Rid}(k) + b_{j3T}C_{\mathrm{h}}(k) + b_{j4T}\mathrm{Bru}(k) \\ &\quad + b_{j5T}\mathrm{Ray}(k) + b_{j6T}V_{\mathrm{v}}(k) + b_{j7T}T_{\mathrm{e}}(k) \\ &\quad + b_{j8T}H_{\mathrm{e}}(k) + \alpha_{j1} \end{aligned} \tag{28}$$

$$\begin{aligned} H_i^j(k+1) &= a_{j1H}T_i(k) + a_{j2H}H_i(k) + b_{j1H}\mathrm{Ouv}(k) \\ &\quad + b_{j2H}\mathrm{Rid}(k) + b_{j3H}C_{\mathrm{h}}(k) + b_{j4H}\mathrm{Bru}(k) \\ &\quad + b_{j5H}\mathrm{Ray}(k) + b_{j6H}V_{\mathrm{v}}(k) + b_{j7H}T_{\mathrm{e}}(k) \\ &\quad + b_{j8H}H_{\mathrm{e}}(k) + \alpha_{j2} \end{aligned} \tag{29}$$

with $a_{j1T}, a_{j2T}, b_{j1T}, \ldots, b_{j8T}, \alpha_{j1}$: consequence parameters for the temperature and for the rule j and $a_{j1H}, a_{j2H}, b_{j1\mathrm{H}}, \ldots, b_{j8\mathrm{H}}, \alpha_{j2}$: consequence parameters for the hygrometry and for the rule j.

In order to select a model, a set of possible models, with different n_y, n_u and n_d, must first be chosen, and then the best model selected. There are several criteria that can be employed, but the most commonly used is Akaike's Information theoretic Criterion (AIC).

Here, we choose the structure below on the one hand so as not to increase the number of inputs and on the other hand, to refer to previous works [36]:

$$n_y = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}; \quad n_u = n_d = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{30}$$

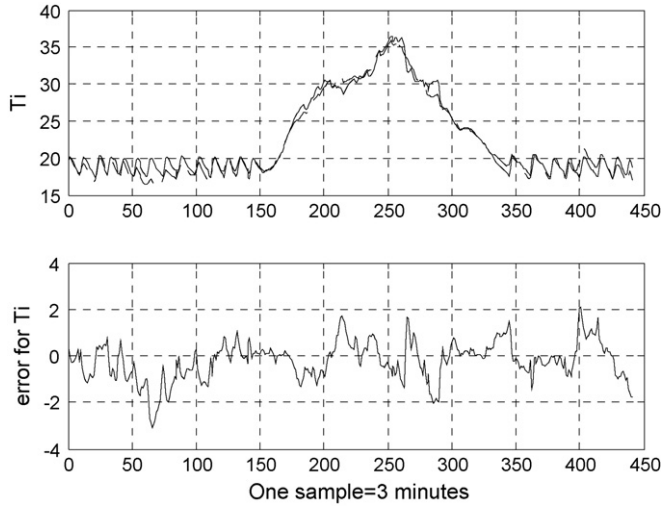


Fig. 4. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for identification of T_{i}.

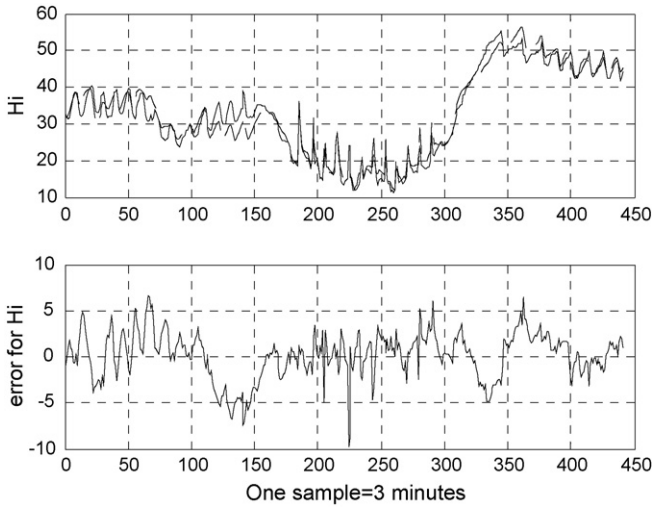


Fig. 5. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for identification of H_{i}.

We define:

$$x(k) = \begin{bmatrix} T_i(k) & H_i(k) & \mathrm{Ouv}(k) & \mathrm{Rid}(k) & C_{\mathrm{h}}(k) & \\ & \mathrm{Bru}(k) & \mathrm{Ray}(k) & V_{\mathrm{v}}(k) & T_{\mathrm{e}}(k) & H_{\mathrm{e}}(k) \end{bmatrix}^{\mathrm{T}} \tag{31}$$

The data matrix is defined by Eq. (12).

Once the data matrix is available, we compute the matrix U, V and F according to Eqs. (14) and (15) by Gustafson–Kessel algorithm and then we determine the consequence parameters of each rule generated by fuzzy clustering according to Eq. (17).

The responses and the error of the process and the model output for data file of the day (January 10, 2004) are shown in Figs. 4 and 5.

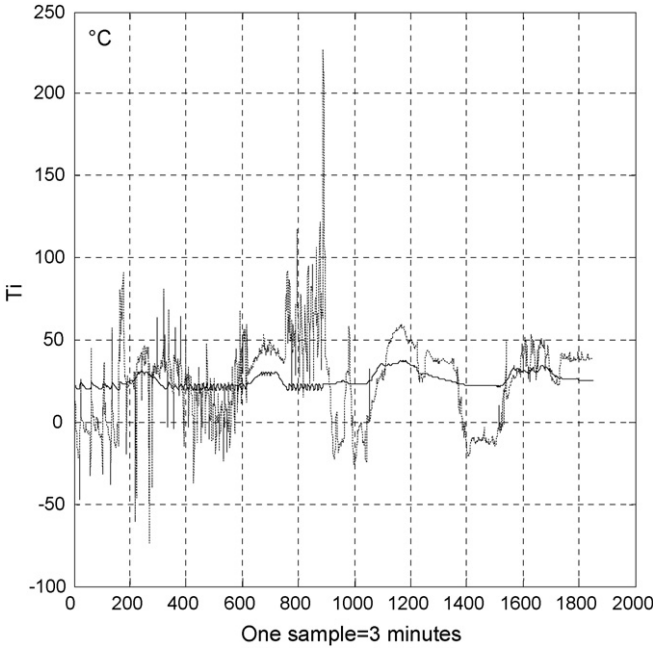


Fig. 6. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for validation without aggregation for T_{i}.

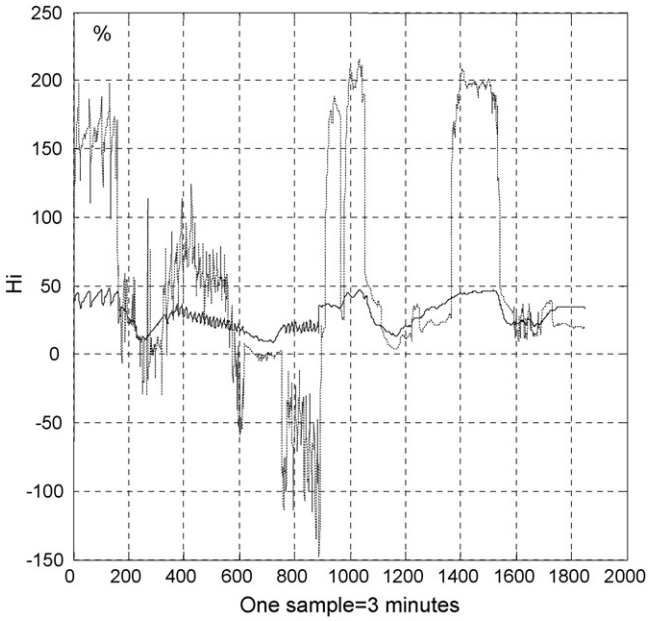


Fig. 7. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for validation without aggregation for H_i.

To valid the established model, we apply other data comprising four data files (from January 14–15 to September 14–15).

In the first stage, we identify linear local models and apply them as such single in the region, without aggregation and without adaptation. The responses and the corresponding error of the process and model output without aggregation to these new data are shown in Figs. 6 and 7.

Now, we apply the fuzzy aggregation (Eq. (8)). The responses and the corresponding error of the process and model output with aggregation and without adaptation to these new data are shown in Figs. 8 and 9.

These responses are obtained from local models computed by fuzzy clustering without adaptation of the consequence parameters. According to the errors resulting between the two

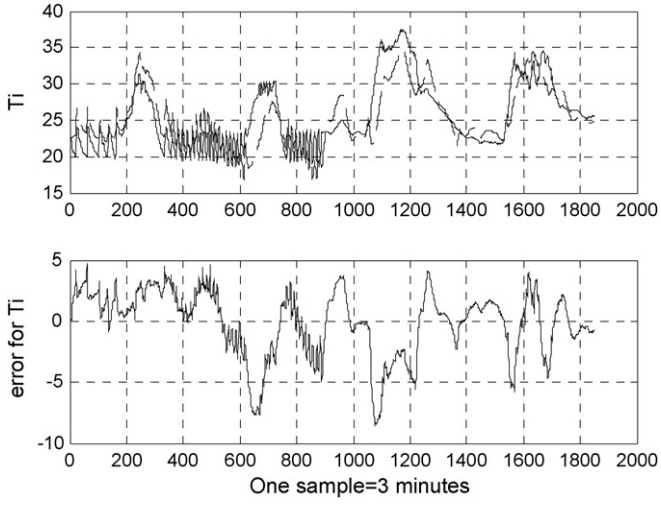


Fig. 8. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for validation with aggregation for T_i.

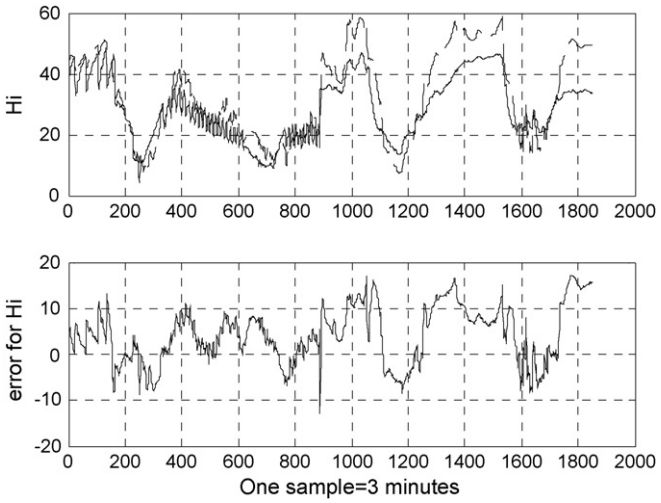


Fig. 9. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for validation with aggregation for H_i.

outputs in Figs. 6 and 8 and Figs. 7 and 9, we can remark that the estimated outputs with fuzzy aggregation are better than those without aggregation. It is also noticed that these outputs cannot follow the process' outputs well and the error resultant is rather big. To improve the quality of the established fuzzy model, the parameters of the rules consequences are adapted by a recursive least squares algorithm with forgetting factor ($\lambda = 0.99$) according to Eq. (20).

So, the rules conclusions are written as, for example, for the rule j:

$$\begin{aligned} T_i^j(k+1) = {} & a_{j1T}(k)T_i(k) + a_{j2T}(k)H_i(k) + b_{j1T}(k)\mathrm{Ouv}(k) \\ & + b_{j2T}(k)\mathrm{Rid}(k) + b_{j3T}(k)C_\mathrm{h}(k) + b_{j4T}(k)\mathrm{Bru}(k) \\ & + b_{j5T}(k)\mathrm{Ray}(k) + b_{j6T}(k)V_\mathrm{v}(k) + b_{j7T}(k)T_\mathrm{e}(k) \\ & + b_{j8T}(k)H_\mathrm{e}(k) + \alpha_{j1}(k) \end{aligned} \tag{33}$$

$$\begin{aligned} H_i^j(k+1) = {} & a_{j1H}(k)T_i(k) + a_{j2H}(k)H_i(k) + b_{j1H}(k)\mathrm{Ouv}(k) \\ & + b_{j2H}(k)\mathrm{Rid}(k) + b_{j3H}(k)C_\mathrm{h}(k) \\ & + b_{j4H}(k)\mathrm{Bru}(k) + b_{j5H}(k)\mathrm{Ray}(k) \\ & + b_{j6H}(k)V_\mathrm{v}(k) + b_{j7H}(k)T_\mathrm{e}(k) + b_{j8H}(k)H_\mathrm{e}(k) \\ & + \alpha_{j2}(k) \end{aligned} \tag{34}$$

Figs. 10 and 11 show the responses and the corresponding error of the process and the model output for the same 4 files data but with adaptation.

We define a function VAF which computes the percentile Variance Accounted For between two signals as follows [37]:

$$\mathrm{VAF} = \max\left\{1 - \frac{\mathrm{var}(y_1 - y_2)}{\mathrm{var}(y_1)}, 0\right\} \times 100\% \tag{35}$$

y_1 is the output of the process and y_2 is the output of the model.

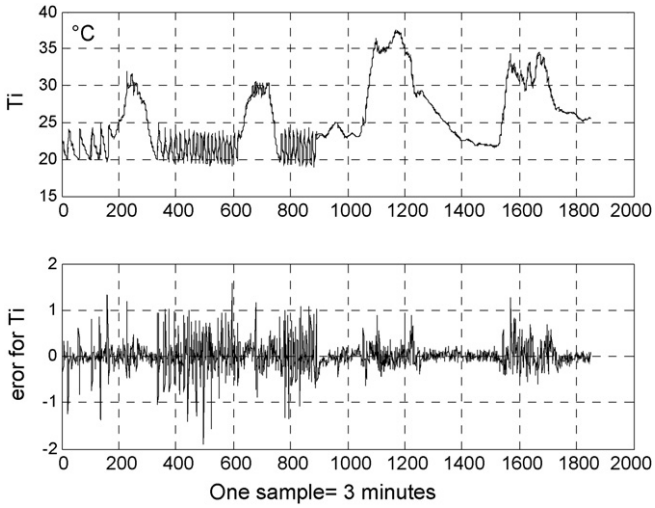


Fig. 10. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for validation with adaptation for T_i.

Table 1
Comparison of the prediction accuracy of the TS fuzzy model in three phases

	Identification	Validation without adaptation	Validation with adaptation
	February 09	January 14–15 to September 14–15	January 14–15 to September 14–15
VAF_T_i (%)	97.66	63.58	99.63
VAF_H_i (%)	94.94	64.67	99.50

The VAF of two equal signals is 100%. If the signals differ, VAF is lower.

Table 1 gives the VAF performance index for the responses of the process and the identification model in the identification phase, the validation phase without adaptation and in the validation phase with adaptation.

From Table 1, we can see that the adapted fuzzy TS model is more accurate than a nonadapted fuzzy TS model.

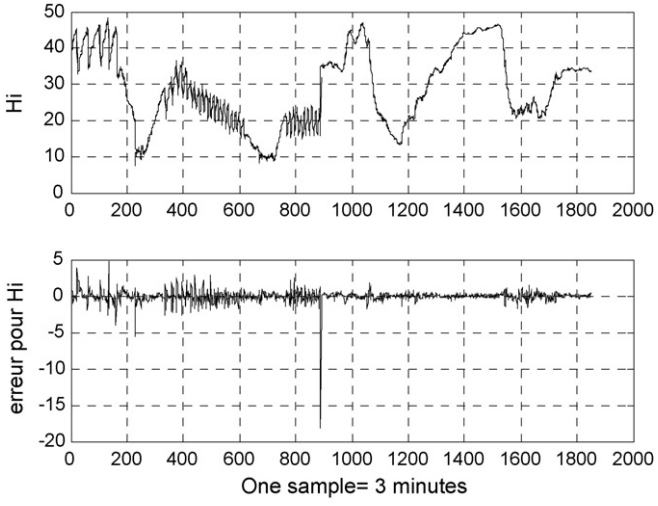


Fig. 11. Comparison of the process output (solid line) with the fuzzy model output (dashed-dotted line) for validation with adaptation for H_i.

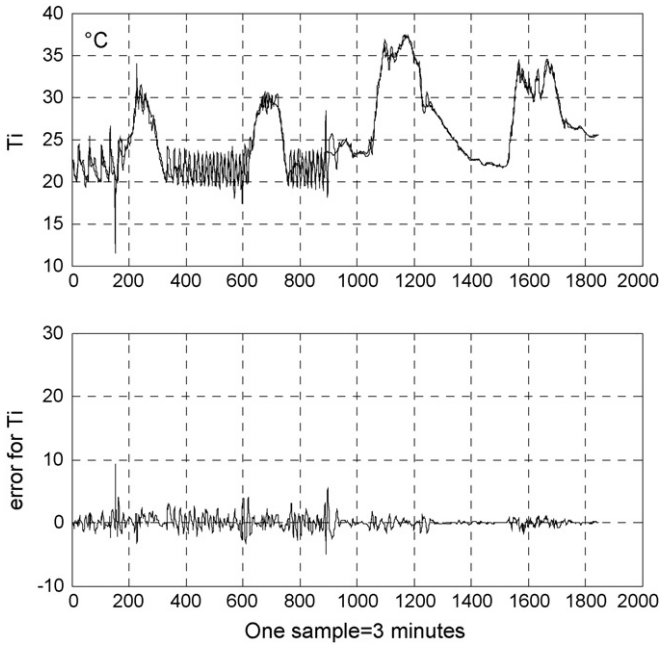


Fig. 12. Comparison of the process output (solid line) with the Sigrimis model output (dashed-dotted line) for T_i.

To justify the use of the fuzzy model, we compare it with the linear "Sigrimis model" (Eq. (24)) with the data file (from January 14–15 to September 14–15).

Figs. 12 and 13 show the responses and the corresponding error of the process and the "Sigrimis model" output.

Table 2 gives the VAF performance index for the responses of the process and the identification model in the validation phase with adaptation for the fuzzy model and the "Sigrimis model".

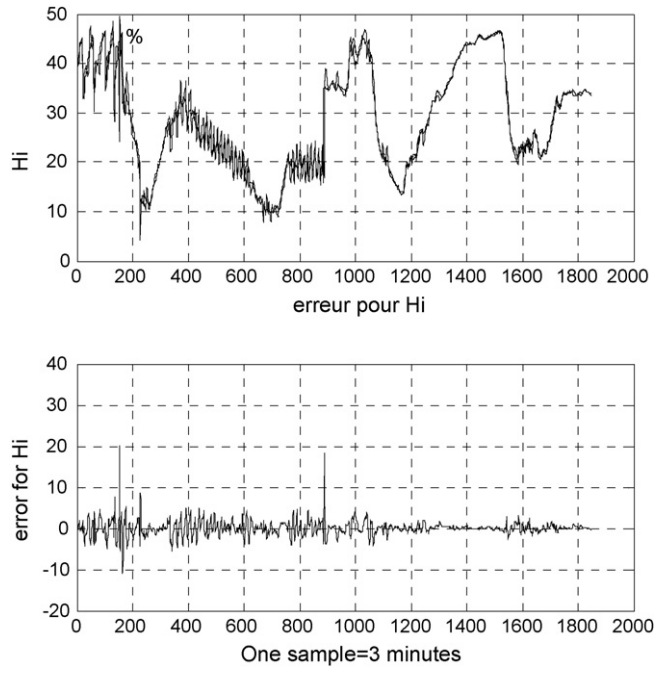


Fig. 13. Comparison of the process output (solid line) with the Sigrimis model output (dashed-dotted line) for H_i.

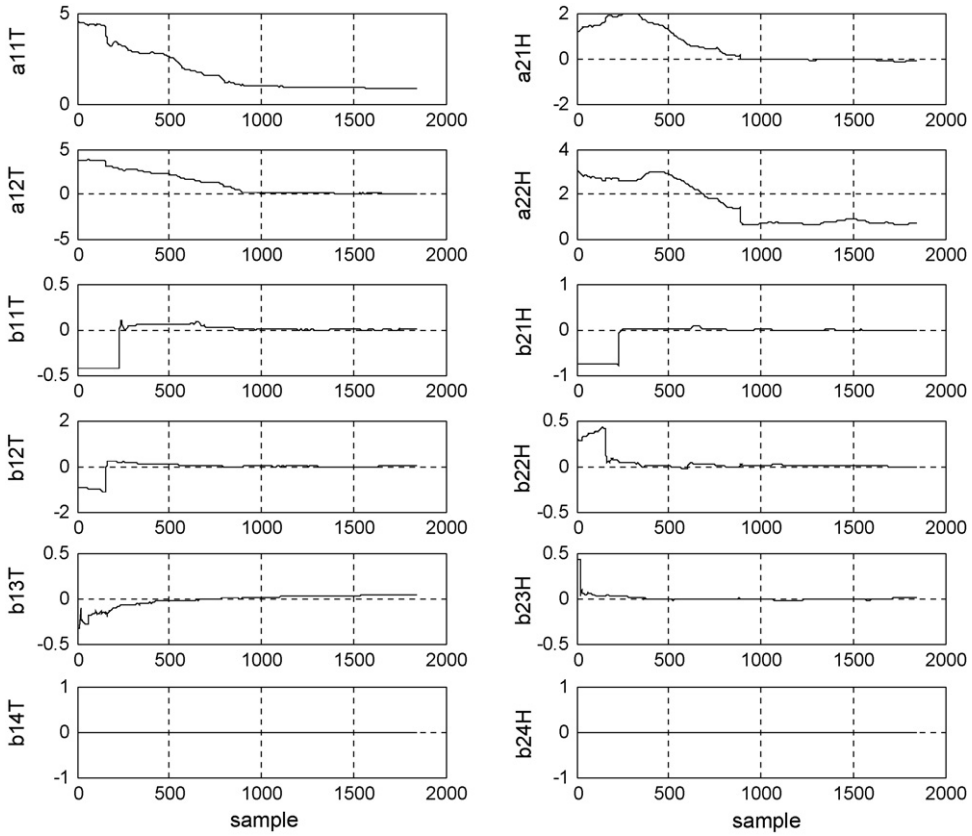


Fig. 14. Evolution of some parameters of polynomial A_i and B_i for T_i and H_i.

Table 2
Comparison of the prediction accuracy of the TS fuzzy model and "Sigrimis model"

	Fuzzy model ($\lambda = 0.99$)	Sigrimis model ($\alpha = 2$, $\delta = 0.99$)	Sigrimis model ($\alpha = 5$, $\delta = 0.99$)	Sigrimis model ($\alpha = 10$, $\delta = 0.99$)	Sigrimis model ($\alpha = 2$, $\delta = 1$)	Sigrimis model ($\alpha = 2$, $\delta = 0.98$)
	January 14–15 to September 14–15	January 14–15 to September 14–15	January 14–15 to September 14–15	January 14–15 to September 14–15	January 14–15 to September 14–15	January 14–15 to September 14–15
VAF_T_i (%)	99.63	93.05	93.04	93.05	85.25	93.61
VAF_H_i (%)	99.50	95.75	95.69	95.68	94.64	93.95

From Table 2, Figs. 12 and 13, we can see that the "Sigrimis model" is less accurate than our fuzzy model for different values of α and δ. In fact, the Sigrimis model is a particular case of our fuzzy model which corresponds to one cluster. We equally remark that Sigrimis model outputs for the best values of α and δ present for both temperature and humidity some peaks which are not suitable.

To verify the applicability of our method to the real process, we must show the evolution of some parameters of polynomial A_i and B_i during the adaptation. In fact, the parameters of polynomial B_i correspond to the gain of command (Ouv, C_h, Rid, Bru) which must not change their values in great proportions.

Fig. 14 shows the evolution of some parameters of polynomial A_i and B_i for the outputs T_i and H_i, respectively, for the first cluster (rule) of the four files data.

From Fig. 14, we can notice that each linear parameter varies until a certain value. After that, they are practically constant and it seems to be null. In reality, these parameters began with their values in the identification phase, then they are adapted and converge to their final values, which are weak.

6. Conclusion

This paper proposes a study on the application of the fuzzy method to the identification problem of MIMO process. This method is based on the fuzzy clustering technique using the Takagi–Sugeno (TS) fuzzy models. The local models automatically obtained, are adapted by the weighted recursive least squares algorithm with forgetting factor.

The performance of the proposed technique is demonstrated on the air temperature and humidity inside greenhouse modelling.

The obtained results are satisfactory and we think to insert the elaborated model in an adaptive control scheme to ensure an increased production and quality of the horticultural products and reducing pollution and energy consumption. This will be the object of our forthcoming work.

Acknowledgment

I would like to thank Mr. Duplaix, professor in automatic control to have provided us the greenhouse data files containing the different data which are used in our simulation for greenhouse modelling.

References

[1] K. Zeng, N.-Y. Zhang, W.-L. Xu, A comparative study on sufficient conditions for Takagi–Sugeno fuzzy system as universal approximation, IEEE Trans. Fuzzy Syst. 8 (2000) 773–780.
[2] J.C. Bakker, G.P.A. Bot, H. Challa, N.J. van de Braak, Greenhouse Climate Control: An Integrated Approach, Wageningen Pers, Wageningen, the Netherlands, 1995.
[3] A. Fink, M. Fischer, O. Nelles, Supervision of nonlinear adaptive controllers based on fuzzy models, Control Eng. 8 (2000) 1093–1105.
[4] P. Salgado, J. Boaventura Cunha, Greenhouse climate hierarchical fuzzy modelling, Control Eng. Pract. 13 (2004) 613–628.
[5] F. Lafont, J.F. Balmat, Fuzzy logic to the identification and the command of the multidimensional systems, Int. J. Comput. Cognition 2 (3) (2004) 21–47.
[6] A. Trabelsi, F. Lafont, M. Kamoun, G. Enéa, Identification of nonlinear multivariable systems by adaptive fuzzy Takagi–Sugeno model, Int. J. Comput. Cognition 2 (3) (2004) 137–153.
[7] F. Lafont, J.F. Balmat, Optimized fuzzy control of a greenhouse, Fuzzy Sets Syst. 128 (2002) 47–59.
[8] T. Takagi, M. Sugeno, Fuzzy identification of systems and its application to modeling and control, IEEE Trans. Syst. Man Cybernetics 15 (1) (1985) 16–132.
[9] P.Y. Glorennec, Algorithmes d'apprentissage pour systèmes d'inférence floue, Hermes Sciences Publications, Paris, 1999.
[10] F. Lafont, J.-F. Balmat, Fuzzy control of the greenhouse climate, in: International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications, ACIDCA'2000 Vol. Systems Analysis and Automatic Control, Monastir, Tunisia, 22–24 March, (2000), pp. 118–122.
[11] O. Nelles, A. Fink, R. Isermann, Local linear model trees (LOLIMOT) toolbox for nonlinear system identification, in: 12th IFAC Symposium on System Identification (SYSID), Santa Barbara, USA, 2000.
[12] M. Sugeno, G.T. Kang, Structure identification of fuzzy model, Fuzzy Sets Syst. 28 (1987) 15–33.
[13] J.S.R. Jang, ANFIS: adaptive-network based fuzzy inference system, IEEE Trans. Syst. Man Cybernetics 23 (3) (1993) 665–685.
[14] R. Babuska, H.B. Vebruggen, Identification of composite linear models via fuzzy clustering, in: Proceedings European Control Conference, vol. 4, Rome, Italy, (1995), pp. 1593–1606.
[15] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[16] I. Gath, A.B. Geva, Unsupervised optimal fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 7 (1989) 773–781.
[17] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with fuzzy covariance matrix, in: Proceedings IEEE CDC, San Diego, (1979), pp. 761–766.
[18] R. Babuska, H.B. Vebruggen, An overview of fuzzy modeling for control, Control Eng. Pract. 4 (11) (1996) 1593–1606.
[19] R. Isermann, K.H. Lachmann, D. Matko, Adaptive Control Systems, Prentice Hall, Englewood Cliffs, 1992.
[20] N. Sigrimis, G. Papageorgiou, Intelligent electronic leaf sensor, J. Agric. Eng. Res. 58 (1994) 169–180.
[21] J. Boaventura Cunha, C. Couto, A.E. Ruano, Real-time parameter estimation of dynamic temperature models for greenhouse environmental control, Control Eng. Pract. 5 (1997) 1473–1481.
[22] N.A. Sigrimis, K.G. Arvanitis, R.S. Gates, A learning technique for a general purpose optimiser, Comput. Electron. Agric. 26 (2000) 83–103.
[23] G.D. Pasgianos, K.G. Arvanitis, P. Polycarpou, N. Sigrimis, A nonlinear feedback technique for greenhouse environmental control, Comput. Electron. Agric. 40 (2003) 153–177.
[24] L.D. Albright, Environment Control for Animals and Plants, ASAE Publishers, Saint Joseph, MI, 1990.
[25] L.D. Albright, A.-J. Both, A.J. Chiu, Controlling greenhouse light to a consistent daily integral, Trans. ASAE 43 (2000) 421–431.
[26] K.P. Ferentinos, L.D. Albright, D.V. Ramani, Optimal light integral and carbon dioxide concentration combinations for lettuce in ventilated greenhouses, J. Agric. Eng. Res. 77 (3) (2000) 309–315.
[27] P.M. Ferreira, E.A. Faria, A.E. Ruano, Real-time data acquisition system for the identification of dynamic temperature models in a hydroponic greenhouse, Acta Hort. 519 (2000) 191–196.
[28] A.J. Udink Ten Cate, Modeling and simulation in greenhouse climate control, Acta Hort. 174 (1985) 461–467.
[29] T. Boulard, A. Baille, A simple greenhouse climate control model incorporating effects on ventilation and evaporative cooling, Agric. Forest Meteorol. 65 (1993) 145–157.
[30] G.P.A. Bot, Greenhouse climate: from physical processes to a dynamic model, Ph.D. Thesis, Agricultural University, Wagemingen, Nederlands, 1983.
[31] I. Seginer, T. Boulard, B.J. Bailey, Neural network models of the greenhouse climate, J. Agric. Eng. Res. 59 (1994) 203–216.
[32] J. Boaventura Cunha, Greenhouse climate models: an overview, in: EFITA 2003 Conference, Debrecen, Hungary, July 5–9, (2003), pp. 823–829.
[33] R. Linker, P.O. Gutman, I. Seginer, Robust controllers for simultaneous control of temperature and CO_2 concentration in greenhouses, Control Eng. Pract. 7 (7) (1999) 851–862.
[34] R. Linker, I. Seginer, Greenhouse temperature modelling: a comparison between sigmoid neural networks and hybrid models, Mathematics Comput. Simul. 65 (2004) 19–29.
[35] N. Sigrimis, N. Rerras, A linear model for greenhouse control, Trans. ASAE 39 (1) (1996) 253–261.
[36] A. Trabelsi, M. Chaabane, M. Kamoun, Estimation structurelle d'une serre agricole par les réseaux de neurones, Actes des journées en Génie Electrique et Informatique (GEI), Hammamet, Tunisia, 2001.
[37] J.C. Gomez, A. Jutan, E. Baeyens, Wiener model identification and predictive control of a pH neutralisation process, IEE Proc. Control Theory Appl. 151 (3) (2004) 329–338.

5.3 Multi-modèle neuro-flou

A multi-structure modeling methodology

Nathalie PESSEL, Jean DUPLAIX, Jean-François BALMAT, Frédéric LAFONT

Laboratoire des Sciences de l'Information et des Systèmes (LSIS UMR CNRS 6168 Marseille) USTV BP20132 83957 La Garde Cedex France

Corresponding author : duplaix@univ-tln.fr

1 Introduction

1.1 Context

The aim of this chapter is to present an approach of modeling in the field of the systems with multiple inputs and multiple outputs (MIMO), non-linear, non-stationary and strongly disturbed. . For this class of system, it is very difficult to find a general model. We fix ourselves like objective to obtain a model of behavior to be able, for example, to compare different control laws of this type of system or to detect the sensor faults

The first idea is to define a model for each sample time. However an adaptive model depends on the inputs and its permanent evolution does not allow to compare several control laws.

Another solution is to use more global approaches like neuron non-adaptive networks for example. The model obtained depends then on the training carried out. That requires data files which cover all the possible situations of behavior.

For this class of system, another manner consists to develop an approach multi-models. There are many ways of determining sub-models, we directed ourselves towards the methods of soft-computing and present our methodology to the following paragraph.

1.2 Presentation of the modeling methodology

The diagram of synthesis of our method is represented in figure 1. A fundamental point is to define a strategy of obtaining sub-models. We supposed to have an expert knowledge of the system to be able to describe in time its various structures.

A principal component analysis (PCA) based on the recording data files makes it possible to minimize the number of measured inputs and define the structure of sub-models.

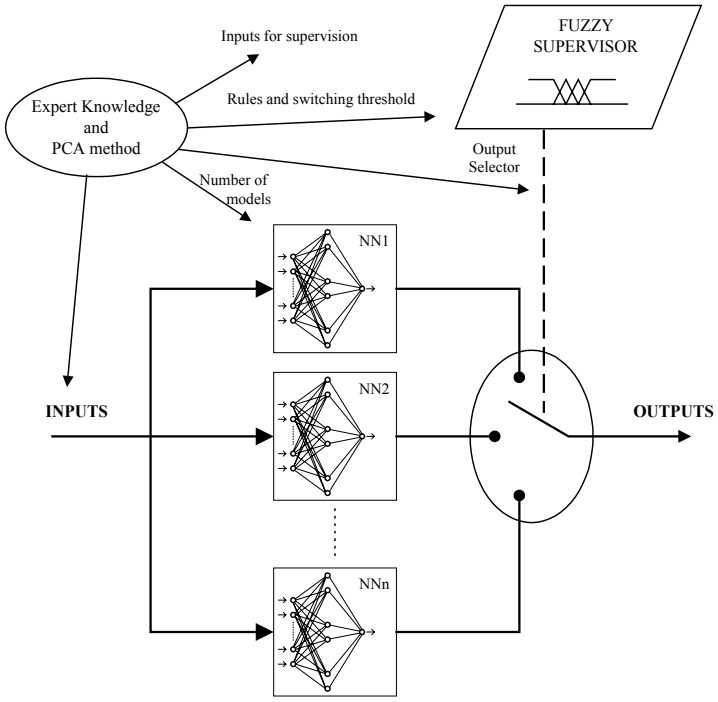


Fig. 1. Synthesis diagram of our method

These results are used to carry out a neural network (NN) modeling of the system.

We must build a block which makes it possible to manage these sub-models and to obtain its outputs. We have choice to have a multi-model system supervised by a Hierarchical Fuzzy Logic (HFL).

2 Principal Component Analysis

The principal component analysis (PCA) is a statistical method, which is included in the more general context of the factorial analysis.This method is used to reduce the number of variables of the considered system.

The PCA transforms a set of variables into a set of uncorrelated variables that represents most of the information in the original set of variables [1].

After the presentation of the PCA principle, we present the two principal steps of the PCA: the choice of the number of principal component and the clustering variable analysis.

2.1 Principle

When there are correlation between the *m* descriptive variables of a data distribution, the *m* dimensions of the data space exceeds the *l* number of characteristic variables necessary to describe these data. The higher the correlation between data descriptive variables, the smaller the number of useful characteristic variables for their representation.

Let a set of the initial data represented by the matrix X_{data}. The size of this matrix is $q \times p$ (q samples and p variables).

The variables can have different scales and units. However, we wish each variable to have the same weight in the system analysis. The data of each variable are centred and reduced. The principal component analysis is said normed. A new matrix of data X_{norm} is defined by:

$$X_{norm}(i) = \frac{X_i - \overline{X}_i}{S_i} \tag{1}$$

in which X_i is the i^{th} row vector of the matrix X, $\overline{X}_i$ is the mean of this vector with $\forall i \in [1, p]$ and S_i is the standard deviation of the considered variable i.

The identification of the PCA model parameters (i.e. principal components, PCs) is achieved by the estimation of the eigenvalues $\lambda_1, \cdots, \lambda_p$ and eigenvectors $u_1, \cdots, u_p$ of the correlation matrix R_x where $R_x = \frac{1}{q-1} X_{norm}^T X_{norm} = \sum_{i=1}^{p} \lambda_i u_i u_i^T$.

The principal components Y_i are a new set of data estimated by:

$$Y_i = X_{norm} u_i \tag{2}$$

in which u_i is the i^{th} eigenvectors of the correlation matrix R_x.

The k^{th} principal component is the vector which the components are the coordinates of the points on the k^{th} principal axis.

The first choice in the PCA is to determine how many PCs should be used to model the data.

2.2 Choice of the number of PC

Many different approaches have been suggested to carry out the choice of the number of PC [2, 3]. There are two groups of approaches: first based on the eigenvalues (Percentage of explained variance for each variable and Mean of eigenvalues) and second based on PCA model (Cross-validation criterion and Number of PC for a best reconstruction). A concise presentation of these methods is presented here.

2.2.1 Percentage of explained variance for each variable

The percentage of variability of each PC may be explained by:

$$W_{PC}(i) = \frac{\lambda_i}{\sum \lambda} \times 100 \tag{3}$$

The inertia explained by each PC decreases in function of the number of PC. Two methods are used to select the number of PC from the percentage of variability of each PC.

The first method is based on a threshold corresponding to the minimum percentage of inertia to restitute. The number of PC selected is equal to the number of PC necessary to achieve this threshold, generally fixed between 80% and 90%.

The second method is to represent graphically the percentages of variability of each PC in function of the number of the axis. The PCs which the order number is located before the bend of the plot are selected.

2.2.2 Mean of eigenvalues

This method consists to keep only the PC for which the eigenvalue is upper than the arithmetic mean of all the eigenvalues. When the data are centered and reduced, a PC is considered only if the variance is superior than 1.

2.2.3 Cross-validation criterion

This method is based on a PCA models. Several PCA models are computed by varying the number of PC. The squared differences between predicted and observed values are summed to form the Predictive Residual Sum of Squares (PRESS), which is a measure of the predictive power of the tested model. PRESS is computed as:

$$PRESS(\ell) = \sum \left(X_i - \hat{X}_i^{\ell} \right)^2 \tag{4}$$

where $\hat{X}_i^{\ell}$ is the vector of the i^{th} variable estimated by the PCA model determined by the first ℓ PC. The number of PC will correspond to ℓ for which the minimum $PRESS(\ell)$ value appeared.

2.2.4 Number of PC for a best reconstruction

The principle of the reconstruction consists in estimating a variable of vector X, by using measurements of the other variables and the PCA model defined by the first ℓ PC. This estimated variable will be called $\hat{X}_i^{\ell}$. The number of PC is determined by minimising the variance of the reconstruction error for variable set. We search ℓ such as:

$$J(\ell) = \sum_{k=1}^{m} \mathrm{var}\left(X_i - \hat{X}_i^k \right) \tag{5}$$

with m is the number of descriptive variables lets be minimum.

The choice of the number of significant PCs allows to study the correlation of the initial variables which is deduced from the correlations of principal components with each initial variable.

2.3 Clustering variable analysis

The study of the correlation of the initial variables can be achieved by analyzing the variable clusters.

The correlation of PC with each initial variable is explained by coefficients of correlation. These coefficients C_i are obtained by the multiplication of each eigenvector by the square root of the eigenvalue associated. They are between -1 and 1.

$$C_i = u_i \sqrt{\lambda_i} \tag{6}$$

The coefficients of correlation are used for variable graphic representations on the correlation circle formed by two principal components. This circle is called the "factorial space". Two variables are correlated if the projections of their vector are both close to the circle and themselves. The angle between two variables projected on the correlation circle is equal to the coefficient of correlation between these variables. This angle is measured by its cosine:

$$\cos\alpha = angle\left(C_i, C_j\right) \tag{7}$$

with i and $j \in [1, p]$. Thus, if the projection of two variables are both close to themselves (α little different of $2k\pi$), so the variables X_i and X_j are correlated. By opposition, if α is equal to 90°, the variables X_i and X_j are not correlated.

In this study, the modeling is achieved with multilayer neural networks.

3 Neural networks for modeling

In this section, we discuss of the modeling problems and the interest to use a neuronal approach. First, we explain the choice of the neural networks in the modeling techniques. Then, we present the different types of representation model which can be constructed according to the knowledge of the system. Finally, we tackle the conception of the neuronal structures through the PCA analysis.

3.1 Why a neuronal model ?

In a first stage, to design a model, several criteria must be defined. These criteria allow to choose the type of the model, and they depend of the type of the system. Thus, different elements must be considered such as the class of the model (simulation or prediction, static or dynamic), the knowledge of the different phe-

nomena (physical, biological,…) which occur in the process, the order of the system, the type and the effects of the disturbances, the different sensors and actuators of process, and so on. In this way, with these different elements, it is possible to define an appropriated model structure.

Neural networks are parsimonious universal approximations and they can approximate any non-linear function with the required accuracy using a smaller number of parameters. They can be used effectively for identification of non-linear static and dynamic complex systems [4]. Thus, the neural networks are known to possess characteristics that make them applicable to problems of large dimension [5, 6], where model isn't perfectly known and can't be created exclusively from physical knowledge. On the other hand, one of the important feature is their ability to learn by the example (a set of input/output data). For these reasons, using models based on neural networks is often interesting.

The quality of the neuronal model depends of the different neural network characteristics : choice of the learning method (supervised or unsupervised, training algorithm) and choice of the structure (topology, number of neurons, feedback loops,…).

3.2 "Black-Box" and "Grey-Box" models

As we have explained in the previous section, different types of models are distinguished according to the amount of physical prior knowledge considered. Thus, when a physical model is known, it is possible to design a knowledge model (*"White-Box"* model) which is the perfect representation of the system. This type of model is defined, thanks to an analyze of the phenomena brought into the process. Generally, it is interesting to design a knowledge model but according to the system complexity it is very difficult to obtain it. In this case, it is necessary to use the empirical knowledge by using the experimental measurements.

So, system identification is usually accomplished not only by processing measured data, but also by applying an expert knowledge. Like this, the two types of models are : the *"Grey-Box*" model or the *"Black-Box"* model.

The *"Black-box"* model is designed only with the experimental data while *"Grey-Box"* model takes into account the theoretical equations and the empirical knowledge of the process.

The *"Black-Box"* model must characterize the relations between the inputs and the outputs, thanks to the input/output measures, and by using a learning algorithm. Neural networks are associated, generally, with this type of model and they are often efficient for modeling the complex processes.

3.3 PCA and structure of neuronal model

The main goal of this part is to establish several models based on neural networks. The design of the models includes an important stage of selection and analyze of the set of the variables. Indeed, determining the relevance of the inputs

is of great importance in practical modeling problems [7, 8]. This first study allows to choose the best structure of the model. So, the right choice of the architecture is crucial for the application of neural nets in process identification.

Each neural network is used to identify and to model the process during a specific operating range. We choose a set of training data which are statistically significant and representative of the system during the considered range. Thus, we obtain several neuronal models with different structures, which depend of the number of inputs. The input pattern structure has therefore been fixed in accordance with the experimental knowledge of the system and by using PCA method. The purpose is to design several local neuronal models in reducing the number of inputs and, so, simplifying the structure of each neural network to improve the apprenticeship quality and, like this the quality of the model.

Finally, we obtain different networks (static or dynamics) in accordance with the choice of the inputs. The static networks realise a non linear algebraic function of the inputs while the dynamic networks is governed by a recurrent equation. Figure 2. shows, for example, a form of neuronal structure for the identification of the process. In this case, during the training stage the input vector consists of the outputs and the inputs of the process.

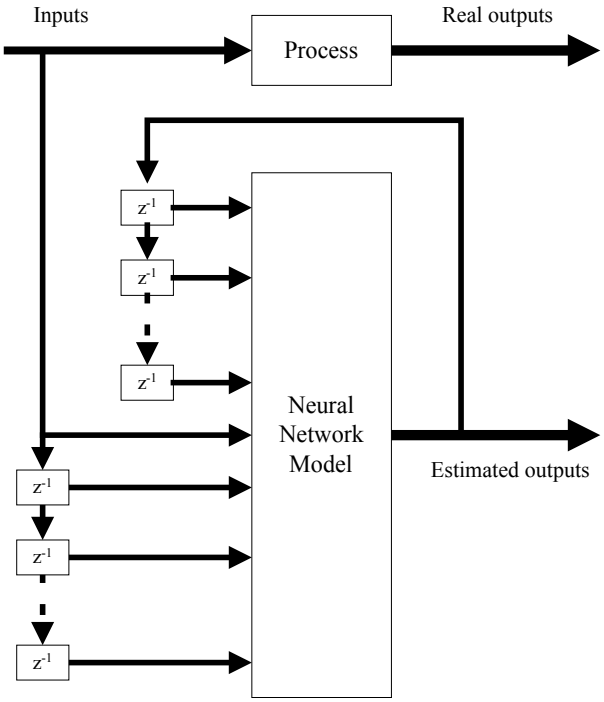


Fig. 2. Recurrent Neural Networks for identification (after training)

After training (**Fig. 2.**), the outputs of the neuronal model are fed back to its inputs through time delay units.

4 Supervision of a multi-model structure by a Hierarchical Fuzzy Logic

The aim is to build a fuzzy supervisor that selects one sub-model. The supervision manages the multi-model system. The knowledge of system enables us to choose certain data as inputs of the supervisor. The rules of a classical fuzzy supervisor (**Fig. 3.**) are:

$$If\ (E_1\ is\ F_1^1\ and\ E_2\ is\ F_2^1 \cdots and\ E_r\ is\ F_r^1)\ then\ (the\ output\ S\ is\ C_1) \tag{8}$$

E_r : r^{th} input of the fuzzy supervisor

F_r^1 : membership functions of the r^{th} input for the first rule

S : output of the fuzzy supervisor

C_1 : membership function of the output for the first rule.

The disadvantage of the classical fuzzy supervisor is the number of fuzzy rules grows exponentially with the number of input variables *r*.

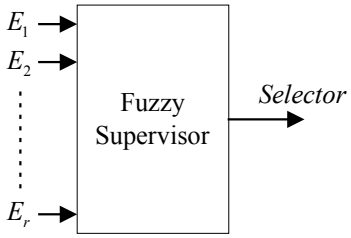


Fig. 3. Fuzzy supervisor

To overcome the problem, the idea of using hierarchical structure in designing a fuzzy system has been reported by Raju et al. [9, 10], Joo et al. [11], Lee et al. [12]. Thanks to this particular structure, the multivariable fuzzy supervisor can be decomposed into a collection of low-dimensional fuzzy logic units (FLUs). The set defines the hierarchical fuzzy system (HFS).

4.1 Hierarchical fuzzy logic system

Raju et al. proposes to treat, on a hierarchical basis, the rule base to reduce the number of rules. The aim is this number increases linearly and either exponentially with the number of variables. The set of rules is built in a hierarchical way: the variables in input of the fuzzy supervisor aren't treated in parallel but are distributed according to different levels of reasoning. The problem of selection is solved sequentially.

The rules of the first level are:

$$If\ (E_1\ is\ F_1^1\ and\ E_2\ is\ F_2^1 \cdots and\ E_n\ is\ F_n^1)\ then\ (the\ output\ S_1\ is\ C_1) \tag{9}$$

The rules of the level $i\,(i>1)$ are:

$$\begin{aligned} &If\ (E_{N_i+1}\ is\ F_1^1\ and\ \cdots and\ E_{N_i+n_i}\ is\ F_{n_i}^1\ and\ S_{i-1}\ is\ G_{i-1}) \\ &\qquad then\ (the\ output\ S_i\ is\ C_i) \end{aligned} \tag{10}$$

where $N_i=\sum_{j=1}^{i-1} n_j \le r$, with n_j the number of variables at the level *i*, and S_{i-1} the output of the precedent level $(i-1)$ considered as an input variable for the level *i* and G_{i-1} the membership function corresponding to the input S_{i-1} for the i^{th} rule.

At each level $i\,(i>1)$, one or more variables are added at the output of the precedent level to develop the set of the rules of the level *i*.

Thanks to the hierarchical structure, the rules number is a linear function of the variables number *r*.

Furthermore, Raju et al. show that the total number of rules is minimal if, in each successive levels, one and only one additional variable are considered.

The hierarchical grading remains nevertheless a difficult stage and requires an expert knowledge or a qualitative analysis of the process. There is not any general rule to organize the variables. However, it seems more natural to place the significant variables in bottom of the hierarchy because the last FLU constitutes the most direct chain. Indeed, the last variables influence more the final decision. The hierarchical structure, composed of the FLUs at two inputs, is that which permits a maximal reduction of the rules number (**Fig. 4.**).

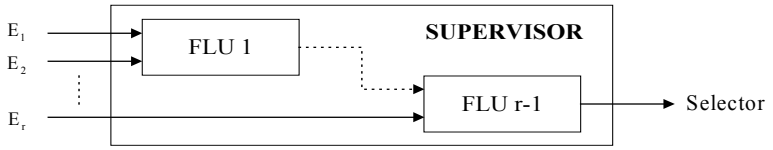


Fig. 4. Hierarchical fuzzy system

4.2 Selection

The estimate output of the process can be obtained in two manners: the switching or the fusion of the models. The output of the HFS is calculated by a defuzzyfication method. The defuzzyfication consists to transform the fuzzy set, given by the aggregation, in a precise value of selection. There are several methods such as the height, the centre of gravity, the centers of the surfaces and of the maximum,…

4.2.1 Switching

The switching relates to the selection of only one model at the same time.

In this case, the used method for the defuzzyfication is the method of the height (**Fig. 5.**) with $Selector = \arg \left\{ \max \mu_S (y) \right\}$

where μ_S is the degree of membership function and y the range of the output selector. This method of defuzzyfication is applied only at the last FLU. The others FLUs have, as defuzzyfication method, the centre of gravity.

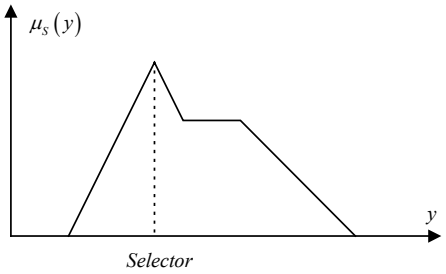


Fig. 5. Defuzzyfication: method of height

This method consists to choose the value of the maximum as the output variable.

If the membership functions of the output are the singletons then the fuzzy supervisor selects one model.

4.2.2 Fusion

In this category, several models can be concerned to elaborate the estimated output of the process.

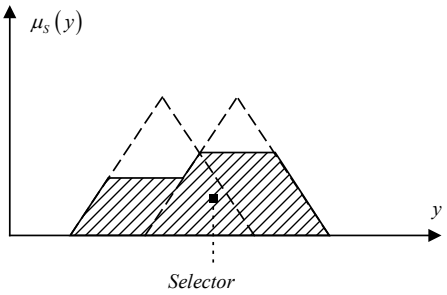


Fig. 6. Defuzzification: method of the centre of gravity

The centre of gravity, which is the most used method, provides intuitively the most significant value of the fuzzy set resulting from the aggregation (**Fig. 6.**)

with $Selector = \dfrac{\int_y \mu_S (y)\, y dy}{\int_y \mu_S (y)\, dy}$.

5 Application and results

The applicability is the management of the microclimate of an experimental greenhouse, first level of the control of a plant industrial production.

We defined the principal inputs (actuators and disturbances) and the outputs to be regulated. Our work is based on a great data base of measurable variables recorded during days, months, seasons and years.

In a first study, we used as Cunha [13] parametric methods of identification (recursive less square methods and its variants): the idea being of going towards adaptive control. One of the difficulties met is that these models contain the dynamics of adaptation: it is thus not easy to build control laws. For this reason, we developed methods of global solution by using neural networks, fuzzy logic or genetic algorithms [14]. This approach enabled us to note the difficulties to have a global model.

We direct ourselves now towards the search for multi-models [15, 16]. Our physical knowledge of the system must be coupled with a principal component analysis (PCA) to allow a sorting of the data in order to isolate a reduced number of models [17].

5.1 Description of the system

The essential goal of a greenhouse is to install a shelter to improve the conditions weather: this system must be open to make it possible to exploit the advantages of the external disturbances as well as possible (radiation, temperature…) but also to filter the disadvantages (wind, rain…).

This step allows us by controlling the internal temperature, the internal hygrometry and the carbon dioxide to create optimal conditions for the plant on the level of photosynthesis.

To manage the root part of the plant, industrial resource of action consists in carrying out a culture out-ground in order to optimise, through the nutritive solution, the development of the plant.

The problem is thus complex on the level of modeling and definition of control laws. We quote some recent references and limited to the management of the microclimate: for modeling, J B.Cunha and Salgado [13, 18], F Lafont [19] and for control K.G. Arvantis [16], F Lafont [20], N Bennis [21], JF. Balmat [15].

Our system comprises eight inputs and two outputs:

- 4 actuators (heating Ch (boolean), opening Ov (%), shade Om (%), misting system Br (boolean));
- 4 meteorology disturbances (external temperature Te (°C), external hygrometry He (%), radiation Rg (W/m^2), speed of the wind Vv (km/h));
- 2 controlled outputs (internal temperature Ti (°C), internal hygrometry Hi (%)).

The figure 7 presents the model of microclimate:

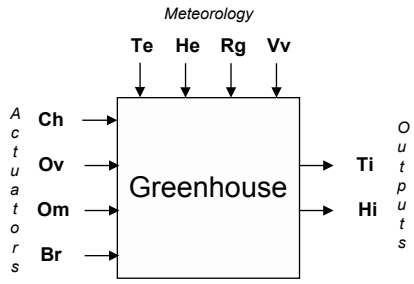


Fig. 7. The greenhouse model

In this developpement, the controlled output is only the internal temperature Ti.

5.2 Application of the methodology over one day of training

Manipulations were carried out using experimental greenhouse data of a daytime in March (**Fig. 8., Fig. 9.**). This month is very interesting because it presents a lot of external disturbances (rain, wind in two different directions). These variations assign automatically the internal climate.

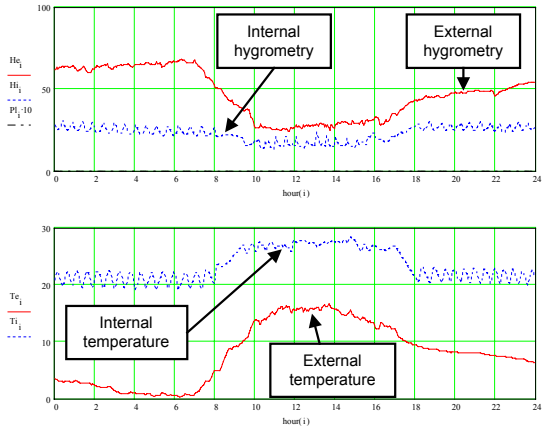


Fig. 8. Hygrometry and temperature of daytime March 10th

The sample time is about three minutes. So the size of the file is 10 × 453 (10 variables and 453 samples). The 10 variables are 6 sensors and 4 actuators. The 4 actuators are "actionable" independently in function of the part of daytime. They will not consider during the study of correlation of variables (but by the expert analysis).

Previously, we have seen that the parts of the daytime condition the command's activity. The expert knowledge of the experimental greenhouse allowed defining three parts of the daytime according to the global radiation.

The first part defined is the night. The global radiation must be inferior to 7 W/m^2. This threshold is fixed at this value because the radiation in the course of the night can be 6 watts in the presence of the moon. For example, for the studied daytime, the night file is composed of 233 samples.

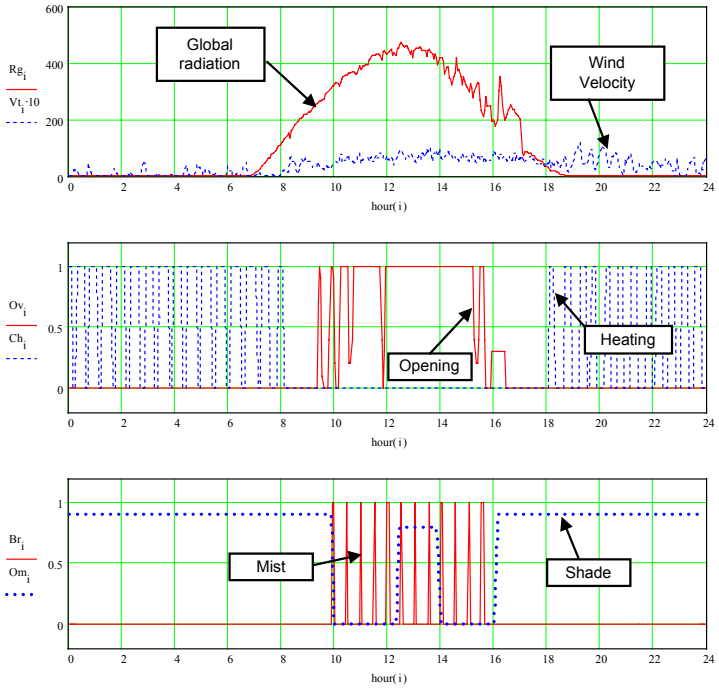


Fig. 9. Disturbances and actuators of daytime March 10th

The second part defined is the sunrise (daybreak). This is an intermediary part, very short but which presents a typical variation of sensors. Through the expert knowledge, we know that there is interesting range between the night and the day parts. This part begins when the global radiation is higher than 7 W/m^2. Its end is more difficult to define. The PCA allowed us to detect a correlated variable change as soon as the radiation is higher than 170 W/m^2. So the daybreak is represented by only 29 samples.

The third part is the day. This part begins after the sunrise and continues until the radiation is higher than 7 watts. For the considered day, the day file is composed of 191 samples.

The study of the daytime parts gives us three files. We have independently applied a PCA on the 6 variables for each file. The PCA is achieved on the identical way. So the following results present the experimentation relating to the day file. The correlation between variables is defined by projecting the variables in the correlation circle on the plane of the must representative principal components.

The first result given by the PCA allows determining the number of principal components necessary to represent the system. This choice depends of the percentage of variability of each PC. For the studied day, 90 % of the variability is explained by the first two principal components. The first one explains 77 % and the second 13%. Moreover the plot of the variability percentage of each PC (**Fig. 10.**) presents a bend which allows to select the two first PC. The principal components selected create a new space of visualization that allows evaluating correlation between variables. Figure 11 shows the projections of the variables in the correlation circle on the plane PC1-PC2 for the day part.

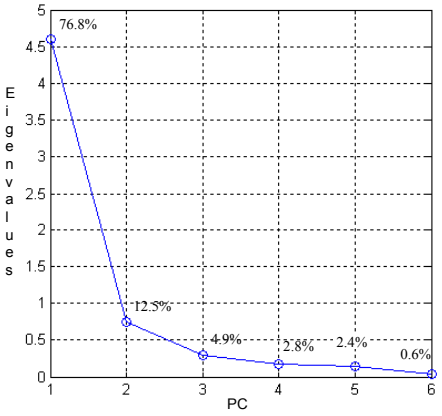
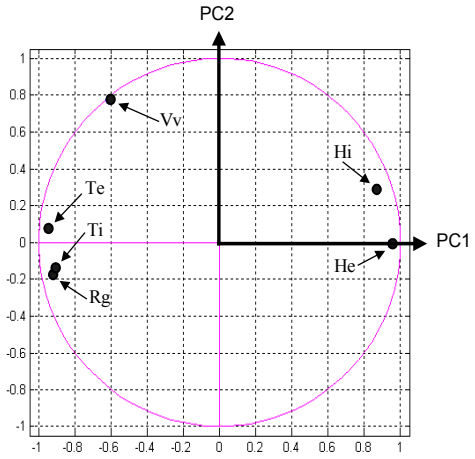


Fig. 10. Percentage of variability of each PC



	Ti	*Te*	*Hi*	*He*	*Vv*	*Rg*
Ti	–	4.8°	175.0°	157.1°	47.6°	6.4°
Te		–	170.3°	152.3°	42.8°	1.7°
Hi			–	17.9°	127.4°	168.6°
He				–	109.5°	150.6°
Vv					–	41.2°
Rg						–

Table of correlation coefficients

Application of equation (7)

Fig. 11. Correlation circle PC1-PC2 for day part

Two groups of variables can be identified. These groups allow defining the correlated variables. Indeed, two variables are correlated if they are close both to

themselves and to the circle, and if their angle in the center is small (**Fig. 12.**). We observe that Te, Ti and Rg are three correlated variables. In the same way Hi and He are correlated. The same study is achieved by using data of night and sunrise parts. The correlation circles on planes PC1-PC2 for these parts allow concluding that for the daybreak, Te and Rg are correlated. The night is a particular part where no correlation is detected.

The PCA allows to detect correlation between variables and thereby to reduce the number of inputs. Indeed, when two or more variables are correlated only one will be selected as input of the model. This choice is very important. It is based on the expert knowledge and on the possibilities to simplify the model structure. In this way, for our experimental conditions, for the day Te will be selected to represented the first group {Te, Ti, Rg} and He for the second {He, Hi}, and for the daybreak Te will be conserved from the group {Te, Rg}.

These results allow to define the new inputs of the three models. The following figure presents the structure of the global model and all the sub-models (**Fig. 12.**).

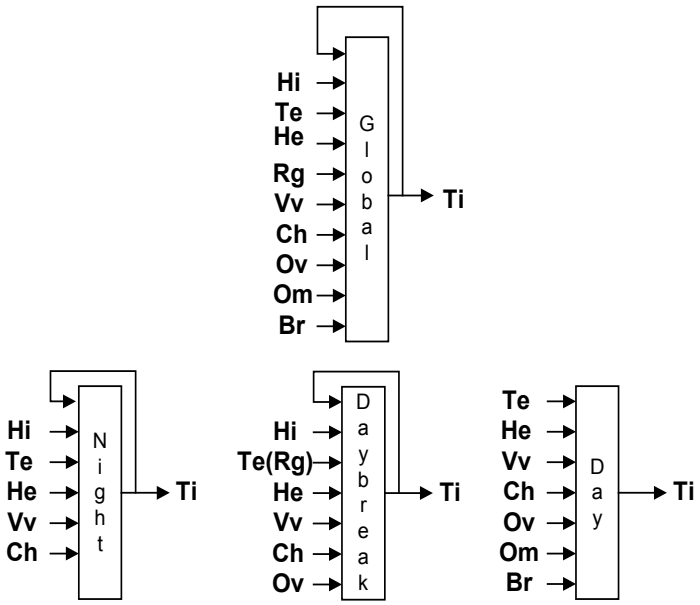


Fig. 12. Global model and sub-models for a daytime

For our system, model isn't perfectly known and can't be created exclusively from physical knowledge. However, we have a lot of files with input/output data pairs. According to the previous results, we can establish three neuronal sub-models with different structures which depend of the choice of the inputs.

Like this, we obtain the neuronal models with the reduction of the input variables number. The reduction of the dimension of the state inputs allows to transform and simplify the model structure.

In all case, we use typical multilayer networks with one input layer, one hidden layer and one output layer (**Fig. 13.**). The activation function is a sigmoid for the hidden cells and a linear function for the output cell. The training is based on the back propagation algorithm and the inputs (X_i) of the neural networks are normalized between 0 and 1. For each structure model, we keep the same number of neurons in hidden layer.

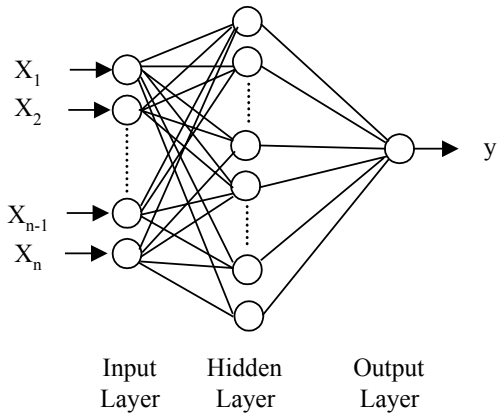


Fig. 13. Feed-forward neural network architecture

The three neuronal sub-models correspond to different types of neural networks. Indeed, we obtain dynamic networks for "Night" and "Daybreak" sub-models, and a static network for "Day" sub-model. The three local sub-models simplified contain less parameters than the global model (**Table 1.**)

Model	Global	Night	Daybreak	Day
Number of inputs	9	6	7	7
Number of neurons in hidden layer	8	5	5	5
Number of parameters (weights and bias)	89	41	46	46

Table 1. Number of parameters for each model

In figure 14, we depict, the internal temperature curves (real and estimated) for the learning stage (by using the March 10th) .

These results will be compared with the results obtained by using a multi-model approach.

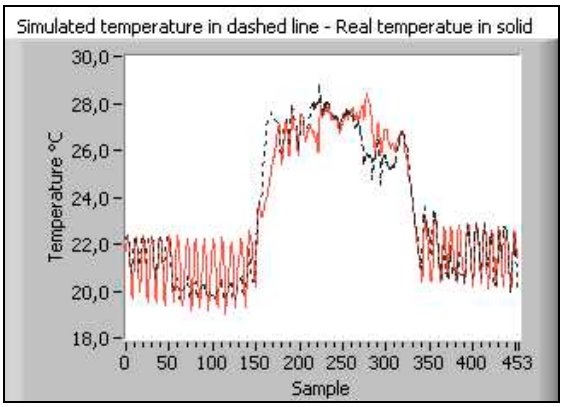


Fig. 14. Global model (March 10th) – Apprenticeship data

In Table 2, we compute the Mean Error (ME) and the Variance Error (VE) for each model in the learning stage.

Model	Global	Night	Daybreak	Day
Mean error	0.72	0.73	0.40	0.36
Variance error	0.49	0.31	0.36	0.10

Table 2. Mean and Variance Errors for each model (March 10th)

We obtain better results for the local models daybreak and day when the night model error is similar to global model error.

We have three neuronal models that represent three operating ranges (Night, Daybreak, and Day). Thus, for our application, the used HFS is composed by two FLUs (**Fig. 15.**).

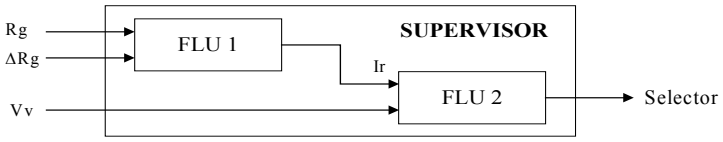


Fig. 15. Hierarchical fuzzy system

The two systems are of Mamdani type with the max-min method for the fuzzy inference. An expert of the greenhouse system defines the rules bases.

The first FLU has two inputs Rg and ΔRg and one output Ir. The radiation Rg has been defined with three membership functions: one for the null radiation, one for the means radiation and one for the strong radiation. ΔRg is a mean value of the Rg variation, on several samples (five samples is equal to fifty minutes), which represents the tendency and allows to eliminate the problems which could bring the furtive clouds (considered as noise). This input has been defined with three

membership functions: one for the negative variation, one for the constant variation and one for the positive variation. Thus, this FLU has nine rules. The output is an index of the global radiation (Ir) with three membership functions: weak, small and strong.

The second FLU has two inputs Ir and Vv. The wind speed Vv has been defined with two membership functions: one for weak wind and one for strong wind. Thus, this FLU has six rules. The membership functions of the output are the singleton. The output selector determines the model to choose (Night, Daybreak, and Day).

The FLU1 has the centre of gravity for the deffuzzyfication method. For the second FLU, the deffuzzyfication method changes according to the choice of selection. For the switching, the method of the height is used as we have seen in the section 4.2.1.

For the fusion, the method is the centre of gravity as in section 4.2.2:

$$Ti = \frac{\mu_1 . Ti_1 + \mu_2 . Ti_2 + \mu_3 . Ti_3}{\mu_1 + \mu_2 + \mu_3} \tag{11}$$

For this application, the fusion doesn't make improvements because the different models have been well defined. The models represent all the operating ranges of the greenhouse system.

In figure 16, we present the results of the learned daytime. In the first window, we can see the evolution of the selector during the daytime of March 10th. The selector switches between the night model, the daybreak model and the day model.

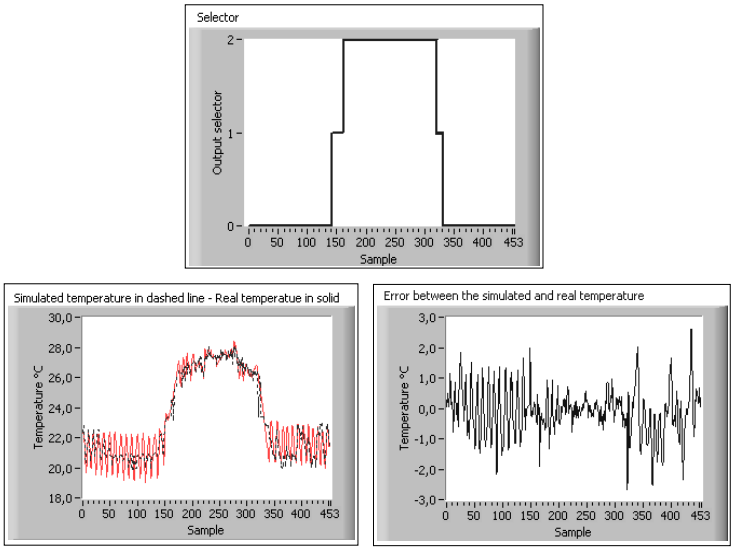


Fig. 16. Learned daytime March 10th selection by switching

The second window presents the evolution of the real temperature and the simulated temperature. Finally, the error between these two temperatures is shown in the third window.

With these results, we are now validating on other days: one day comparable on March 11th and a very different day on March 12th.

5.3 Validation over other days

We have the principal characteristics of the day selected in Table 3. For the outputs and disturbances, the mean, min and max are given. For the actuators, only the percentage of use over 24 hours is noted.

			March 10th	March 11th	March 12th
Outputs	Ti (°C)	Mean	23.4	23.7	21.7
		Min	19.0	19.2	19.9
		Max	28.4	27.7	23.3
	Hi (%)	Mean	22.7	21.2	37.0
		Min	13.3	11.3	24.1
		Max	30.5	30.5	59.9
Meteorology	Te (°C)	Mean	8.0	8.6	8.4
		Min	0.3	2.5	5.7
		Max	16.6	16.8	10.5
	He (%)	Mean	46.7	39.5	68.0
		Min	24.4	18.2	47.6
		Max	68.0	47.6	92.6
	Rg (W/m^2)	Mean	129.7	138.5	12.2
		Min	0	0	0
		Max	472.5	456.9	59.4
	Vv (km/h)	Mean	3.9	4.6	5.3
		Min	0	0	0
		Max	11.9	12.2	12.6
	Rain	(% of 24h)	0	0	15.9
Actuators	Ch	(% of 24h)	35.5	31.8	46.4
	Ov	(% of 24h)	23.1	24.4	0
	Om	(% of 24h)	72.1	73.1	66.6
	Br	(% of 24h)	3.5	3.9	3.3

Table 3. Principal characteristics for the three days selected

For the whole of the data, the day of March 11th is very close to that of March 10th (day of training: fine weather with few clouds and sun). The second validation day (March 12th) is very different: bad weather (overcast sky, no sun and rain).

The figures 17 and 18 depict the results with the global model and with a selection by switching over the two validation days.

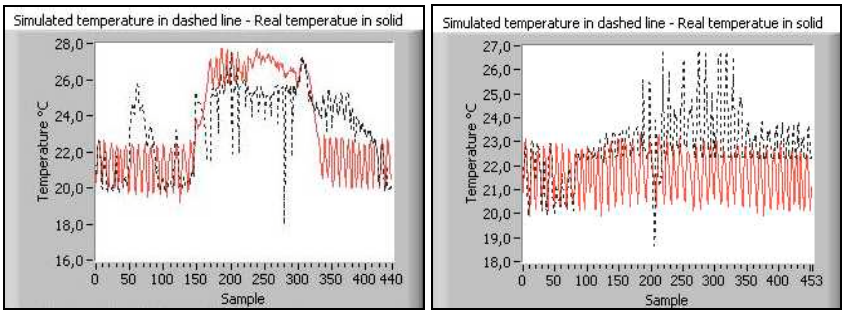


Fig. 17. Global model (March 11th and March 12th) – Validation data

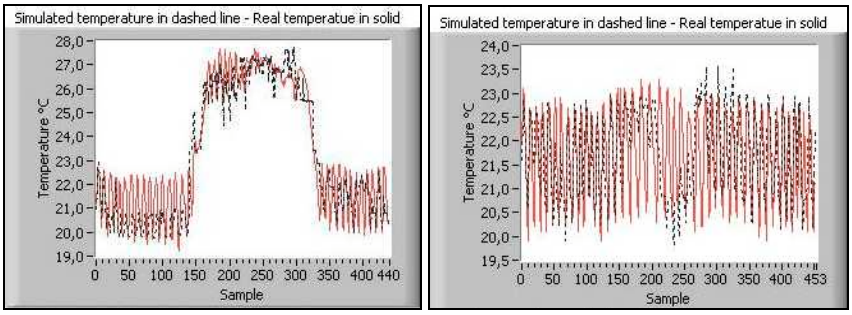


Fig. 18. Selection by switching (March 11th and March 12th) – Validation data

The simulation over the 3 days with the 3 approaches (global model, supervisions with selection by switching or fusion) leads us to the results (**Table 4.**).

		Global Model	Switching HFL	Fusion HFL
March 10th Learning	Mean	0.72	0.60	0.60
	Variance	0.49	0.29	0.29
March 11th Validation	Mean	1.58	0.74	0.74
	Variance	1.52	0.39	0.38
March 12th Validation	Mean	1.45	0.62	0.62
	Variance	1.44	0.34	0.34

Table 4. Mean and variance for the 3 days studied

For the global model, the results are much worse as well over the day similar to the day earned as for the day very different. That confirms that for such a class of system that it is difficult to have a total model.

The approach multi-model on the other hand led to results virtually identical that the day is similar or different.

We do not detect a difference between the selection by switching or fusion as we noted in the preceding section.

6 Conclusion

This chapter presents a modeling methodology of the complex systems. The system class considered put together multivariable, non-linear, non-stationary and strongly disturbed systems. These processes aren't easy to modelise and it's generally difficult to find a global model. Thus, we propose to modelise these systems with multi-model at variable structure. The methodology is based on a supervision of a multi-model structure defined by a statistical analysis and validated by the expert knowledge of the system.

The association of the expert knowledge and the data analysis allows to detect the correlation between variables and to select the most significant in each group of correlated variable. In our methodology, we use the Principal Component Analysis. The selected variables define the neuronal structures.

Neural Networks are efficient for the complex non-linear dynamic systems modeling. In this study, we present two types of neural networks which correspond to the dynamic networks and the static networks. These neuronal structures are obtained from data statistic analysis. So, we define several local models which allow to represent the working in large field. The sub-models are selected by a supervisor.

The management of the multi-model system is carried out by supervision with a Hierarchical Fuzzy Logic. The fuzzy allows processing some uncertain and/or imprecise information present in this system class. Moreover, a fuzzy supervisor can describe the behavior of the process by the switching or the fusion of models. The advantage of the selection by fusion is the possibility to combine some models when an operating range is not defined. However, the fuzzy approach presents a limit: if the inputs number of the supervisor is great, there is too difficult to implement it. The expert knowledge isn't able to tune the fuzzy selector.

The multi-structure modeling methodology presented refers to an expert knowledge of the system. This knowledge steps in the data analysis and more exactly in the definition of the operating ranges number and in the selection of the pertinent variable in a group of correlated variables. The supervision definition needs to this expert knowledge to describe the rule bases.

The several experiments presented in this chapter show the performances of this multi-structure modeling methodology.

7 References

1. Thalib L, Kitching RL, Bhatti MI (1999) Principal Component Analysis for grouped data – a case study. Environmetrics 10:565-574
2. Andrade JM, Gomez-Carracedo MP, Krzanowski W, Kubista M (2004) Procrustes rotation in analytical chemistry, a tutorial. Chemometrics and Intelligent Laboratory Systems 72:123-132
3. Emre Celebi M, Alp Aslandogan Y (2003) Content-based Image Retrieval Incorporating Models of Human Perception. Department of Computer Science and Engineering, University of Texas at Arlington, Technical Report CSE-2003-21
4. Norgaard M, Ravn O, Poulsen NK, Hansen LK (2000) Neural networks for modelling and control of dynamic systems, Springer-Verlag, London
5. Magali R, Meireles G, Paulo E, Almeida M, SimÄoes MG (2003) A comprehensive review for industrial applicability of artificial neural networks. IEEE Transactions on Industrial Electronics 50, no 3:585-601
6. Kumpati S, Narendra F, Parthasarathy K (1990) Identification and control of dynamical systems using neural networks. IEEE Transactions on neural networks1, no. 1:4-27
7. Koskela T (2003) Neural network methods in analyzing and modeling time varying processes - Dissertation for the degree of Doctor of Science in Technology - Department of Electrical and Communications Engineering, Helsinki University of Technology – Finland
8. Afrashteh R (2000) Modeling, fault detection and diagnosis of an automotive engine using artificial neural networks. Thesis submitted in partial fulfillment of the requirements for the degree of Master of Applied Science in the School of Engineering Science - Simon Fraser University, Canada
9. Raju GVS., Zhou J, Kisner R.A (1991) Hierarchical fuzzy control. Int. J. Control 54:1201-1216
10. Raju GVS., Zhou J (1993) Adaptive Hierarchical fuzzy controller. IEEE Trans. Systems, Man Cybernet. 23: 973-980
11. Joo MG, Lee JS (2002) Universal approximation by hierarchical fuzzy system with constraints on the fuzzy rule. Fuzzy Sets and Systems 130:175-188
12. Lee ML, Chung HG, Yu FM (2003) Modeling of hierarchical fuzzy systems. Fuzzy Sets and Systems 138:343-361
13. Cunha JB, Couto C, Ruano AE (1997) Real-time parameter estimation of dynamic temperature models for greenhouse environmental control. Control Eng. Practice 5, n° 10:1473-1481
14. Bouchouicha M, Lafont F, Balmat JF (2002) Neural networks, Fuzzy logic and Genetic algorithms for greenhouse identification. Second International Conference – Tunisian Conference on Electro-Technical and Automatic Control (JTEA'2002), Sousse, pp. 356-362
15. Balmat JF, Lafont F (2003) Multi-model architecture supervised by Kohonen map. Sciences of Electronic, Technology of Information and Telecom (SETIT'03), Sousse, pp 98–104
16. Arvantis KG, Paraskevopoulos PN, Vernados AA (2000) Multirate adaptative temperature control of greenhouses. Comput. Electron Agric. 26:303-320
17. Pessel N, Balmat JF (2005) Principal Component Analysis to the Modeling of Systems – Application to an Experimental Greenhouse.The 3rd IEEE International Conference

on Systems, Signals & Devices, SSD'2005, Sousse, Cd-rom, Systems Analysis Automatic Control
18. Salgado P, Cunha JB (2005) Greenhouse climate hierarchical fuzzy modelling. Control Eng. Practice 13, n°5:613-628
19. Lafont F, Balmat JF, Taurines M (2005) Fuzzy forgetting factor for system identification. SSD'2005, Sousse, Cd-rom, Vol. Systems Analysis Automatic Control
20. Lafont F, Balmat JF (2002) Optimized fuzzy control of a greenhouse. Fuzzy Sets and systems 128, n°1:47-59
21. Bennis N, Duplaix J, Enéa G, Haloua M, Youlal H (2005) An advanced control of greenhouse climate. 33rd International Symposium"Actual Tasks on Agricultural Engineering", Opatija, Croatia, pp. 265-277

5.4 Aide à la décision : statique

Ocean Engineering 36 (2009) 1278–1286

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

MAritime RISk Assessment (MARISA), a fuzzy approach to define an individual ship risk factor

Jean-François Balmat [a,*], Frédéric Lafont [a], Robert Maifret [b], Nathalie Pessel [a]

[a] LSIS, UMR CNRS 6168, University of South-Toulon-Var, B.P. 20132, F-83957 La Garde Cedex, France
[b] DCNS, Division SIS, BP 403, F-83055 Toulon Cedex, France

ARTICLE INFO

Article history:
Received 14 November 2008
Accepted 15 July 2009
Available online 21 July 2009

Keywords:
Maritime risk assessment
Maritime safety
Fuzzy risk factor

ABSTRACT

This paper presents a fuzzy approach for the MAritime RISk Assessment (MARISA) applied to safety at sea. The aim of this work is to define automatically an individual ship risk factor which could be used in a decision making system. To achieve this purpose, a modular and hierarchical structure using fuzzy logic has been developed. It allows us to obtain a fuzzy risk factor (FRF) composed of a static risk factor (SRF) and a dynamic risk factor (DRF). The static risk factor assessment takes into account several static data relative to the ship (age, flag, gross tonnage, number of companies, duration of detention and type). The dynamic risk factor is evaluated by considering the meteorological conditions (sea state, wind speed and visibility) and the moment of the day. Moreover, the MARISA graphic interface developed with the Labview software is presented. This interface allows several simulations to be carried out to validate the fuzzy method proposed. Simulation results are presented.

Contents

* Corresponding author. Tel.: +33 0494142039.
E-mail address: balmat@univ-tln.fr (J.-F. Balmat).

0029-8018/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oceaneng.2009.07.003

1. Introduction

This paper presents a new approach of the maritime risk assessment for safety at sea based on a risk factor determined by a fuzzy expert system. The objective is to design a new and flexible decision tool to be fitted to existing vessel traffic monitoring and information systems (VTMIS) or naval communication and information system (CIS). More precisely, we propose an approach to evaluate the casualty risk.

The identification of risky ships is, nowadays, an important research theme. Like this, many studies have been realised to identify high risk ships (Degré, 2003; Glansdorp, 2004; Regelink et al., 2004; Van der Heijden et al., 2004; Wang et al., 2004; Sage, 2005; Haj-Salem et al., 2006).

Degré (2004) proposed a risk factor which is an individual index allowing the risk rating to be quantified for each ship. He defined an individual ship risk index for safety at sea (IRIS). Moreover, Degré and Benabbou (2005, 2004) determined the general expression of this index and showed that it was possible to assess risk in real time. So, automatic detection of high risk vessels can be realised and decision processes of authorities in charge of safety at sea be improved. In his study, Degré made a data analysis of maritime accidents listed by the International Maritime Organization (IMO) for several years. Thanks to this analysis, he proposed to take into account not only the static parameters (such as the ship's age, type, flag) but also the dynamic parameters (such as the meteorological conditions). We used this work to design our system.

The objective of this article is to present a new system for MAritime RISk Assessment (MARISA) and to give some simulation results to evaluate the proposed approach. First of all, we briefly describe the principles of fuzzy logic. Then, we present some studies dealing with fuzzy logic applied to safety at sea. In the fourth section, we propose the architecture of the MARISA system and we detail the different modules which constitute it. Next, we describe MARISA graphic interface and we give some results about the fuzzy risk factor which is computed for various kinds of ships and for various meteorological conditions. Finally, the conclusion summarises the main results and objectives of our approach.

The research perspectives are very attractive because this kind of system allows an individual ship's risk factor to be determined to generate visual alarms in a maritime surveillance zone, in real time. This is an essential issue for environment protection. Next step will be to apply these results to oil pollution prevention at sea. In this case, it will be necessary to consider other dynamic parameters such as, for example, the speed evolution of ships.

2. Background of the fuzzy logic approach

An expert system is a tool which imitates the cognitive mechanisms of a human expert. It is composed of three parts which are: the event base, the rule base and the inference engine. Most expert systems use a formal logic mechanism and deductive reasoning. The design of these systems is based on the information given by humans. Generally, this information is vague, ambiguous or imprecise. For this reason, it can be worthwhile to use fuzzy logic which was introduced by Zadeh (1965). The fuzzy logic approach is based on the definition of a set of logic rules derived from human knowledge and reasoning. A fuzzy system is able to compute a non-fuzzy result from several non-fuzzy variables but with a fuzzy reasoning process. The first stage is fuzzification which performs the transformation of data into fuzzy values. The input/output variables are linguistic variables which are defined by a name, a reference set (universe of discourse) and a set of fuzzy normalised subsets (membership functions).

The system is based on fuzzy reasoning. The inference is described by the knowledge base with a set of fuzzy rules. The reasoning consists in applying the fuzzy rules and aggregating the results. At last, the defuzzification module transforms the results into a single value. The diagram depicted in Fig. 1 sums up the fuzzy reasoning principle.

x is the input data and x_{res} is the output of the fuzzy system. A fuzzy system is defined by a set of values and it is characterized by a membership function $\mu(x)$ which quantifies the degree of membership.

3. Fuzzy for safety assessment in the maritime domain

There are many works about risk assessment by using fuzzy logic in several industrial fields. In this section, we present some studies dealing with the specific problem of safety assessment in the maritime domain.

Sii et al. (2001, 2004) studied about the interest of using fuzzy logic in safety problems. In these papers, the authors proposed a qualitative safety modelling for maritime systems. They presented an example of fire due to a fuel system failure in the engine room of an offshore support ship. They explained two concepts for maritime risk analysis using classical fuzzy logic (using the Mamdani's inference system) and adaptive fuzzy logic approaches. In offshore safety assessment, one of the main difficulties is to take into account incomplete and vague information. In this study, the authors have not considered the incomplete information to establish the rule bases. Thus, when the complexity of expert analysis increases, the knowledge representation

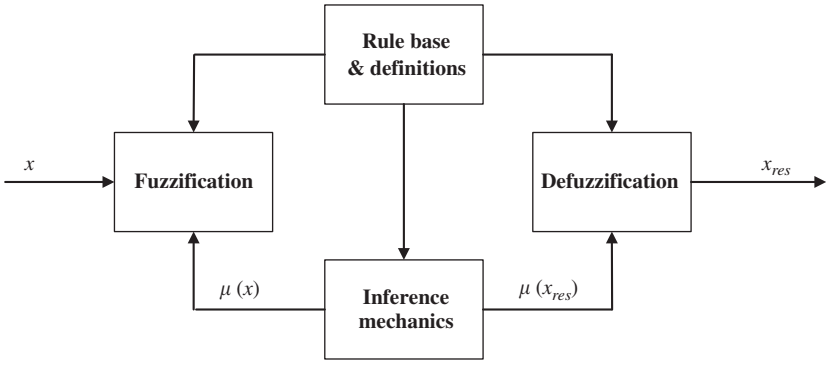


Fig. 1. Fuzzy reasoning principle.

power of the classical fuzzy rule base may be very limited (Yang et al., 2008).

Liu et al. (2005) studied a method to analyse safety and the procedures of safety assessment. They pointed out the difficulties of defining a mathematical model representing the safety of the maritime system. Thus, the authors used fuzzy logic to model a floating production storage offloading (FPSO) system. A study case of a collision risk between an FPSO and a shuttle tanker has been described. The authors showed the advantages of fuzzy logic to translate situations and knowledge. From the Dempster–Shafer theory of evidence, they proposed a hybrid safety model based on fuzzy logic and an evidential reasoning approach. The main difficulty is to deal with complex calculations in the fuzzy rule-based evidential reasoning (FRBER) approach. Moreover, it is not easy to obtain a good compromise between the "precise" risk results and the loss of information in the process of fuzzy inference (Yang et al., 2008).

Hu et al. (2007) proposed a model based on relative risk assessment (MRRA) with fuzzy functions. Their approach considered five factors including detailed information about accident characteristics in ship navigation. This study has been applied to the assessment of safe piloting in Shanghai harbour, China. Knowing the difficulty of identifying the major risk distribution in pilotage safety in Shanghai Harbour, the authors introduced the MRRA for analysing the risks involved in the ship's pilot. The conclusion is that the results are equivalent with those using a generic risk model (like formal safety assessment (FSA)) but the proposed method shows more detailed information.

Eleye-Datubo et al. (2008) explained an offshore safety assessment by incorporative risk modelling in a fuzzy-bayesian network. This kind of approach enables the modelling and the reasoning about uncertainty that can be due to a combination of inherent vagueness and randomness. In the maritime domain, it is very important to take the effects of human performance into account. Thus, the proposed study concerns a flexible risk modelling approach by combining the advantages of fuzzy logic and Bayesian networks. The potential of this approach to model safety knowledge/assessments/practices in the maritime industry has been described by the authors.

Recently, Yang et al. (2008) proposed a fuzzy rule-based Bayesian reasoning approach for prioritizing failures in failure mode and effects analysis (FMEA). The methodology includes five steps to develop the criticality analysis. This approach has been applied to the offshore engineering domain and more particularly to collision risks in maritime domain. The authors showed the ability of this method to deliver risk criticality values when data is subjective.

In these articles, the authors confirm the interest of using fuzzy logic techniques for risk assessment in the maritime domain. As the literature shows, according to the applications and to improve the results, it may be necessary to add a Bayesian approach. However, the fuzzy approach seems efficient to process the various kinds of data available in maritime safety to design a risk assessment system.

Thus, we propose to use a fuzzy approach to define an individual ship risk factor which could be used for a decision making system. In our study, the considered data is not subjective and the knowledge is obtained from the analysis of variables influencing the casualty rate (Degré, 2004). First, in the next section, we present the data that could be used.

4. Data used for MARISA

To improve maritime safety, it is necessary to implement anticipative methods based on risk assessment techniques. Generally, formal safety assessment (FSA) methodology is adopted to identify the risk, of quantifying the risk level, and estimating the cost–benefit of a new management procedure to reduce risk. Many parameters interact with the maritime traffic situation and safety.

Maritime safety for environment protection depends on four main elements/criteria: the ship, the crew's qualification, the environment and meteorology. The recognition of a criterion is strongly related to the availability of large amount of data. For this reason, the data analysis stage is very important for risk assessment.

Several databases supply some information related to the ships. They are three kinds of data: data related to accidents, data related to ship inspections and to ship condition, and data related to the fleet. In this part, we present four databases Lloyd's Register–Fairplay (LRF), International Maritime Organization (IMO), European Quality in Shipping Information Service (EQUASIS) and Memorandum Of Understanding of Paris (Paris MOU).

The LRF database supplies information about the ships' features: for example, length, year and place of construction, registry, ownership, number and duration of detentions.

Each year, the IMO provides a list of all the accidents related to ships over 100 gross tonnage (GT). This list is established thanks to declarations of each country which declares the number of accidents or incidents for their flag.

EQUASIS is an information system related to the world fleet of ships. This database makes it possible to promote shipping quality and to fight against the use of ships which do not respect the statute law. The information system is based on the reports of inspections carried out by Port State Control (PSC).

In addition to these databases, the Paris MOU organization has defined procedures for checking ships and flags more generally. These inspections allow two factors to be defined: the target factor for the ships and the excess factor for the flags. Ship evaluation is carried out in several stages (Sage, 2005; Paris MOU, 2006). From these different processing stages, the Paris MOU organization then draws up an annual list of ships banned from ports of the signatory countries of the Paris MOU. To classify flags, the excess factor is introduced. The evaluation of a flag is based on the data collected after all the ships of the fleet have been checked. Once the excess factor is determined for all flags, they can be ordered in descending order in three lists: black, grey and white (Paris MOU, 2006).

The crew qualification criterion must take into account nationality, the name of the employment agency, the years of service in the agency and in the rank, and level of English language for each shipmate. This data is not included in databases (presented previously) that we used in this study.

Likewise, the environment is an element which includes maritime traffic and the presence of natural or artificial obstacles. In this study, these elements are not taken account but maritime traffic could be included in a future work. The last important criterion for maritime safety is the meteorology. The meteorological conditions are given by the weather bureau.

As we have explained in the introduction, many relevant static and dynamic variables exist to define an individual ship's risk factor. These variables influence the casualty rates. In our study, thanks to Degré's works (2004) and Paris MOU's study (2006), we have defined a static risk factor which depends on static data. Selected data concerns the ship's characteristics (type, size in gross tonnage, flag…) and the accident factors such as history of detentions after inspections or the number of company changes (Paris MOU, 2006). The static data is used in MARISA, it is picked up from LRF database (for the ship's data).

The dynamic risk factor is evaluated thanks to the dynamic data associated with the maritime weather conditions of the navigation area.

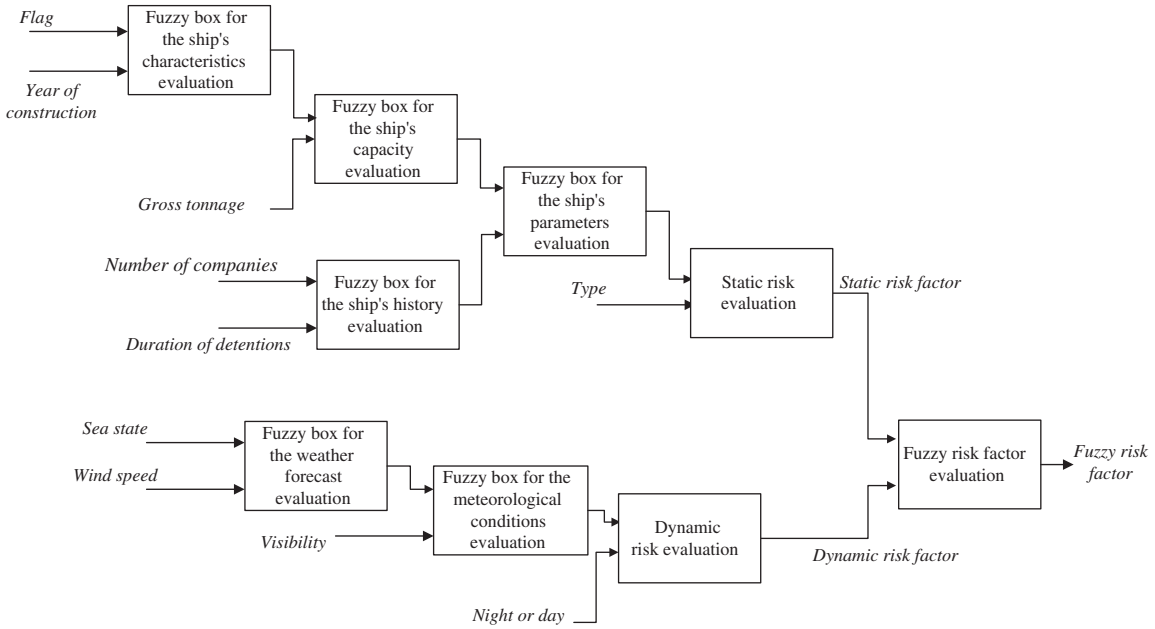


Fig. 2. Architecture of the MARISA system.

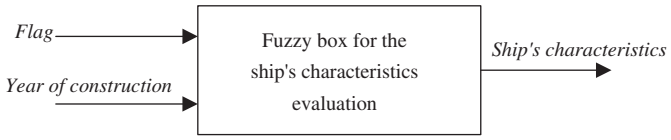


Fig. 3. *Ship's characteristics* classifier definition.

5. Architecture of the MARISA system

As we explained in Section 4, the proposed architecture presented in Fig. 2 must take into account the individual static data of the ship but also of the meteorology and oceanographic data.

For the static part, we chose six parameters about the ship: flag, year of construction, number of companies, gross tonnage, duration of detentions and type.

For the dynamic part, we took the meteorological and oceanographic conditions into account. We chose four parameters: sea state, wind speed, visibility and moment of the day (night or day).

One of the difficulties in the design of decision making systems is the large number of input variables to consider. This is a problem which makes it difficult to design systems based on expert knowledge. For this reason, we decided to define a modular and hierarchical architecture consisting of several simple modules with a maximum of two inputs and one output.

The architecture of the MARISA system consists of eight units (five for the static risk factor and three for the dynamic risk factor). The importance of selected parameters is revealed by Fig. 2. Indeed, the most influential parameters are the flag, the year of construction and also the number of companies and the duration of detentions. As we have chosen a hierarchical structure, it is possible to check the importance of selected parameters. If the user wishes a more reasonable decision making methodology, he can limit himself to the first parameters (those which appear at the top in left). In fact, the most important parameters in this hierarchical architecture are the first ones in chronological order of appearance.

The fuzzy risk factor is calculated from the static and dynamic factors. In the following paragraphs, we detail this architecture.

5.1. Static risk factor

The first four units (ship's characteristics, ship's capacity, history of ship, and ship's parameters) are fuzzy classifiers with two inputs. The last unit (static risk) is not a fuzzy classifier but a weighting function according to the type of the ship. For each variable input, the number and the type of the membership functions have been determined by considering Degré's article (2004).

First, to illustrate our approach, we describe the fuzzy box for the *Ship's characteristics* evaluation and show the simplicity of the design. The other fuzzy classifiers are designed similarly.

5.1.1. Ship's characteristics evaluation

The *Ship's characteristics* evaluation is based on the *Year of construction* of the ship and its *Flag*. Fig. 3 shows the *Ship's characteristics* classifier definition. The first stage is the fuzzification of input variables. For the *Flag* input, we used the excess factor of the Paris MOU report. In this report, the universe of discourse is comprised between 0 and 13. Thus, we defined four membership functions: low risk, middle risk, strong risk and very strong risk. Fig. 4 presents the fuzzification of the *Flag* input.

For the second input *Year of construction*, we considered an old ship, a ship built before the 1990s. Thus, as shown in Fig. 5, we defined two membership functions: old and recent. As shown in Fig. 6, the output of the fuzzy unit is made of three singletons: small, middle and high.

The fuzzy inference process is the second stage. The knowledge is represented by a set of fuzzy rules listed in Table 1. In this unit, we defined eight rules which connect premises and conclusions.

IF–THEN rules are used to represent the relationship between inputs and the output.

For example: if *Flag* is very strong risk and *Year of construction* is old then *Ship's characteristics* is high.

It means that the risk is higher if the ship is old and if the ship's flag is in the black list of the Paris MOU. The final stage is defuzzification which translates the linguistic values into numerical values. For all units, we use the center-of-gravity defuzzification method with Mandani's max–min inference method.

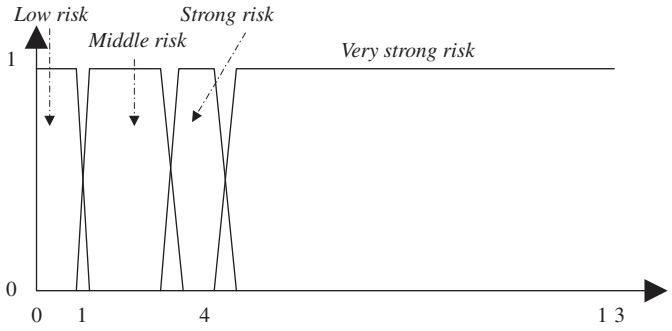


Fig. 4. Fuzzification of the *Flag* input.

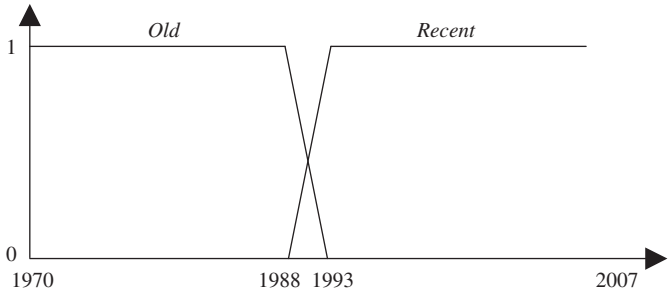


Fig. 5. Fuzzification of the *Year of construction* input

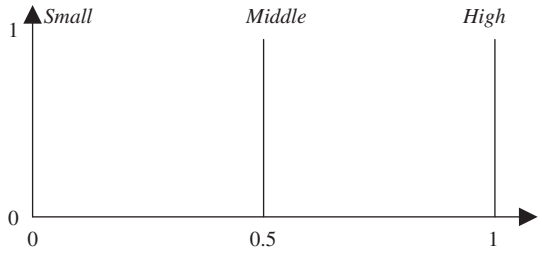


Fig. 6. Fuzzification of the *Ship's characteristics* output.

Table 1
Set of rules of *Ship's characteristics* classifier.

Flag	Year of construction	Ship's characteristics
Low risk	Old	Middle
Low risk	Recent	Small
Middle risk	Old	Middle
Middle risk	Recent	Middle
Strong risk	Old	High
Strong risk	Recent	Middle
Very strong risk	Old	High
Very strong risk	Recent	High

5.1.2. Ship's capacity evaluation

Gross tonnage is a parameter which appears in the statistic study about maritime accidents. The *Ship's capacity* classifier takes into account the *Gross tonnage* and the *Ship's characteristics* evaluation. The definition of the *Ship's capacity* classifier is illustrated in Fig. 7. This last input is the output of the *Ship's characteristics* classifier as shown in Fig. 2.

The *Ship's characteristics* input is defined like the output of the *Ship's characteristics* classifier by three membership functions: small, middle and high risk.

The *Gross tonnage* input is defined by five membership functions: small, middle, middle-large, large and very-large tonnage.

The output of the *Ship's capacity* classifier is divided into three membership functions: small, middle and high risk. This unit is governed by a set of fifteen rules described in Table 2.

5.1.3. Ship's history evaluation

The static risk study is also related to the ship's history. Indeed, it is important to control the *Number of companies* and the *Duration of detentions*. Fig. 8 shows the definition of *Ship's history* classifier.

The *Number of companies* input is the number of company changes of the ship. It is composed of three classes: small, middle and large number of companies.

The *Duration of detentions* input is the number of days of detention after the ship has been inspected. This input is defined by three membership functions: null, middle and very-large number of detention days.

The output of the *Ship's history* classifier is decomposed into three membership functions: small, middle and high risk. So, as shown in Table 3, there are nine rules.

5.1.4. Ship's parameters evaluation

The fuzzy box for the *Ship's parameters* evaluation makes it possible to define the risk for maritime safety with respect to the ship only (*Flag, Year of construction, Gross tonnage, Number of companies* and *Duration of detentions*). As shown in Fig. 2, this evaluation is carried out by using outputs of the *Ship's capacity* and *Ship's history* fuzzy boxes. In Fig. 9, the *Ship's parameters* classifier is presented.

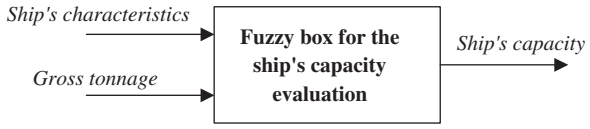


Fig. 7. *Ship's capacity* classifier definition.

Table 2
Set of rules of *Ship's capacity* classifier.

Ship's characteristics	Gross tonnage	Ship's capacity
Small	Small	Small
Small	Middle	Middle
Small	Middle-large	Middle
Small	Large	Small
Small	Very-large	Small
Middle	Small	Middle
Middle	Middle	High
Middle	Middle-large	High
Middle	Large	Middle
Middle	Very-large	Middle
High	Small	High
High	Middle	High
High	Middle-large	High
High	Large	High
High	Very-large	High

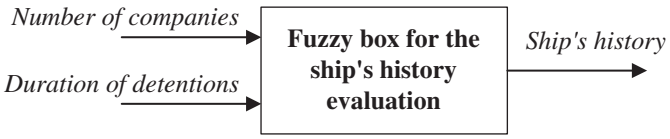


Fig. 8. *Ship's history* classifier definition.

Table 3
Set of rules of *Ship's history* classifier.

Number of companies	Duration of detentions	Ship's history
Small	Null	Small
Small	Middle	Middle
Small	Very-large	Middle
Middle	Null	Small
Middle	Middle	Middle
Middle	Very-large	High
Large	Null	Small
Large	Middle	High
Large	Very-large	High

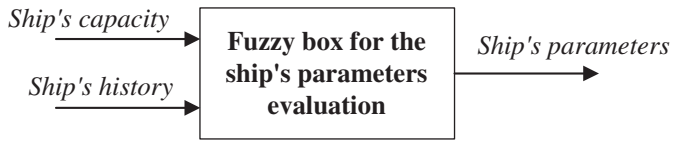


Fig. 9. *Ship's parameters* classifier definition.

Table 4
Set of rules of *Ship's parameters* classifier.

Ship's capacity	Ship's history	Ship's parameters
Small	Small	Small
Small	Middle	Middle
Small	High	Middle
Middle	Small	Middle
Middle	Middle	High
Middle	High	High
high	Small	Middle
high	Middle	High
high	High	High

The *Ship's capacity* and *Ship's history* inputs, like the output of the *Ship's capacity* classifier and *Ship's history* classifier respectively, have been decomposed into three classes: small, middle and high risk.

The output of the *Ship's parameter* classifier is divided into three membership functions: small, middle and high risk. This decomposition uses a set of nine rules listed in Table 4.

5.1.5. Static risk evaluation

The *Static risk* box is not a fuzzy box. As shown in Fig. 10, the static risk evaluation is based on the *Ship's parameter* and the *Type of ship*.

Indeed, the *Type of ship* is not fuzzy data (to simplify a ship is either a passenger ship or a tanker or a container ship or a cargo ship). According to Degré (2004), we determined the accident rate values according to the type of ship. By normalising the values compared with the rate of passenger ships, we defined a new correction function. The results are as follows: the risk for bulk carriers is 1.97 times higher than for passenger ships, for cargo ships it is 1.69 times higher, for container ships it is 1.28 times higher and for tankers it is 1.03 times higher.

So, the value of the static risk box output is multiplied by this coefficient.

5.2. Dynamic risk factor

The dynamic risk factor depends on weather conditions and the current moment of the day. Weather information (*Wind speed*, *Sea state* and *Visibility*) are extracted from the Meteo-France's data (Guide Marine, 2007).

To estimate this factor, we define two fuzzy classifiers (*Weather forecast* and *Meteorological conditions*) and a multiplier (dynamic risk) according to the moment of the day.

5.2.1. Weather forecast evaluation

In Fig. 11, we present the *Weather forecast* evaluation which is based on the *Sea state* and the *Wind speed*.

For the fuzzification of the *Sea State* input, we considered three membership functions: calm, choppy and rough. The universe of discourse for the sea state input is defined between 0 and 9 (Douglas Sea Scale).

The *Wind speed* input is decomposed by using three membership functions: calm, breeze and strong. The universe of discourse for the wind speed input is defined between 0 and 12 (Beaufort wind force scale).

The output of the *Weather forecast* classifier is defined by three membership functions: null, middle and high. As shown in Table 5, the decomposition implies a set of nine rules.

5.2.2. Meteorological conditions evaluation

As shown in Fig. 12, the *Meteorological conditions* evaluation is estimated by using the output of the *Weather forecast* classifier and *Visibility*.

The first input, *Weather forecast*, is decomposed into three membership functions: good, middle and bad.

For the fuzzification of the second input, *Visibility*, we considered four classes: fog, poor, moderate and good.

The output of the *Meteorological conditions* classifier is decomposed into three membership functions: good, middle and bad. As shown in Table 6, this decomposition uses a set of twelve rules.

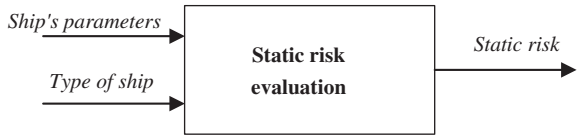


Fig. 10. *Static risk* evaluation.

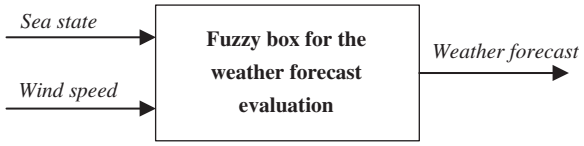


Fig. 11. *Weather forecast* classifier definition.

Table 5
Set of rules of *Weather forecast* classifier.

Sea state	Wind speed	Weather forecast
Calm	Calm	Good
Calm	Breeze	Middle
Calm	Strong	Middle
Choppy	Calm	Good
Choppy	Breeze	Middle
Choppy	Strong	Bad
Rough	Calm	Middle
Rough	Breeze	Bad
Rough	Strong	Bad

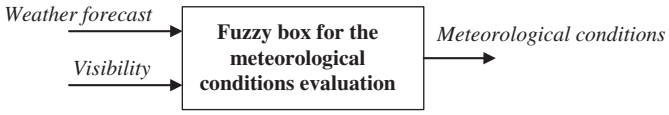


Fig. 12. *Meteorological conditions* classifier definition.

Table 6
Set of rules of *Meteorological conditions* classifier.

Weather forecast	Visibility	Meteorological conditions
Good	Fog	Bad
Good	Poor	Middle
Good	Moderate	Middle
Good	Good	Good
Middle	Fog	Bad
Middle	Poor	Bad
Middle	Moderate	Middle
Middle	Good	Middle
Bad	Fog	Bad
Bad	Poor	Bad
Bad	Moderate	Bad
Bad	Good	Bad

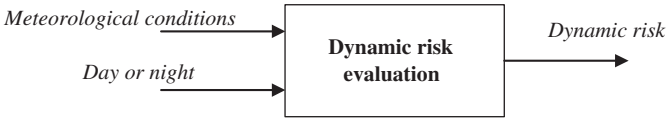


Fig. 13. *Dynamic risk* evaluation.

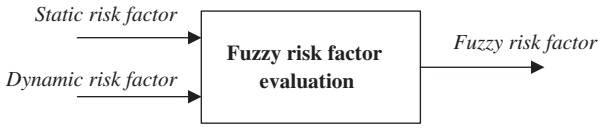


Fig. 14. *Fuzzy risk factor* evaluation.

5.2.3. Dynamic risk evaluation

In Fig. 13, we present the dynamic risk evaluation which is estimated by taking into account the result of the *Meteorological conditions* fuzzy classifier and the period of the day (*Day or night*).

This evaluation is not carried out with a fuzzy box. It is a multiplier activated only if it is dark. Indeed, we assume that at night the risk is higher. If it is night-time, we multiply the output of the previous fuzzy box (*Meteorological conditions*) by 1.5 (50% more risk than day). If it is day-time, we do not modify the output value of the previous fuzzy box.

5.3. Fuzzy risk factor

The fuzzy risk factor (FRF) evaluation is shown in Fig. 14. This factor is computed by combining the static and dynamic risk factors.

Obviously, it is necessary to study in more details this association and the respective weight of the static risk factor (SRF) and of the dynamic risk factor (DRF). The choice of the weighting is based on the fact that the dynamic output increases in proportion with the static output. The dynamic output is included between 0 (nice weather and day) and 1.5 (bad weather and night). Thus, the multiplying coefficient is equal to

$$k = \frac{DRF}{1.5} \tag{1}$$

We obtain,

$$FRF = SRF(1 + k) \tag{2}$$

Case 1: if the user considers that it is better to increase the night risk, it is necessary to replace 1.5 by an another higher value (v). In this case, the ratio Eq. (1) will be equal to

$$k = \frac{DRF}{v} \tag{3}$$

Case 2: if the user considers that the dynamic output has more importance than in the formula Eq. (2), a new expression can be taken as:

$$FRF = SRF(1 + nk) \tag{4}$$

n is an integer.

The choice of the parameters (v and n) is realised according to expert knowledge. Thus, the static risk factor weight can be adjusted in comparison to the dynamic risk factor to compute the fuzzy risk factor.

In the following section, we present several simulation results to validate this algorithm and to show the interest of this architecture.

6. Simulation and results

6.1. Simulation interface

A graphic interface was developed with the Labview software to display the risk assessment value according to input data. The inputs are fixed with pull-down menus but in a real system the information would be generated automatically.

On the interface, as shown in Fig. 15, we can see three sub-windows which concern the static risk factor, the dynamic risk factor and the fuzzy risk factor.

The fuzzy modules compute the risk factor of a single ship with regard to weather conditions inputted. Each risk factor is composed of five levels which correspond to a risk percentage: zero (0%), weak (25%), average (50%), high (75%) and very high (100%).

6.2. Results

6.2.1. Static risk factor

In this subsection, we present some illustrative examples. So, we chose three ships and for each of them we gave the Paris MOU's target factor and we computed the static risk factor.

In Table 7, we present the results obtained from the six input parameters (ship type, gross tonnage, duration of detentions, year of construction, flag and number of companies). Of course, we know that the target factor and the fuzzy static factor are not calculated with exactly the same inputs. Moreover, the target factor is used within the Paris MOU on PSC as a tool for selecting ships eligible for an inspection only. The Paris MOU defines a scale of priority of inspection based on the target factor, and composed of four levels: low priority (less than 15 points), standard priority (less than 25 points), medium (less than 35 points) and high (more than 35 points). The target factor is not an indication of the quality of the ship.

In this table, we study three different ship types (bulk carrier, cargo and oil tanker) and show that the fuzzy static risk is coherent with the target factor.

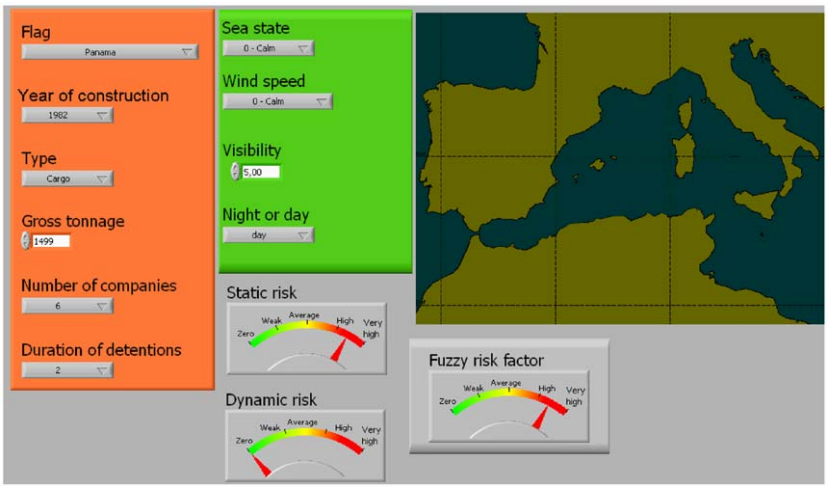


Fig. 15. Graphic interface for simulation.

Table 7
Static risk values for three examples of ship.

Ship name	**LAILA Queen**	**ALADIN I**	**ERIKA**
Ship type	Bulk carrier	Cargo	Oil tanker
Gross Tonnage	13015	1499	19666
Duration of detentions (days)	0	2	1
Year of build	1977	1982	1975
Flag	Cambodia	Panama	Malta
Companies number	2	8	9
Target factor of Paris MOU	45	177	12
Static risk	**Average (50 %)**	**High (86 %)**	**Weak (26 %)**

Table 8
Dynamic risk values.

Sea state (height of the waves)[a]	Calm–glassy	Calm–glassy	Calm–glassy	High	Phenomenal
Wind speed[b]	Calm	Calm	Calm	Strong breeze	Hurricane
Visibility[c]	Good	Poor	Good	Poor	Poor
Period of the day	Day	Day	Night	Day	Night
Dynamic risk	**Zero(0%)**	**Weak (33%)**	**Average (50%)**	**Average to high (67%)**	**Very high (100%)**

[a] Sea state (meters): calm–glassy: 0 m; high: 6–9 m; phenomenal: ≥14 m.
[b] Wind speed (knot): calm: <1 kt; strong breeze: 22–27 kt; hurricane: ≥64 kt.
[c] Visibility (nautical miles): poor: <2 and ≥0.5 NM; good: ≥5 NM.

For the ERIKA ship, we note that the target factor (value: 12) and the fuzzy static risk (26%) do not lead to the conclusion that this ship was dangerous. These results confirm that taking the static data exclusively into account is not sufficient.

In the following paragraphs, we present results taking the dynamic factor into account. So, the fuzzy risk factor is calculated by considering the meteorological conditions.

6.2.2. *Dynamic risk factor*

This section focuses on the evolution of the dynamic risk factor with respect to the weather variations. For this, five weather conditions have been simulated. The results are presented in Table 8.

The dynamic factor changes according to the meteorological conditions. It is contained between zero and very high (100%). For example, we consider two cases more precisely (columns 1 and 3, in Table 8). For these two cases, the sea is glassy, the wind speed is calm and visibility is good. The difference between these two simulations is the period of the day. When these meteorological conditions appear during the day, then the dynamic risk factor is zero. However, when these conditions occur at night, then the dynamic risk factor increases to average (50%).

The several results obtained are consistent. They are validated from the marine guide (Guide marine, 2007).

6.2.3. *Fuzzy risk factor*

The fuzzy risk factor is calculated by combining the static and dynamic risk factors. So we come back to the same examples (same ships) as the static risk factor estimation and we take into account the same weather conditions as the dynamic risk factor estimation.

The results are presented in Table 9.

Table 9
Fuzzy risk factor values.

LAILA QUEEN—static risk factor: average (50%)					
Dynamic risk factor	Zero (0%)	Weak (33%)	Average (50%)	Average to high (67%)	Very high (100%)
Fuzzy risk factor	**Average (50%)**	**Average to high (67%)**	**High (75%)**	**High to very high (84%)**	**Very high (100%)**
ALADIN—static risk factor: high (86%)					
Dynamic risk factor	Zero (0%)	Weak (33%)	Average (50%)	Average to high (67%)	Very high (100%)
Fuzzy risk factor	**High (86%)**	**Very high (100%)**	**Very high (100%)**	**Very high (100%)**	**Very high (100%)**
ERIKA—static risk factor: weak (26%)					
Dynamic risk factor	Zero (0%)	Weak (33%)	Average (50%)	Average to high (67%)	Very high (100%)
Fuzzy risk factor	**Weak (26%)**	**Weak (35%)**	**Weak(39%)**	**Weak to average(44%)**	**Weak to average (52%)**

Two observations can be made. First, when the dynamic risk factor is zero (the first case), the fuzzy risk factor is equivalent to the static risk factor. Next, if we study, for each ship, the fuzzy risk factor obtained with the various weather conditions simulated, we note the fuzzy risk factor increases if the weather conditions get worse.

So, for all cases except for the first (dynamic risk factor is zero), the fuzzy risk factor is higher than the static risk factor. From these results, according to the ships' characteristics (known from the databases) and meteorological conditions, we can consider that the fuzzy risk factor is efficient to give an individual risk assessment factor.

7. Conclusions

The use of an individual ship risk index for safety at sea (IRIS) has been suggested in an article proposed by Degré (2004). In its conclusion, the author suggests that this indicator value could be used, in real time to detect a suspect ship.

In this article, we have proposed to take into account an individual ship risk factor which concerns the ship's characteristics while considering the weather conditions. In this way, we can obtain a risk factor for each ship.

To calculate this risk factor, we propose a new and flexible risk modelling approach using the fuzzy logic advantages. Thus, we have defined a modular and hierarchical structure based on the definition of several simple units. Through this approach, interactive risk scenarios in situations of highly uncertain data can be facilitated.

In order to evaluate our approach of risk assessment for maritime safety, we developed a graphic simulation interface which allows maritime traffic scenarios to be built easily. The case study of the risk assessment of various kinds of ships (bulk carrier, passenger, cargo and oil tanker) has demonstrated that the fuzzy risk factor can be used efficiently in risk analysis.

The objectives of future works will be to take into account the trajectory of the ships (speed, position compared with shipping lanes) and maritime traffic conditions.

Another objective is to integrate the FRF computation library into a maritime monitoring system such as NAOS MS2 (Naval Operation System) developed by DCNS to use real maritime traffic data.

Acknowledgments

This work has been supported by DCNS (expert in naval systems)—Project no. 5049433.

The English version has been reviewed by Michel Dufour (University of South-Toulon-Var).

References

Degré, T., Benabbou, Z., 2005. Vers une automatisation de la détection des navires à haut risque avec IRIS, un indicateur de risque individuel de navires pour la sécurité en me. Etude de l'effet de certaines variables sur les taux d'accidents. Recherche Transports Sécurité 86, 1–15.

Degré, T., Benabbou, Z., 2004. Automatisation de la détection des navires à haut risque avec IRIS, un Indicateur de Risque Individuel de navires pour la Sécurité en mer. Actes des 5ièmes Journées Scientifiques et Techniques du Cetmef, France, pp. 71–80.

Degré, T., 2004. IRIS: un indicateur de Risque Individuel de navire pour la Sécurité en mer fondé sur les concepts des modèles d'évaluation des risques. Revue de l'Electricité et de l'Electronique 3, 23–29.

Degré, T., 2003. L'importance d'une approche de la sécurité maritime fondée sur les modèles d'évaluation des risques. Revue Recherche Transports Sécurité 78, 21–32.

Eleye-Datubo, A.G., Wall, A., Wang, J., 2008. Marine and offshore safety assessment by incorporative risk modeling in a fuzzy-bayesian network of an induced mass assignment paradigm. Risk Analysis 28 (1), 95–112.

Haj-Salem, H., Ramananjaona, C., Kates, R., 2006. EUropean RAmp Metering Project (EURAMP). Safety Critical Issues—Project no.: 507645.

Glansdorp, C.C., 2004. European Research Projects—Embarc and MarNIS. Actes des 5ièmes Journées Scientifiques et Techniques du Cetmef, France, pp. 81–93.

Guide marine, 2007. 〈http://www.meteo.fr/meteonet/decouvr/guides/marine/mar1.htm〉.

Hu, S., Fang, Q., Xia, H., Xi, Y., 2007. Formal safety assessment based on relative risks model in ship navigation. Reliability Engineering and System Safety 92, 369–377.

Liu, J., Yang, J.B., Wang, J., Sii, H.S., 2005. Engineering system safety analysis and synthesis using the fuzzy rule-based evidential reasoning approach. Quality and Reliability Engineering International 21, 387–411.

Paris MOU on Port State Control, 2006. Port State Control steady she goes. Annual Report.

Regelink, H., Glansdorp, C., Jarvis, D., 2004. Safety at sea and efficiency in ports; a European project to improve the advanced use of VTM. In: 10th International Symposium on Vessel Traffic Services, Hong Kong, pp. 10–13.

Sage, B., 2005. Identification of high risk vessels in coastal waters. Marine Policy 29, 349–355.

Sii, H.S., Ruxton, T., Wang, J., 2001. A fuzzy-logic-based approach to qualitative safety modelling for marine systems. Reliability Engineering and System Safety, 19–34.

Sii, H.S., Wang, J., Ruxton, T., Yang, J.B., Liu, J., 2004. Application of fuzzy logic approaches to safety assessment in maritime engineering applications. Journal of Marine Engineering and Technology A5, 45–58.

Van der Heijden, W., Glansdorp, C., Sage, B., Trant, G., Veldhuyzn, W., Degré, T., 2004. Embarc WP3: Tracking and Tracing; SWP3.2: Monitoring of High Risk Vessels at Medium and Long Ranges. Contract no. GRD1-2000-25500.

Wang, J., Sii, H.S., Yang, J.B., Pillay, A., Yu, D., Liu, J., Maistralis, E., Saajedi, A., 2004. Use of advances in technology for Maritime Risk Assessment. Risk Analysis 24 (4), 1041–1063.

Yang, A., Bonsall, S., Wang, J., 2008. Fuzzy rule-based Bayesian reasoning approach for prioritization of failures in FMEA. IEEE Transactions on Reliability 57 (3), 517–528.

Zadeh, L., 1965. Fuzzy sets. Information and Control 8, 338–353.

5.5 Aide à la décision : global

Ocean Engineering 38 (2011) 171–176

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

A decision-making system to maritime risk assessment

Jean-François Balmat [a,*], Frédéric Lafont [a], Robert Maifret [b], Nathalie Pessel [a]

[a] LSIS, UMR CNRS 6168, University of South-Toulon-Var, B.P. 20132, F-83957 La Garde Cedex, France
[b] DCNS, Division SIS, B.P. 403, F-83055 Toulon Cedex, France

ARTICLE INFO

Article history:
Received 25 May 2010
Accepted 17 October 2010
Editor-in-Chief: A.I. Incecik
Available online 11 November 2010

Keywords:
Maritime risk assessment
Maritime safety
Fuzzy risk factor

ABSTRACT

In this study, we propose a fuzzy approach in order to evaluate the maritime risk assessment applied to safety at sea and more particularly, the pollution prevention on the open sea. The work is based on the decision-making system, named MARISA, presented in Balmat et al. (2009). This system allowed defining a risk factor for each ship according to ship's characteristics and weather conditions. In this novel paper, the proposed system takes into account the ship speed evolution and the ship position with respect to maritime shipping lanes is developed. To validate the method, we present an example of results with real data.

1. Introduction

Nowadays, the maritime risk assessment is an important research theme. Like this, many studies have been realised to identify high risk ships (Degré, 2003; Glansdorp, 2004; Regelink et al., 2004; Van der Heijden et al., 2004; Wang et al., 2004; Sage, 2005; Haj-Salem et al., 2006). The maritime risk modelling is a subject who takes into account several notions such as maritime safety relative to the traffic or the environment protection. The objective of this present work is to evaluate the maritime risk assessment within the framework of the environment protection and more precisely, to prevent the oil pollution.

This paper presents a study about a fuzzy approach for the MAritime RISk Assessment (MARISA) applied to safety at sea which is an extension of previous article proposed by Balmat et al. (2009). In this first study, the authors defined an individual risk factor for each ship determined from a fuzzy approach. This factor has been obtained from several static data relative to the ships (age, flag, gross tonnage, number of companies, duration of detention and type) and by considering the meteorological conditions (sea state, wind speed and visibility). To design a flexible decision-making tool, the authors have designed a modular and hierarchical structure. Furthermore, to evaluate the approach, a simulator has been developed and some tests of several ships have been presented. In this context, a risk assessment study of various kinds of ships (bulk carrier, passenger, cargo and oil tanker) has demonstrated the efficiency of the approach in risk analysis.

However, to improve this system, it is necessary to take into account some other dynamic parameters such as ships speed and their positions compared with shipping lanes. This is the purpose of this novel article. The paper is organized as follows. Firstly, a brief introduction of risk factor definition is proposed. In this part, the choice of relevant data, and the proposed fuzzy approach are presented. In Section 3, the architecture of the new MARISA system is described and the three risk factors (static, meteorological and dynamic) allowing to compute a global risk factor are presented. Section 4 details more precisely the design of the fuzzy classifier about the ship's speed evolution. In the last section, a simulation from a scenario on a passenger ship which navigates in Mediterranean Sea on the shipping lane Fos-Napoli is depicted. The obtained results have been validated by a human expert and show the interest of this system.

2. Risk factor definition

The aim of this work is to design a decision-making system enable to evaluate an individual maritime risk factor in the oil pollution prevention context. For this, in a first stage, by performing an expert analysis of the problem, the relevant input data have been selected. In fact, the knowledge acquisition is based on statistical data and information analysis (Degré and Benabbou, 2005) mixed with human expert experience. Therefore, a decision-making system based on fuzzy classifiers has been developed to define the risk factor.

2.1. Choice of the relevant data

In several papers, T. Degré et al. (2003–2005) defined an individual ship risk index for safety at sea (IRIS). In the papers, the authors analysed the data about accidents listed by the

* Corresponding author. Tel.: +330494142039.
E-mail address: balmat@univ-tln.fr (J.-F. Balmat).

0029-8018/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oceaneng.2010.10.012

International Maritime Organization (IMO) for many years. From these works, a decision-making system for the Maritime Risk Assessment has been developed (Balmat et al., 2009). Maritime safety for environment protection depends on many elements and criteria concerning, for example, the ship's characteristics, the ship's history elements, the ship's trajectory and the meteorological conditions. Data about these criteria are very important to design an efficient decision-making system. For this reason, the relevant available data which can be used in our system have been listed. These input data allow computing the individual risk factor for each ship. The set of data can be obtained from a set of database (Lloyd's Register, IMO, EQUASIS, Paris MOU). Like this, the data on the ship's characteristics (type, flag, year built, gross tonnage, simple or double hull), the ship's historic elements (number of companies, duration of detentions), the trajectory (position and speed of the ship, last known port and destination) and the meteorology parameters (sea state, wind speed, visibility, night or day) have been chosen (Sage, 2005; Paris MOU, 2006; Guide marine, 2007). Moreover, the environmental risk assessment for different type of ships is taken into account according to accident rate study (Degré, 2004).

2.2. Background of the fuzzy classifier

The fuzzy theory was introduced by Zadeh (1965). In this paper, several fuzzy classifiers are defined to design the decision-making system. When a training data set is not available, a classifier can be designed from prior knowledge and expertise. Experts are able to give the class labels using the if-then rules, their membership functions and the rules base.

For the fuzzy classifier inputs, each linguistic value is represented by a membership function. The consequent part of the rule may also contain linguistic values (Mamdani, 1977) or functions (Takagi and Sugeno, 1985). To calculate the output, several methods of defuzzification exist but the most used method is the centre-of-gravity. The maritime risk evaluation can find an interest in the fuzzy logic approach because much data are linguistic variables. Indeed, fuzzy classifier allows decision making with estimated values under incomplete or uncertain information, e.g., small, large, far, fast, slow, etc.

2.3. Fuzzy for maritime risk assessment

The fuzzy logic is an efficient approach for design a decision-making system in maritime domain. This technique allows solving a lot of problems related to dealing the imprecise and uncertain data. Moreover, fuzzy logic enables to take into account the insufficient information and the evolution of available knowledge. Naturally, this is not the only method to deal with the uncertainty analysis and there are many different approaches.

Therefore, Liu et al. (2005) reviewed the uncertainty reasoning approaches for decision-making. The authors discussed of the different formal techniques and they described their possible application in maritime risk assessment. They presented the Bayesian theory of probability, the Dempster–Shafer theory of evidence and the fuzzy set theory. They compared these three approaches and they expressed their strengths and weaknesses. In fact, the choice of the approaches depends on the several elements including the availability data (qualitative and qualitative information), the degree of interrelationships complexity, the causes of "uncertainty" and the languages required by the operator. Depending on the type of applications it could be necessary to combine the approaches.

Wang et al. (2004) studied several risk modelling and decision-making approaches. Therefore, the authors discussed the potential applicability of fuzzy set theory to uncertainty analysis of accident. They explained that uncertainties in maritime risk assessment are relevant to impression associated with the complexity of a system as well as vagueness of human judgments.

Other authors studied the safety assessment in maritime domain by using fuzzy logic approach (Sii et al., 2001, 2004; Liu et al., 2005; Hu et al., 2007; Eleye-Datubo et al., 2008; Yang et al., 2008). In these articles, all authors confirm the fuzzy logic interest for maritime risk assessment.

Thus, in the proposed study, we used the fuzzy approach in order to determine an individual ship risk factor. Indeed, data used are more or less vague and should be used with an approximate reasoning and thus with an imprecise logic. For example, there will be uncertainty elements to determine the environment conditions or in knowledge of the ship. In this maritime context, human experts may have also difficulties to define an accurate value, although they can express the rules with vague words. In fact, in order to define fuzzy rules, we considered the expert knowledge combined with Degré's data analysis which enables to determine the relationships between several parameters and the accident rate.

Finally, due to the large number of input variables and, knowing the complexity design of expert systems, it will be necessary to define an architecture enable to simplify the system structure. For this, a modular and hierarchical architecture has been defined.

3. Architecture of the MARISA system

This architecture is modular because the rules modification must be easily realizable; it is hierarchical in the aim to simplify the fuzzy block's development. The system consists in 12 simple modules and one block for decision logic as shown in Fig. 1. The simple modules contain two inputs and one output.

The maritime risk assessment will be appreciated for each ship, from four risk factors relating to static characteristics, meteorological conditions, ship speed evolution and its position compared with shipping lanes.

3.1. Static risk factor

For the static part, Degré's analysis of the maritime accidents data listed by the International Maritime Organization (IMO) for several years is used (Degré, 2004). From this, a fuzzy structure is designed with which we would determine a static risk factor (Balmat et al., 2009). The input data about this factor are the flag, year of construction, number of companies, gross tonnage, duration of detentions and type of ships and type of hulls. The static risk factor is estimated from four fuzzy classifiers with two inputs linked at two weighting blocks. As described in Balmat et al. (2009), with the fuzzy classifier *Ship's characteristics* it makes possible to give a first estimated value of risk concerning the ship. This value depends on the *Year of construction* (recent or old) and the *Flag* (using the Excess Factor of Paris MOUs). The parameter *Gross tonnage* is determined from the study of maritime accidents given by Degré. The fuzzy classifier *Ship's capacity* adds the gross tonnage to the evaluation of ship's characteristics. To complete this risk study, it is necessary to evaluate the *Ship's history* in taking into account of the *Number of companies* (small, middle or large) and of the *Duration of detentions* (null, middle or very-large). A fourth fuzzy classifier (*Ship's potential*) allows evaluating directly the risk corresponding to the ship. Finally, two other characteristics of the ship (*Type of ship* and *Type of hull*) have been taken into consideration by two weighting blocks which are not fuzzy (*Ship's potential* 1 and *Ship's potential* 2). According to Degré (2004), the accident rate values can be determined in relation to the type of ship. By normalising the values compared with the rate of passenger ships, a correction function is defined. The results are as follows: the risk

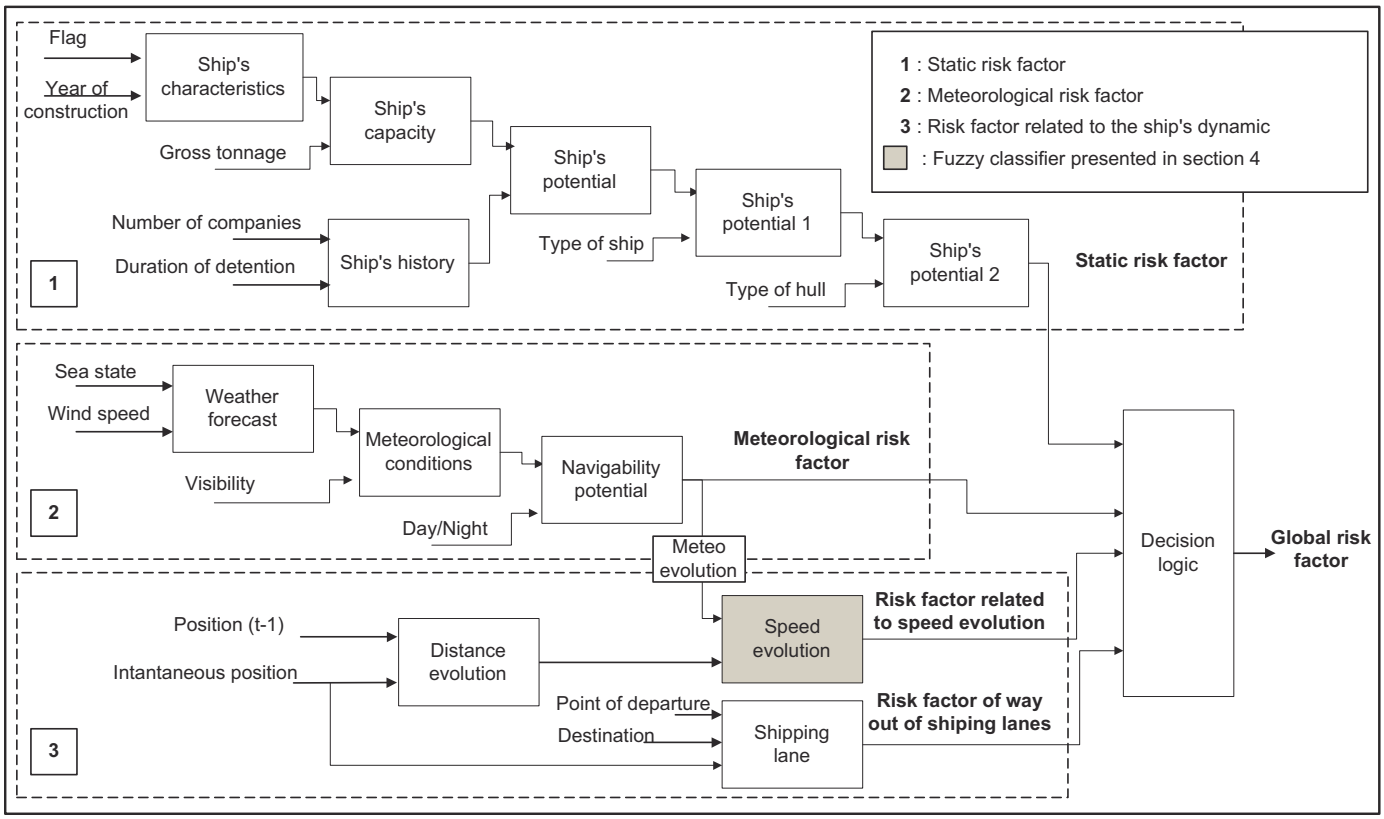


Fig. 1. Maritime risk assessment architecture.

for bulk carriers is 1.97 times higher than for passenger ships, for cargo ships it is 1.69 times higher, for container ships it is 1.28 times higher and, for tankers it is 1.03 times higher. Therefore, the value of the block output *Ship's potential* 1 is multiplied by these coefficients. Finally, to obtain static risk factor the *Type of hull* is considered for weighting the output *Ship's potential* 2. There are both pros and cons associated with double hull tanker designs. In our study, the static risk factor is considered higher for a simple hull ship (for example: 1.20 times higher).

3.2. Meteorological risk factor

The meteorological risk factor depends on the meteorological conditions and on the day period. The meteorology data (wind speed, sea state and visibility) are given by Meteo France from his marine guide. The meteorological marine scales are defined in this guide (Beaufort, sea state or the visibility scales) and allow defining value intervals. For example, force 6 of Beaufort scale is related to a strong breeze which is equal to a wind speed between 21 and 26.45 knots. With these values, we defined the intervals of the membership functions. We used the same principle for other scales. At last, the *meteorological risk factor* is estimated with two fuzzy classifiers (*Weather forecast* and *Meteorological conditions*) after weighting according to the period of the day (*Day* or *Night*) (Balmat et al., 2009). This is the output of the *Navigability potential* block.

3.3. Risk factors related to the ship's dynamic

The study of ship's dynamic emphasizes two kinds of suspect behaviours. Firstly, the distance between two successive instants is insufficient (due to a drift controlled or a damage) or too large. Secondly, due to a bad sea state or a pollution action, the ship changes its trajectory. Considering these behaviours two risk factors are defined: the *Risk factor related to speed evolution* and the *Risk factor of way out of shipping lanes*. When the behaviour of the ship is suspect, these factors are estimated with logics of decision depending on several criteria.

The *Risk factor related to speed evolution* is related to meteorological change and to the distance between two successive instants. In this case, the hypothesis that the ship's speed at open sea can be considered as more or less constant for stable meteorological conditions is taken. This rule permits to define the fuzzy classifier *Speed evolution*.

The *Risk factor of way out of shipping lanes* allows verifying trajectory of the ship compared with the shipping lane which is defined by the points of departure and destination. To estimate this factor, the shipping lane is supposed known and that the trajectory of the ship is supposed described by line segments. The exact trajectory is estimated by computing, at time t (instantaneous position), the coordinates of the ship's position projected on the shipping lane (symbolised by a line segment). The radius projection is calculated with the bloc *Shipping lane* which is not fuzzy. Therefore, if a ship is inside a circle with radius inferior to a threshold which is defined according to the nominal trajectory then it is not considered suspect. The threshold value depends on the difference between the real and ideal trajectories.

3.4. Global risk factor

Finally, a global risk factor is determined from the four previous risk factors thanks to decision logic block. A first operator which realises a weighting between the *Static risk factor* and the *Meteorological risk factor* is defined. Moreover, combination logic between the output of

the previous operator is realised i.e., between the *Risk factor related to speed evolution* and the *Risk factor of way out of shipping lanes*. So, three possibilities can arise:

- If all factors are false then the ship is no suspect.
- If only one factor is true then the ship is suspect.
- If two or three factors are true then the ship is very suspect.

The structure is the same for all the fuzzy blocks; they are designed with the same principle (two inputs, one output and fuzzy method). Therefore, we use the centre-of-gravity defuzzification method with Mandani's max–min inference method (Jager, 1995). In the following section, to depict our approach, the fuzzy classifier *Speed evolution* evaluating the ship's trajectory is detailed.

4. Fuzzy classifier *Speed evolution* design

Our study is focused on oil pollution prevention at sea. In this case, some rules can be defined to depict a risk ship. For example, it is possible to consider several suspect maneuvers as zigzags or successive accelerations/decelerations. Considering these kinds of examples and according to the human knowledge, some rules about the dynamic ship behaviour are described. More particularly, the two suspect behaviours following are studied:

Case 1: An insufficient distance between two instants (decrease of velocity).

Case 2: A sudden trajectory modification.

As shown in Fig. 2, the fuzzy box *Speed evolution* enables to evaluate the ship's trajectory according to the two inputs: *Meteo evolution* and *Distance*.

As shown in Fig. 3, for the fuzzification of the *Meteo evolution* input, three membership functions are considered: Negative (Neg), Null (N) and Positive (Pos). This input is the meteorological risk factor at two successive instants.

Therefore, we have the three following cases:

Case 1: If the weather has improved then the difference is negative (Neg).

Case 2: If the weather has worsened then the difference is positive (Pos).

Case 3: If the weather is stationary then the difference is null (N).

The fuzzification of the *Distance* input (evolution of distance covered between two instants) is defined by three membership functions as shown in Fig. 4: Short (S), Normal (N) and Long (L). As the positions are taken into account in latitude/longitude coordinates, the reference set (universe of discourse) is comprised between 0 and 1.

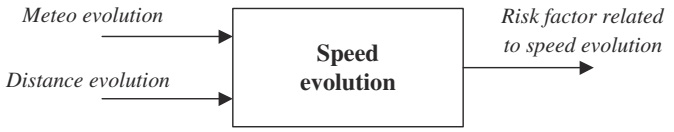


Fig. 2. Speed evolution classifier.

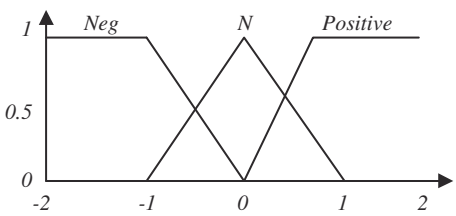


Fig. 3. Fuzzyfication of the *Meteo evolution* input.

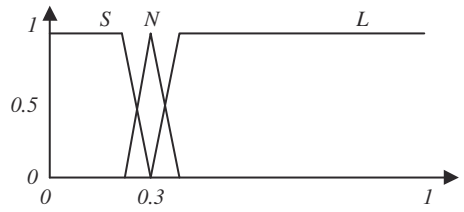


Fig. 4. Fuzzyfication of the *Distance* input.

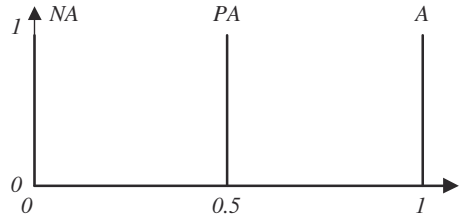


Fig. 5. Fuzzyfication of the Speed evolution classifier.

Table 1
Set of rules of *Speed evolution* classifier.

Meteo evolution	Distance	Speed evolution
Neg	S	A
Neg	N	NA
Neg	L	NA
N	S	A
N	N	NA
N	L	NA
Pos	S	NA
Pos	N	PA
Pos	L	A

The last geographic position P_{geo} is stored; it is defined by the following relation (1):

$$P_{geo}(t-1) = (latitude(t-1), longitude(t-1)) \quad (1)$$

From this position and with the instantaneous geographic position we calculate the orthorhombic distance by Eq. (2):

$$Distance = 60\arccos[\cos(lat_{inst})\cos(lat_{t-1})\cos(long_{inst} - long_{t-1}) + \sin(lat_{inst})\sin(lat_{t-1})] \quad (2)$$

with $lat_i = latitude(i)$ and $long_i = longitude(i)$

The output of the *Speed evolution* classifier is divided into three singleton membership functions (Fig. 5): Not Alarm (NA), Possibility of Alarm (PA) and Alarm (A). As described in Table 1, this decomposition uses a set of nine rules.

5. Simulation and results

In this section, an example of obtained results from the MARISA simulator is presented. This decision-making system has been developed to validate the approach through the implementation of several scenarios. Results have been obtained by using real data. Thus, several files with real trajectories and ship's characteristics corresponding have been used. These files contain the latitude/longitude coordinates and the meteorological conditions in the navigation zone of the ships.

The real maritime traffic data stem from AIS (Automatic Identification System) real data recorded by DCNS (expert in naval

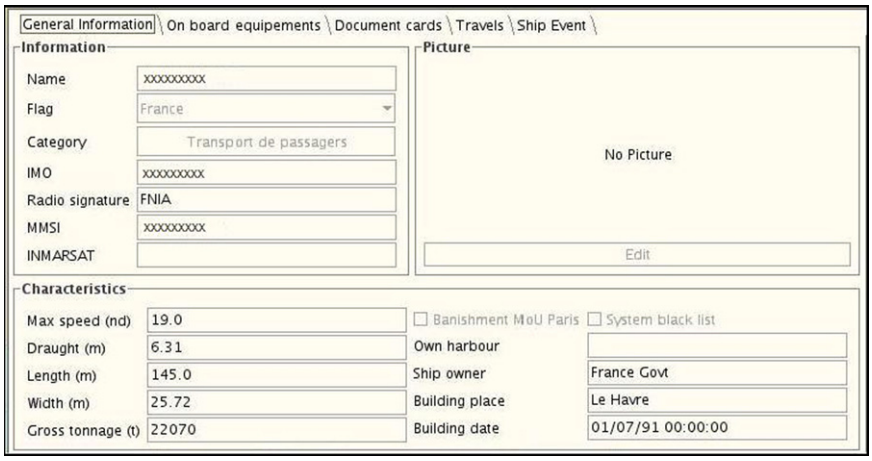


Fig. 6. General information about the ship.

systems). For each ship, the following data are stored: date, hour, latitude, longitude, speed and course. The meteorological data have been stored from Meteo France data related to the date, hour and zone. The time sampling depends on the maritime regulation. So, the ships must give their position with an imposed frequency.

This example of simulation has been validated by a human expert. Because of the confidentiality problem, the name and the IMO numbers of the ship are canceled. To carry out the simulations, the static database Lloyd's Register-Fairplay (numbers of companies, duration of detentions, simple or double hull), the meteorological conditions and visibility in the ship's navigation zone (data of Meteo France), and ships trajectories data given by DCNS are considered.

5.1. Simulation parameters

For this simulation the general information about the ship shown in Fig. 6 is used. This is a passenger ship, built in 1991, with a French flag, and its gross tonnage is equal to 22,070 tons.

The database Lloyd's Register-Fairplay gives the following static data: no company changes, no detentions and double hull. These data will be used to compute the *Static risk factor*.

The ship navigates in Mediterranean Sea on the shipping lane Fos-Napoli. The ship's positions, the ship's speed and its course are given in Table 2. Therefore, the meteorological data of this zone as shown in Table 3 are used.

5.2. Commented results

Fig. 7 shows the scenario results. The time sampling is equal to 3 h. The first three positions of the ship are not suspect because the evolution and the trajectory are correct. Therefore, the instantaneous positions of the ship are shown in green.

For the four following positions, the ship moves away from the shipping lane. According the fixed threshold, the decision-making system detects this trajectory fault and, for this reason the instantaneous positions of the ship are shown in red. At the same time, an alarm message window describes the causes as shown in Fig. 8.

Moreover, the visibility has worsened between the sixth and seventh positions of the ship. Considering the maximum ship's speed, the time between the last two positions, and the slight worsening weather then the last position of the ship seems suspect. A second alarm is activated indicating that the speed evolution of the ship is suspect (Fig. 9).

Table 2
Ship's trajectory (01/10/2008).

Hour	Latitude	Longitude	Speed (knot)	Course (deg)
12:14:19	43°08′15″N	005°23′10″E	19.1	122.8
13:14:31	42°57′53″N	005°44′07″E	18.8	121.9
14:14:43	42°47′23″N	006°05′04″E	18.2	124.7
15:14:12	42°37′15″N	006°25′51″E	19.3	123.7
16:14:53	42°26′30″N	006°47′46″E	18.9	123.4
16:46:24	42°21′06″N	006°58′44″E	18.8	123.2
19:33:03	41°52′27″N	007°56′25″E	18.6	123.6

Table 3
Meteorological data (01/10/2008).

Hour	Sea state	Wind speed	Visibility	Period (day=1 and night=0)
12:14:19	4	3	4	1
13:14:31	4	3	4	1
14:14:43	3	3	4	1
15:14:12	3	4	4	1
16:14:53	3	4	4	1
16:46:24	3	4	4	1
19:33:03	3	4	3	1

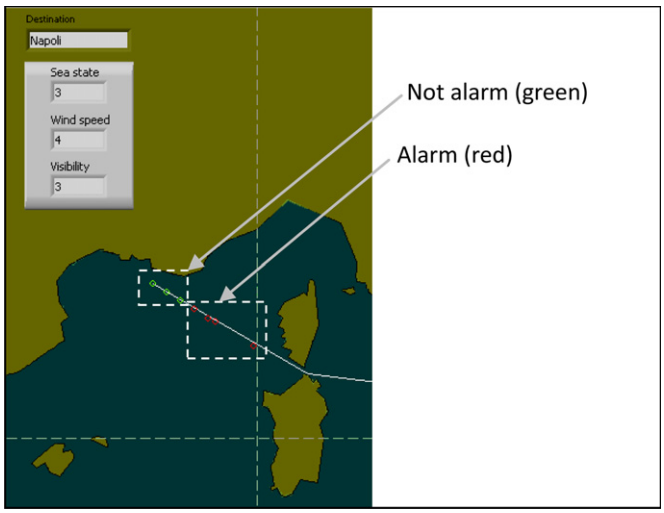


Fig. 7. *Scenario results.*

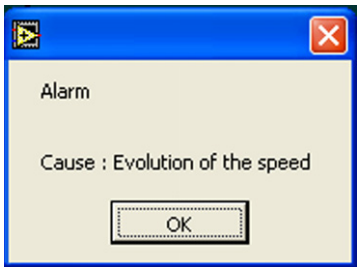


Fig. 8. *First alarm.*

Fig. 9. *Second alarm.*

This scenario has been validated by a human expert. As a member of French Navy, he defined realistic scenario for the ship route and also typical suspect behaviours of infringing ships according to his experience in maritime surveillance missions. He also validated that MARISA system was taking into account these rules in a wide range of weather conditions and routes. Unfortunately we had not enough time with him to investigate some other kind of scenario. However, the obtained results, according to our decision-making system, showed that MARISA can be considered as an efficient system to give an individual risk assessment factor. Nowadays, any comparison with other decision making systems is available. Indeed, maritime risk assessment systems exist but it is difficult to compare with MARISA system because the objectives are not really the same.

6. Conclusions

In this paper, a fuzzy approach is proposed to define an individual ship risk factor. In a first study (Balmat et al., 2009), it had been designed a decision-making system (MARISA) which allowed to define a risk factor based only on static parameters and meteorological conditions. In this novel study, the proposed work considers also trajectory of the ships and more particularly the ship's speed and the ship's position compared with shipping lanes. Easily, because the architecture of our decision-making system is hierarchical and modular, these new functions according to several fuzzy blocks have been added.

In order to evaluate our approach, with real data, several scenarios for several ships have been simulated. The set of simulations have been approved by a human expert. In this paper, a scenario relating to a passenger ship has been developed and the corresponding results are given. These results are coherent and make it possible to confirm interest to use this fuzzy approach for this kind of problems. With this architecture and to improve our system, new data are considered such as, for example, maritime traffic density or maritime risk zones.

Acknowledgments

This work was supported by DCNS (expert in naval systems)—Project no. 5049433.

References

Balmat, J.F., Lafont, F., Maifret, R., Pessel, N., 2009. Matitime RISk Assessment (MARISA), a fuzzy approach to define an individual ship risk factor. Ocean Engineering 36, 1278–1286.
Degré, T., Benabbou, Z., 2005. Vers une automatisation de la détection des navires à haut risque avec IRIS, un indicateur de risque individuel de navires pour la sécurité en mer ? Etude de l'effet de certaines variables sur les taux d'accidents. Recherche Transports Sécurité 86, 1–15.
Degré, T., Benabbou, Z., 2004. Automatisation de la détection des navires à haut risque avec IRIS, un Indicateur de Risque Individuel de navires pour la Sécurité en mer. Actes des 5ièmes Journées Scientifiques et Techniques du Cetmef, France, 71–80.
Degré, T., 2004. IRIS: un indicateur de Risque Individuel de navire pour la Sécurité en mer fondé sur les concepts des modèles d'évaluation des risques. Revue de l'Electricité et de l'Electronique 3, 23–29.
Degré, T., 2003. L'importance d'une approche de la sécurité maritime fondée sur les modèles d'évaluation des risques. Revue Recherche Transports Sécurité 78, 21–32.
Eleye-Datubo, A.G., Wall, A., Wang, J., 2008. Marine and offshore safety assessment by incorporative risk modeling in a fuzzy-Bayesian network of an induced mass assignment paradigm. Risk Analysis 28 (1), 95–112.
Glansdorp, C.C., 2004. European research projects—Embarc and MarNIS. Actes des 5ièmes Journées Scientifiques et Techniques du Cetmef, France, 81–93.
Guide marine, 2007. Available from: 〈http://www.meteo.fr/meteonet/decouvr/guides/marine/mar1.htm〉.
Haj-Salem, H., Ramananjaona, C., Kates, R., 2006. EUropean RAmp Metering Project (EURAMP). Safety Critical Issues—Project number: 507645.
Hu, S., Fang, Q., Xia, H., Xi, Y., 2007. Formal safety assessment based on relative risks model in ship navigation. Reliability Engineering and System Safety 92, 369–377.
Jager, R., 1995. Fuzzy logic in control. Ph.D., ISBN:90-9008318-9.
Liu, J., Yang, J.B., Wang, J., Sii, H.S., 2005. Engineering system safety analysis and synthesis using the fuzzy rule-based evidential reasoning approach. Quality and Reliability Engineering International 21, 387–411.
Mamdani, E.H., 1977. Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Transactions on Computers 26 (12), 1182–1191.
Paris MOU on Port State Control, 2006. Port State Control steady she goes. Annual Report.
Regelink, H., Glansdorp, C., Jarvis, D., 2004. Safety at sea and efficiency in ports; a European project to improve the advanced use of VTM. In: Proceedings of the 10th International Symposium on Vessel Traffic Services, Hong Kong, pp. 10–13.
Sage, B., 2005. Identification of high risk vessels in coastal waters. Marine Policy 29, 349–355.
Sii, H.S., Ruxton, T., Wang, J., 2001. A fuzzy-logic-based approach to qualitative safety modelling for marine systems. Reliability Engineering and System Safety, 19–34.
Sii, H.S., Wang, J., Ruxton, T., Yang, J.B., Liu, J., 2004. Application of fuzzy logic approaches to safety assessment in maritime engineering applications. Journal of Marine Engineering and Technology A5, 45–58.
Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its application to modeling and control. IEEE Transactions on Systems, Man and Cybernetics 15, 116–132.
Van der Heijden, W., Glansdorp, C., Sage, B., Trant, G., Veldhuyzn, W., Degré, T., 2004. Embarc WP3: tracking and tracing; SWP3.2: monitoring of high risk vessels at medium and long ranges. Contract no. GRD1-2000-25500.
Wang, J., Sii, H.S., Yang, J.B., Pillay, A., Yu, D., Liu, J., Maistralis, E., Saajedi, A., 2004. Use of advances in technology for maritime risk assessment. Risk Analysis 24 (4), 1041–1063.
Yang, A., Bonsall, S., Wang, J., 2008. Fuzzy rule-based Bayesian reasoning approach for prioritization of failures in FMEA. IEEE Transactions on Reliability 57 (3), 517–528.
Zadeh, L., 1965. Fuzzy sets. Information and Control 8, 338–353.

5.6 Observateur grand gain adaptatif

Journal of Process Control 21 (2011) 893–900

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

An adaptive high-gain observer for wastewater treatment systems

Frederic Lafont[a,*], Eric Busvelle[b], Jean-Paul Gauthier[a]

[a] Universite du Sud-Toulon-Var, LSIS, UMR CNRS 6168, B.P. 20132, 83957 La Garde Cedex, France
[b] IUT Dijon-Auxerre, LE2I, UMR CNRS 5158, Route des plaines de l'Yonne, 89000 Auxerre, France

ARTICLE INFO

Article history:
Received 5 October 2010
Received in revised form 31 January 2011
Accepted 14 March 2011
Available online 9 April 2011

Keywords:
Nonlinear observer and filter design
Wastewater treatment processes

ABSTRACT

The purpose of this paper is twofold: (1) we apply the adaptive observer developed in Boizot et al. [1] to a wastewater system, following two cascade steps. First, we apply it to a simplified model of the system. Second, we use this "simplified" estimation as a measurement for the full system. (2) Although the observability analysis is trivial, the equations contain rather complicated terms. Therefore, it is not reasonable to change coordinates for those of the required observability canonical form. Hence, we have to establish and work with the "unusual" equations of the observer in natural coordinates.

Let us point out that the simulations are done taking into account the small number of measurements (three) available in practice.

1. Introduction

The present work deals with the observer design of non linear dynamical systems, and application to a wastewater treatment system.

The need to develop observers or "software sensors" for Activated Sludge Processes in perspective of on-line monitoring is due to the following facts, among others:

(1) Although sensors for measuring chemical and biological variables are widespread and very advanced, such measurements are still unreliable and noisy.
(2) The implementation and maintenance costs of these advanced sensors are high.

A lot of work has been developed on the synthesis of nonlinear observers for (bio)chemical processes [2–13]. Here, we have chosen an adaptive high-gain observer as proposed in the paper [1] for the following reasons. This observer is high-gain, but it is also extended-Kalman-filter based: first, in the context of large transitions, it is an high-gain (HG) observer which guarantees theoretical convergence with arbitrary rate, under certain observability assumptions. Second, for small enough initial estimation error, it behaves like a classical extended Kalman filter (EKF), i.e. it is more or less optimal w.r.t. noise. Moreover, in a deterministic setting, it has good convergence properties [14].

Here, transition from HG mode to EKF mode is performed via an adaptation procedure based upon the level of innovation (i.e. the level of new information appearing through the "recent" observations).

For the general theory of high-gain nonlinear observers (see [15,1,7,16]).

The EKF is widely used and works rather well in practice. The main disadvantage for the EKF algorithm is that it requires an approximate knowledge of the initials conditions. Conversely, the HG-EKF algorithm converges whatever the initial guess but is rather sensitive with respect to noise. Then, the idea is to switch between the EKF and the HG-EKF algorithm. If the estimation error of the HG-EKF becomes sufficiently small then the EKF is used. The switching between these two modes can be done by having the high-gain parameter θ evolving between 1 and θ_{max}. The adaptation is made by using a differential equation driven by the "innovation".

Usually this method is applied by previously changing coordinates in order to put the system under a certain observability canonical form. In our case, we prefer to write our observer in the natural coordinates in order that it is not necessary to realize online the inverse coordinates change. The counterpart of this choice is that the Riccati equation of the Kalman filter has not the standard form. Detailed computations are provided in Appendix A.

Moreover here, in order to simplify the computations, we use cascade observers (reduced and complete): a first observer of the type above is used on a simplified model to provide an intermediate estimate of the state, this estimation being itself used as the output of the non simplified system.

Actually, for the complete observer with the three practical outputs, the computations are very heavy, even working in natural coordinates.

In Section 2 we recall the structure of our observer, which is just the multi-output version of the one developed in the paper [1]. Section 3 is devoted to the crucial concept of innovation, which

* Corresponding author: Tel.: +33 0 494142078.
E-mail address: lafont@univ-tln.fr (F. Lafont).

0959-1524/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2011.03.006

is used in order to switch between the EKF and HG-EKF modes. Section 4 presents in a few lines the idea of a cascade observer. Section 5 is devoted to the application to a wastewater treatment plant. First we recall the equations of the process, in full and simplified form. Then we perform the observability analysis in both cases. Thirdly we show noisy simulation results for the cascade observer.

2. Systems under consideration and observer equations

2.1. The observability canonical form

We consider a smooth nonlinear system of the form:

$$\begin{aligned} \frac{dx}{dt} &= f(x,u), \\ y &= h(x) = Cx, \end{aligned} \tag{1}$$

which is mapped by a diffeomorphism ψ into the following system:

$$\begin{aligned} \frac{d\xi}{dt} &= F(\xi,u) = A(t)\xi + b(\xi,u), \\ y &= C\xi, \end{aligned} \tag{2}$$

where $x, \xi \in \mathbb{R}^n$ are the state vectors, where u, the control variable belongs to a certain bounded subset of $\mathbb{R}^p$ and the output $y \in \mathbb{R}^{d_0}$.

Note: We have chosen to consider a linear output only, since it corresponds to our practical case and computations are simpler. However the general case is similar.

The matrices $A(t)$, C and the vector $b(\xi,u)$ have a following form:

$$\begin{aligned} A(t) &= \begin{pmatrix} 0 & a_2(t) & 0 & \cdots & 0 \\ 0 & 0 & a_3(t) & \ddots & \vdots \\ \vdots & \cdots & \ddots & \ddots & 0 \\ \vdots & \cdots & \cdots & 0 & a_k(t) \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}, \\ C &= (a_1(t), 0, \ldots, 0) = (Id, 0, \ldots, 0), \end{aligned} \tag{3}$$

where Id is the d_0 identity matrix.

$$b(\xi,u) = \begin{pmatrix} b_1(\xi_1,u) \\ b_2(\xi_1,\xi_2,u) \\ \vdots \\ b_n(\xi_1,\ldots,\xi_n,u) \end{pmatrix}. \tag{4}$$

The state vector $\xi(t)$ is assumed to have a "block" structure $\xi = (\xi_1' \xi_2' \cdots \xi_n')'$, where $\xi_i \in \mathbb{R}^{d_i}$ with $d_0 \geq d_1 \geq \cdots \geq d_k$. The matrices $a_i(t)$ have dimension $d_{i-1} \times d_i$ and belong to a compact subset K_i of the set of $d_{i-1} \times d_i$ matrices of maximum rank d_i. The $f(x,u)$, $a_i(t)$, $b_i(\xi,u)$ are assumed smooth w.r.t ξ, u and t, the b_i depend on ξ in a "block" triangular way and are compactly supported.

Along the paper x (resp. ξ) is called the **natural** coordinate (resp. the **observable** coordinate).

The structure conditions guarantee obviously "uniform" and "uniform infinitesimal" observability in the sense of Gauthier and Kupka [16]. The compact support conditions can be artificially forced outside the "practical" domain where the state is assumed to remain. All the results in Boizot et al. [1] extend without any difficulty to the case of such a structure with such "compact support" assumptions. It is just a matter of rewriting.

It follows from the observability theory in Gauthier and Kupka [16] that this canonical form together with the associated regularity assumptions is pertinent in several situations: For any system meeting strong observability assumptions, coordinates can be changed for "observable coordinates" in which this canonical form is met.

Along the paper TF denotes the tangent mapping to the mapping F: $x \to F(x)$, $\mathbb{R}^n \to \mathbb{R}^n$, i.e. its Jacobian matrix in coordinates. Accordingly T^2F denotes the double tangent, a skew-symmetric bilinear mapping, $\mathbb{R}^n$-valued, and for any $u \in \mathbb{R}^n$ we define the matrix $D^2F(x)\{u\}$ by $T^2F(u,v) = D^2F(x)\{u\} \cdot v$.

We denote by L_b the bound on the Jacobian matrix $Tb(\xi,u)$ of $b(\xi,u)$ (i.e. $||Tb(\xi,u)|| \leq L_b$). Since $b(\xi,u)$ is compactly supported and u is bounded, b is Lipschitz w.r.t. ξ uniformly in u: $||b(\xi,u) - b(\eta,u)|| \leq L_b||\xi - \eta||$.

2.2. Observer structure in observable coordinates

Let $Q(n \times n)$, $R(d_0 \times d_0)$ be symmetric positive definite matrices. Let θ be the high-gain parameter, $\theta \geq 1$. For $\theta = 1$ the observer will just be an ordinary EKF.

Set $\Delta = BD(1, 1/\theta, \ldots, 1/(\theta^{k-1}))$, the block diagonal matrix with diagonal blocks $Id_{d_0}, (1/\theta)Id_{d_1}, \ldots$. Set $Q_\theta = \theta\Delta^{-1}Q\Delta^{-1}$, $R_\theta = \theta^{-1}R$. The equations of the system in observable coordinates are:

$$\begin{aligned} \frac{d\xi}{dt} &= T\psi(\psi^{-1}(\xi))f(\psi^{-1}(\xi),u), \\ \frac{d\xi}{dt} &= F(\xi,u). \end{aligned} \tag{5}$$

$$y = C\xi. \tag{6}$$

The equations for the HG-EKF in the observable coordinates are:

$$\frac{d\hat{\xi}}{dt} = F(\hat{\xi},u) + PC'R_\theta^{-1}(y - C\hat{\xi}), \tag{7}$$

$$\frac{dP}{dt} = TF(\hat{\xi},u)P + PTF(\hat{\xi},u)' + Q_\theta - PC'R_\theta^{-1}CP. \tag{8}$$

In the natural coordinates we have $\hat{x} = \psi^{-1}(\hat{\xi}) = \Phi(\hat{\xi})$, where $\hat{x}$ denotes the estimate of x. In Appendix A, the equations for the HG-EKF become:

$$\frac{d\hat{x}}{dt} = f(\hat{x},u) + pC'(\hat{x},u)R_\theta^{-1}(y - h(\hat{x})), \tag{9}$$

$$\begin{aligned} \frac{dp}{dt} &= Tf(\hat{x},u)p + pTf(\hat{x},u)' + q_\theta(\hat{x}) - pC'R_\theta^{-1}Cp \\ &\quad + T\psi(\hat{x})^{-1}D^2\psi(\hat{x})\{pC'R_\theta^{-1}(h(\hat{x}) - y)\}p \\ &\quad + pD^2\psi(\hat{x})\{pC'R_\theta^{-1}(h(\hat{x}) - y)\}'(T\psi(\hat{x})^{-1})', \end{aligned} \tag{10}$$

where

$$q_\theta(\hat{x}) = (T\psi(\hat{x}))^{-1}Q_\theta((T\psi(\hat{x}))^{-1})'. \tag{11}$$

3. Innovation

The function In_d introduced below and called the innovation reflects the quality measurement of the estimation error on a small moving time interval of size d. The strategy is to adapt the high-gain parameter θ according to In_d. Due to the observability properties of our system, if the estimate $\hat{x}$ is far from x then θ will increase to high-gain mode. Contrarily, if $\hat{x}$ is close to x, innovation will be small and θ will decrease to 1 (Kalman filtering mode). For this, the variable θ will be subject to the differential equation (15) just below.

Let $G_o(\theta)$ be defined as follows:

$$G_o(\theta) = \begin{cases} \frac{1}{\Delta T}\theta^2 & \text{if } \theta \le \theta_1, \\ \frac{1}{\Delta T}(\theta - 2\theta_1)^2 & \text{if } \theta > \theta_1, \end{cases} \quad (12)$$

where $\theta_1 = (1/2)\theta_{max}$ and ΔT small enough is a constant.

The innovation $In_d(t)$, with forgetting horizon d, is:

$$In_d(t) = \int_{t-d}^{t} \|y(\tau) - \hat{y}(\tau)\|^2 \, d\tau, \quad (13)$$

where $\hat{y}(\tau)$ is the prediction from the initial state $\hat{x}(t-d)$.

Let us define

$$G(\theta, In_d) = \mu(In_d)G_o(\theta) + (1 - \mu(In_d))\lambda(1 - \theta), \quad (14)$$

for a $\lambda > 0$ and with $\mu(In_d)$ a smooth function equal to 1 if $In_d \ge \gamma_1$, to 0 if $In_d \le \gamma_0$, with $0 \le \mu(In_d) \le 1$ for $\gamma_0 \le In_d \le \gamma_1$. Another admissible choice for μ is a sigmoid function, $\mu:]0;+\infty[\to]0;1[$, $\mu(In_d) = 1/(1 + e^{-\beta\cdot(In_d - m)})$. The equation for the HG parameter θ is:

$$\dot{\theta} = G(\theta, In_d). \quad (15)$$

The parameters β and m of the sigmoid play the same role as the parameters γ_0 and γ_1. The zero value of the sigmoid function corresponds to "moving towards Kalman filtering mode" with maximum speed, although the value one corresponds to "moving towards the high gain mode" with maximum speed. The duration of the transition part is controlled by the parameter β (the higher β, the shorter the transition). In practice, the best results are obtained for a small transition time, i.e. a large value of β. All details can be found in Boizot et al. [1].

Finally our adaptive observer in original coordinates is given by the set of equations ((9), (10), (13) and (15)).

Comment 1: Roughly speaking, we can summarize the methodology as follows:

(1) A single Extended Kalman Filter equation, depending on a single parameter θ realizes the observers in both modes: for $\theta = 1$, it coincides with the ordinary Extended Kalman Filter. For θ large it is a HG-KF, as proposed for instance in Gauthier and Kupka [16].
(2) Guarantee of convergence of the error is obtained (see 5 below) in observable coordinates only. It is possible to overcome the difficulty of performing this coordinate change on-line, via the equations ((9)–(11)) of the transformed EKF equations to natural coordinates.
(3) The dynamics of the parameter θ is driven by the "innovation" term computed over a slipping window. Small innovation means that the estimation error is close to zero, hence, it is suitable to move θ to ordinary EKF mode. Conversely, large innovation means large estimation error, hence, the strategy is to move to high-gain mode. This is done via the "driving equation" (15).
(4) It is well known that the Riccati matrix P is related to the Gramm observability matrix of the linearized system along the estimate trajectory. Then it reflects the "innovation" relative to this linearized system. However, this "linearized innovation" is not enough for our purposes.
(5) **Guarantee of convergence:** In Boizot et al. [1], the convergence result is as follows: For all noise characteristics (Q, R) (depending on the noise in EKF mode) the parameters (θ, m, d, β) can be chosen in such a way that global arbitrary exponential convergence can be achieved: $\|\epsilon\| \le e^{-\alpha(t-T^*)} \times \|\epsilon_0\|$, ($\alpha$ arbitrary, T^* arbitrary).

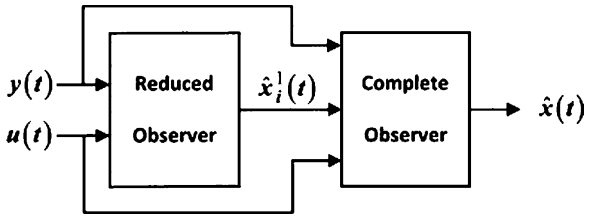


Fig. 1. The cascade observer.

Comment 2: Due to (13) the observer system ((9), (10), (13) and (15)) is not a system of ODE. However existence and uniqueness of solutions is guarantied and it is more or less clear how to proceed numerically.

4. The interest of natural coordinates, and of a cascade observer

It turns out that the change of variable $\psi(x)$ is not so easy to apply. It is the reason why we have chosen to work in natural coordinates. In these natural coordinates according to our observer equations, it is enough to be able to compute the inverse Jacobian $D\psi(\hat{x})^{-1}$. For our application below, this would be still hard in the case of the full equations. It is why we have chosen (in natural coordinates) the following strategy.

We apply first our observer to a simplified model (five states, three outputs). We use the estimate provided by this first observer as the output of the full system. In this way the computation of both inverse Jacobians is easy (see Fig. 1).

5. Application

The process under consideration is a real small-size wastewater treatment plant composed of a unique aeration tank equipped with surface aerators which provide oxygen and mix the influent wastewater with biomass (Fig. 2).

Here, we address the question of online estimation of the effluent quality.

A European Union directive fixed the maximum pollutant concentrations allowed in the effluent of small size wastewater treatment plants: The biochemical oxygen demand over an elapsed period of five days $BOD_5 < 25\ mg\,l^{-1}$, the chemical oxygen demand $COD < 125\ mg\,l^{-1}$ and the total suspended solid $TSS < 35\ mg\,l^{-1}$. These three quantities are defined below in terms of the state of the model.

The model used is based upon the Activated Sludge Model N[1] (ASM 1) from Henze et al. [17]. Then our biodegradation model consists of 12 state variables (Table 1): actually, we consider only biodegradation, the state variables describing the total alkalinity being not included.

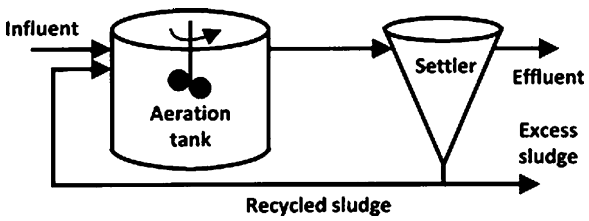


Fig. 2. Wastewater treatment plant.

Table 1
List of variables.

Definition	Notation
1. Soluble inert organic matter ($g\,COD\,m^{-3}$)	S_I
2. Readily biodegradable substrate ($g\,COD\,m^{-3}$)	S_S
3. Particulate inert organic matter ($g\,COD\,m^{-3}$)	X_I
4. Slowly biodegradable substrate ($g\,COD\,m^{-3}$)	X_S
5. Active heterotrophic biomass ($g\,COD\,m^{-3}$)	$X_{B,H}$
6. Active autotrophic biomass ($g\,COD\,m^{-3}$)	$X_{B,A}$
7. Particulate products arising from biomass decay ($g\,COD\,m^{-3}$)	X_P
8. Oxygen ($g\,COD\,m^{-3}$)	S_O
9. Nitrate and nitrite nitrogen ($g\,N\,m^{-3}$)	S_{NO}
10. $NH_4^+ + NH_3$ nitrogen ($g\,N\,m^{-3}$)	S_{NH}
11. Soluble biodegradable organic nitrogen ($g\,N\,m^{-3}$)	S_{ND}
12. Particulate biodegradable organic nitrogen ($g\,N\,m^{-3}$)	X_{ND}

Table 2
Stoichiometric parameters.

Parameter	Unit	Value
Y_A	g cell COD formed $(g\,N\,oxidized)^{-1}$	0.24
Y_H	g cell COD formed $(g\,COD\,oxidized)^{-1}$	0.67
f_p	Dimensionless	0.08
i_{XB}	$g\,N\,(g\,COD)^{-1}$ in biomass	0.08
i_{XP}	$g\,N\,(g\,COD)^{-1}$ in particulate products	0.06

The three quality requirements characterizing the effluent are defined by:

$$\begin{aligned} BOD_5 &= 0.25(S_S + X_S + (1-f_p)(X_{BH}+X_{BA})),\\ COD &= S_S + S_I + X_S + X_I + X_{BH} + X_{BA} + X_P,\\ TSS &= 0.75(X_S + X_I + X_{BH} + X_{BA} + X_P). \end{aligned} \tag{16}$$

Remark 1. The stoichiometric and kinetic parameter values considered are listed in Tables 2 and 3. The complete set of equations and influent conditions can be found on the International Water Association task group on benchmarking of control strategies for wastewater treatment plants website (http://www.benchmarkwwtp.org/, 2010).

Remark 2. Regarding the simplified model, in the paper, for the benefit of the reader, we provide **all explicit formulas and values** of the constants and kinetic functions.

The model assumptions are the following:

- the reactor is well mixed,
- the separation of liquid and solid phases is perfect and no reaction occurs in the settler,
- the sum of all settler flowrates equals the settler influent flowrate.

Our model is of the form $\dot{x} = f(x,u)$, where the control u consists of the state u_b of the turbines and the value Q^{in} of the influent average flow. The input u_b in (19) is a binary sequence switching between 0 and 1 and representing the state of turbines (off/on) that aerate the plant.

Table 3
Kinetic parameters.

Parameter	Unit	Value
μ_H	d^{-1}	4.0
K_S	$g\,COD\,m^{-3}$	10.0
$K_{O,H}$	$g\,COD\,m^{-3}$	0.2
K_{NO}	$g\,NO_3 - N\,m^{-3}$	0.5
b_H	d^{-1}	0.3
$\eta_{NO,g}$	Dimensionless	0.8
$\eta_{NO,h}$	Dimensionless	0.8
k_h	$(g\,cell\,COD\,d)^{-1}$	3.0
K_X	$(g\,cell\,COD)^{-1}$	0.1
μ_A	d^{-1}	0.5
$K_{NH,A}$	$g\,NH_3 - N\,m^{-3}$	1.0
b_A	d^{-1}	0.05
$K_{O,A}$	$g\,COD\,m^{-3}$	0.4
k_a	$m^3\,(g\,COD\,d)^{-1}$	0.05

Natural coordinates are concentrations of the species, i.e. all components x_i of the state vector are the concentrations listed in Table 1. Each equation has the type of a material balance, including kinetic degradation, then the components f_i of the dynamics are as follows:

- For soluble components ($i = 1, 2, 9, 10, 11$)

$$f_i(x) = \frac{Q^{in}}{V}(x_i^{in} - x_i) + r_i(x). \tag{17}$$

- For particulate components ($i = 3, 4, 5, 6, 7, 12$)

$$f_i(x) = \frac{1}{V}\left[Q^{in}(x_i^{in} - x_i) + Q^{rs}\frac{Q^{in}-Q^w}{Q^{rs}+Q^w}x_i\right] + r_i(x). \tag{18}$$

- For dissolved oxygen concentration ($i = 8$)

$$f_8(x) = \frac{Q^{in}}{V}(x_8^{in} - x_8) + r_8(x) + u_b k_L a(S_O^{max} - S_O) \tag{19}$$

where $r_i(x)$, $i = 1, \ldots, 12$ are nonlinear functions not given here (see [17]). They represent the apparent reaction rates depending on the kinetic rates of degradation of the components.

Remark 3. The variables S_I, X_I and X_P, related with the equations corresponding to $i = 1, 3, 7$, do not appear in the other equations. Hence these variables are not observables, and we cannot do better for them than **simple prediction**. Therefore, pertinent dimension of the state space is $n = 9$.

The constant $k_L a$ is the oxygen transfer coefficient ($k_L a = 10\,h^{-1}$) and S_O^{max} is the dissolved oxygen saturation concentration ($S_O^{max} = 8\,mg\,l^{-1}$).

The volume of the aeration tank is ($V = 6000\,m^3$). The settler is a cylindrical tank where the solids are either recirculated to the aeration tank ($Q^{rs} = 18446\,m^3\,day^{-1}$) or extracted from the system ($Q^w = 385\,m^3\,day^{-1}$).

We make here the reasonable assumption of three measurements only: S_O, S_{NO} and S_{NH}, located inside the aeration tank. Although the WWTP with these three outputs is observable, it is too complicated for our purpose. We use first a simplified model of lower dimension that has been developed in Chachuat et al. [8].

5.1. The reduced model

The author in Chachuat et al. [8] proceeds as follows: (1) he regroups the species S_S and X_S into a single one X_{COD} (COD for "chemical oxygen demand"), $X_{COD} = S_S + X_S$. 2. It is known that the dynamics of X_{BH}, X_{BA}, X_{ND} are slow w.r.t. the other. Then, they are assumed to be constant. Hence the variables α_i, $i = 1, \ldots, 8$ defined below are constant. It is also commonly accepted that the ratios X_{ND}/X_S, X_{COD}/S_S, X_{COD}/X_S vary slowly. As a consequence the variables α_9, K_{COD}, K_{ND} below are also assumed constant. Removing the three unobservable variables X_P, X_I, S_I leads to a simplified model with 5 state variables S_O, S_{NO}, S_{NH}, X_{COD}, S_{ND} with the three observable variables S_O, S_{NO}, S_{NH}. All these simplifications provide

Table 4
Constant coefficients.

Coefficient	Value
α_1	−5892
α_2	−875
α_3	−1648
α_4	191
α_5	−957
α_6	150
α_7	−17,855
α_8	830
α_9	561
K_{COD}	574
K_{ND}	296

the following reduced model:

$$\dot{S_O} = \frac{Q^{in}}{V}(S_O^{in} - S_O) + \alpha_1 \frac{X_{COD}}{K_{COD} + X_{COD}} \cdot \frac{S_O}{K_{O,H} + S_O} + \tilde{r}_1(y) + u_b k_L a(S_O^{max} - S_O) \tag{20}$$

$$\dot{S_{NO}} = \frac{Q^{in}}{V}(S_{NO}^{in} - S_{NO}) + \alpha_3 \frac{X_{COD}}{K_{COD} + X_{COD}} \cdot \frac{K_{O,H}}{K_{O,H} + S_O} \frac{S_{NO}}{K_{NO} + S_{NO}} + \tilde{r}_2(y) \tag{21}$$

$$\dot{S_{NH}} = \frac{Q^{in}}{V}(S_{NH}^{in} - S_{NH}) + \alpha_5 \frac{X_{COD}}{K_{COD} + X_{COD}} \cdot \left(\frac{S_O}{K_{O,H} + S_O} + \eta_{NO,g} \frac{K_{O,H}}{K_{O,H} + S_O} \frac{S_{NO}}{K_{NO} + S_{NO}} \right) + \tilde{r}_3(y) + \alpha_6 S_{ND} \tag{22}$$

$$\dot{X_{COD}} = \frac{Q^{in}}{V}\left(X_{COD}^{in} - \frac{K_S}{K_{COD}} X_{COD}\right) + \alpha_7 \frac{X_{COD}}{K_{COD} + X_{COD}} \times \left(\frac{S_O}{K_{O,H} + S_O} + \eta_{NO,g} \frac{K_{O,H}}{K_{O,H} + S_O} \frac{S_{NO}}{K_{NO} + S_{NO}} \right) + \alpha_8 \tag{23}$$

$$\dot{S_{ND}} = \frac{Q^{in}}{V}(S_{ND}^{in} - S_{ND}) - \alpha_6 S_{ND} + \alpha_9 \cdot \frac{X_{COD}}{K_{ND} + X_{COD}} \times \left(\frac{S_O}{K_{O,H} + S_O} + \eta_{NO,h} \frac{K_{O,H}}{K_{O,H} + S_O} \cdot \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \tag{24}$$

The parameters α_1, α_2, α_3, α_4, α_5, α_6, α_7, α_8, α_9, K_{ND} and K_{COD} are defined as follows, their value is given in Table 4, the values of the influent concentrations being listed in Table 5.

$$\begin{aligned}
\alpha_1 &= -\frac{1 - Y_H}{Y_H} \mu_H X_{B,H} \\
\alpha_2 &= -4.57 \frac{\mu_A}{Y_A} X_{B,A} \\
\alpha_3 &= -\frac{1 - Y_H}{2.86 Y_H} \mu_H X_{B,H} \eta_{NO,g} \\
\alpha_4 &= \frac{\mu_A}{Y_A} X_{B,A} \\
\alpha_5 &= -i_{XB} \mu_H X_{B,H} \\
\alpha_6 &= k_a X_{B,H} \\
\alpha_7 &= -\frac{1}{Y_H} \mu_H X_{B,H} \\
\alpha_8 &= (1 - f_p)(b_H X_{B,H} + b_A X_{B,A}) \\
\alpha_9 &= k_h \frac{X_{ND}}{X_S} X_{B,H}
\end{aligned} \tag{25}$$

$$\begin{aligned}
K_{COD} &= K_S \frac{X_{COD}}{S_S} \\
K_{ND} &= K_X \frac{X_{COD}}{X_S} X_{B,H}
\end{aligned} \tag{26}$$

$$\begin{aligned}
\tilde{r}_1(y) &= \alpha_2 \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_O}{K_{O,A} + S_O} \\
\tilde{r}_2(y) &= \alpha_4 \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_O}{K_{O,A} + S_O} \\
\tilde{r}_3(y) &= -\alpha_4 \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_O}{K_{O,A} + S_O}
\end{aligned} \tag{27}$$

Table 5
Influent concentrations.

Concentration	Value
S_I^{in}	30 g COD m^{-3}
S_S^{in}	69.5 g COD m^{-3}
X_I^{in}	51.2 g COD m^{-3}
X_S^{in}	202.32 g COD m^{-3}
X_{BH}^{in}	28.17 g COD m^{-3}
X_{BA}^{in}	0 g COD m^{-3}
X_P^{in}	0 g COD m^{-3}
S_O^{in}	0 g COD m^{-3}
S_{NO}^{in}	0 g COD m^{-3}
S_{NH}^{in}	31.56 g COD m^{-3}
S_{ND}^{in}	6.95 g COD m^{-3}
X_{ND}^{in}	10.59 g COD m^{-3}
X_{COD}^{in}	271.82 g COD m^{-3}

5.1.1. Observability analysis

For the simplified model ((20)–(24)), with outputs S_O, S_{NO}, S_{NH} the "practical" domain is $(\mathbb{R}_+)^5$, the positive orthant in $(\mathbb{R})^5$.

1. The points where S_O, S_{NO} are both zero may appear in practice, this is called "anaerobic behavior". This type of functioning remains nevertheless temporary and is not very frequent since the switch off period of the aerator is limited by the operating constraint $t_{max}^{off} = 120\,min$. In that case, the variable X_{COD} has no influence on the outputs. Therefore the system is not observable in any sense and there is nothing better to do than simple prediction.
2. On the subdomain $\mathfrak{D} \subset (\mathbb{R}_+)^5$, $\mathfrak{D} = \{S_O \neq 0 \text{ or } S_{NO} \neq 0\}$ it is easily computed that the system is uniformly observable and uniformly infinitesimally observable in the sense of Gauthier and Kupka [16]. This is reflected by the fact that the matrix $a_2(t)$ has rank two on $\mathfrak{D}$ (remember that a_2 is a function of t via is dependence on the output variables).

This is enough for the high-gain theory works, and in particular our adaptive algorithm developed in Boizot et al. [1].

5.1.2. Change of variables

The change of variables Ψ that relates natural coordinates to observer coordinates is trivial: it consists of setting just

$$\tilde{X}_{COD} = \frac{X_{COD}}{K_{COD} + X_{COD}}. \tag{28}$$

The state vector $x = (S_O\, S_{NO}\, S_{NH}\, X_{COD}\, S_{ND})'$ is changed for $\xi = (S_O\, S_{NO}\, S_{NH}\, \tilde{X}_{COD}\, S_{ND})'$, therefore our system is almost naturally in observable coordinates. The inverse Jacobian is trivial to compute.

Table 6
Parameters for the adaptation.

Parameter	Value of the reduced observer	Value of the complete observer
θ_{max}	20	10
β	$1664\frac{\pi}{e}$	$1664\frac{\pi}{e}$
m	2	40
ΔT	0.01	0.01
λ	200	200
d	0.1	0.1

5.2. Observer for the complete model

The observability analysis of the full system with the estimate provided by the reduced observer is trivial (after forgetting about the unobservable variables S_I, X_I and X_P, see Remark 3). It leads to similar conclusions of uniform observability and uniform infinitesimal observability.

In that case the state is 9-dimensional and the output is 6-dimensional: Actually the variables X_S, S_S that have been glued together in $X_{COD} = S_S + X_S$, can be splitted out from the reduced model. This is done using the previous assumption that K_{COD} is a constant (26):

$$\frac{X_{COD}}{S_S} = 1 + \frac{X_S}{S_S} = \frac{K_{COD}}{K_S}. \tag{29}$$

5.2.1. Change of variables

The change of variables ψ from natural to observable coordinates is trivial:

$x = (S_O\, S_{NO}\, S_{NH}\, S_S\, X_S\, S_{ND}\, X_{BH}\, X_{BA}\, X_{ND})'$ is changed for
$\xi = (S_O\, S_{NO}\, S_{NH}\, S_S\, X_S\, S_{ND}\, r_8\, r_9\, r_{11})'$.

Of course the functions $r_8\, r_9\, r_{11}$ are such that the change of variables is an embedding.

5.3. Choice of the parameters related to innovation

The choice of the parameters (θ_{max}, β, m, ΔT, λ, d) in the case of our application is given in Table 6.

Remember that the purpose of these parameters is to tune the way the high-gain evolves between 1 (EKF mode) and θ_{max} (HG mode). This choice has been obtained just by successive trials.

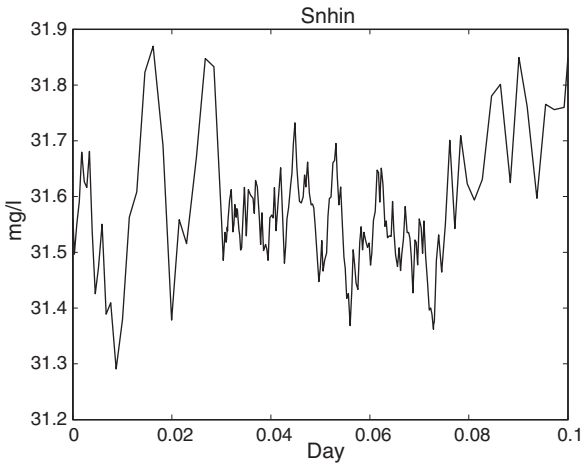


Fig. 3. Variation of S_{NH}^{in}.

5.4. Conditions for a realistic simulation

5.4.1. Input concentrations

In order to perform realistic simulations, the influent concentrations ($S_I^{in}, S_S^{in}, \ldots, X_{COD}^{in}$) (typical values given in Table 5) cannot be considered as constant. We have modelled the variations of these concentrations by an additive noise. In practice, due the length of the feeding pipe (maybe several kilometers), these perturbations should appear rather slowly. However, we have willingly chosen fast dynamics for these noises. An illustrative example of these variations is shown in Fig. 3.

5.4.2. Desadaptation of kinetic rates and stoichiometric coefficients

These parameters are not very well known in practice, and may be subject to large unexpected variation. We have considered simultaneously, for each reaction rate (theoretically given by Henze et al. [17]) a periodic desadaptation of amplitude 20%. Moreover, these desadaptations are realized in a completely asynchrone way.

Here, we consider 3 periods over the 14 days under consideration, with a phase difference uniformly displayed over the 8 reaction rates.

These conditions of simulation 5.4.1 and 5.4.2 above are presumably worse than what may appear in practice.

Table 7
Comparisons between Luenberger, EKF and adaptive HG-EKF.

Variable	Luenberger		EKF		HG-EKF		Range
	m	σ	m	σ	m	σ	
S_{ND}	−0.02	0.07	−0.02	0.07	−0.01	0.08	[0.5–1.2]
S_S	−0.03	0.11	−0.03	0.11	−0.02	0.10	[0.6–1.7]
X_S	0.30	7.85	0.28	7.86	−0.21	7.59	[38.3–110.0]
X_{BH}	96.17	81.39	96.68	81.29	41.64	85.52	[2635.3–3670.7]
X_{BA}	−8.29	4.50	−7.20	4.41	−5.89	3.14	[75.0–154.2]
X_{ND}	0.02	0.61	0.02	0.61	−0.02	0.60	[2.7–7.0]

Table 8
Comparisons between Luenberger, EKF and adaptive HG-EKF.

Variable	Luenberger		EKF		HG-EKF		Range
	m	σ	m	σ	m	σ	
S_I	−0.03	0.12	−0.03	0.12	−0.03	0.12	[29.9–30.1]
X_I	3.93	1.89	3.94	1.88	3.76	1.84	[1224.7–1259.2]
X_P	28.48	8.42	28.51	8.40	20.29	7.82	[270.0–510.4]

Table 9
Quality requirements.

Variable	Luenberger		EKF		HG-EKF		Range
	m	σ	m	σ	m	σ	
BOD_5	0.04	0.04	0.04	0.04	0.01	0.04	[0–2.2]
COD	0.22	0.15	0.22	0.15	0.09	0.15	[0–30.1]
TSS	0.18	0.13	0.19	0.13	0.09	0.10	[0–9.2]

6. Results

As commonly accepted, all simulations shown in this section are done with the outputs perturbed by a realistic additive Orstein–Uhlenbeck process. The alternative control u_b has been chosen as in practice: "On" during 15 min and "off" during 5 min. Our simulation file (dry weather) covers 14 days and the value of the input flow rate Q^{in} come from the benchmark file (http://www.benchmarkwwtp.org/, 2010).

To evaluate the performances of our observer on the WWTP, we compare a Luenberger observer, an ordinary EKF, and our adaptive HG-EKF. No comparison is shown with an ordinary HG-EKF (non adaptive): in that case the results are rather bad, the observer being very sensitive to noise.

The averages and standard deviations are computed over the whole duration of 14 days. However the illustrative figures presented below show the 3 first days only, where the effect of the unknown initial conditions is significant.

Table 7 shows a clear improvement, for our adaptive HG observer.

6.1. Reconstruction of the variables X_I, S_I, X_P

As we said these unobservable variables are reconstructed by simple prediction. The results are shown in Table 8.

6.2. Effluent quality

To validate the method and estimate the effluent outputs, we simulate the complete settler as described in Takacs et al. [18]. This model simulates the solids profile throughout the settling column, including the underflow and effluent suspended solid concentrations. Comparisons of the three quality requirements with their estimates are presented in Table 9. In this table we show again the average and the standard deviation of the estimation error.

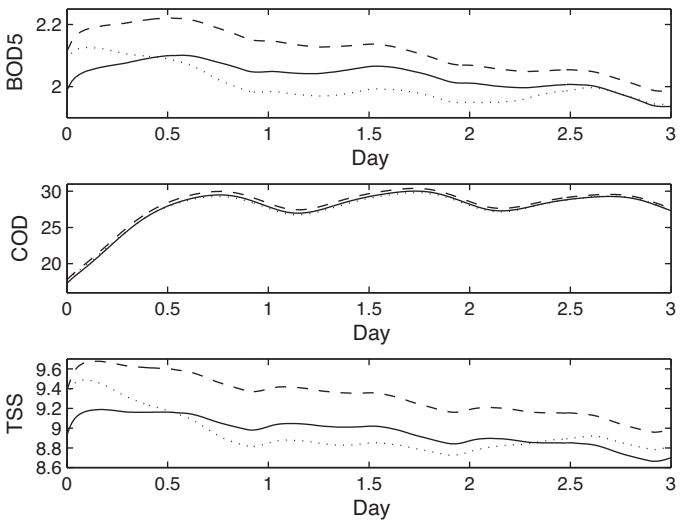


Fig. 4. Effluent quality (continuous: model – dashed: EKF observer – dotted: HG-EKF observer).

Fig. 4 displays the output variables BOD_5, COD, TSS and their estimates, over 3 days. The effect of the high gain at the beginning of the response is very clear. The error really converges quickly to zero, which is not the case for the EKF, for which a significant error remains for long.

7. Conclusion

The method proposed here for state reconstruction of a WWTP seems to be a real improvement with respect to classical methods. It is technically twofold: First the implementation of the adaptive HG-EKF is not too complicated due to the use of natural coordinates which simplifies hugely the computations. Second the use of a cascade observer also leads to reasonable computations for the complete model.

Note that here we have studied only the case of a 20 °C influent temperature (a summer scenario). The kinetic and stoichiometric parameters of the models can be very different for a winter scenario. More generally a possible improvement could be a multi-model strategy taking into account the exterior temperature.

Acknowledgment

This work is supported by ANR program BLANC GCM (Geometric Control methods, Sub-Riemannian geometry and application) and INRIA project GECO.

Appendix A. Change of variables in the HG-EKF

We do the computations in the case of our two applications only, that have special features w.r.t the general cases where the theory applies:

1. The change of variables is of the form $x = \Phi(\xi)$, where x is the original coordinate and ξ is the "observable coordinate". In general it is not the case, but $\xi = \psi(x, u)$.
2. The output consists of the first state coordinates i.e. $C = (I_d, 0)$, $y = Cx = C\xi$, $CT\Phi(\xi) = C$.

Also, in the computations below, equations are $\dot{x} = f(x, u(t))$ and $\dot{\xi} = F(\xi, u(t))$. But we omit the dependence in t and we use $\dot{x} = f(x)$, $\dot{\xi} = F(\xi)$. This has no consequence in the computations. Also, the fact that Q_θ, R_θ depend on θ that itself depends on t has no consequence. In fact, the matrix $A(t)$ above depends on t via a dependence on the outputs S_O, S_{NO} and S_{NH}.

We set $\dot{\xi} = F(\xi, t)$, $x = \Phi(\xi)$, $\hat{x} = \Phi(\hat{\xi})$, then, $T\Phi \circ \Phi^{-1}(x) F(\Phi^{-1}(x), t) = f(x, t)$. Or, dropping t from now on: $f \circ \Phi(\xi) = T\Phi(\xi) F(\xi)$.

It follows that:

$$Tf(\Phi(\xi))T\Phi(\xi) = D^2\Phi(\xi)(F(\xi)) + T\Phi(\xi)TF(\xi).$$

Hence:

$$Tf(x) = D^2\Phi(\xi)\{F(\xi)\}T\Phi(\xi)^{-1} + T\Phi(\xi)TF(\xi)T\Phi(\xi)^{-1} \quad \text{if } x = \Phi(\xi). \tag{30}$$

The equations of the HG-EKF are:

$$\begin{aligned} \dot{\hat{\xi}} &= F(\hat{\xi}) + PC'R_\theta^{-1}(y - C\hat{\xi}) \\ \dot{\hat{x}} &= T\Phi(\hat{\xi})\dot{\hat{\xi}} = T\Phi(\hat{\xi})F(\hat{\xi}) + T\Phi(\hat{\xi})P\,T\Phi(\hat{\xi})'(T\Phi(\hat{\xi})')^{-1} \\ &\quad \cdot C'R_\theta^{-1}(y - C\hat{\xi}). \end{aligned} \tag{31}$$

Setting $p = T\Phi(\hat{\xi})PT\Phi(\hat{\xi})'$,

$$\dot{\hat{x}} = f(\hat{x}) + pC'R_\theta^{-1}(y - C\hat{\xi}). \tag{32}$$

The equation for P is:

$$\dot{P} = TF(\hat{\xi})P + PTF(\hat{\xi})' + Q_\theta - PC'R_\theta^{-1}CP, \tag{33}$$

which produces:

$$\dot{p} = T\Phi(\hat{\xi})\dot{P}T\Phi(\hat{\xi})' + \overbrace{T\Phi(\hat{\xi})}^{\cdot}PT\Phi(\hat{\xi})' + T\Phi(\hat{\xi})P\overbrace{T\Phi(\hat{\xi})}^{\cdot\,\prime}, \tag{34}$$

$$\begin{aligned} \dot{p} &= T\Phi(\hat{\xi})TF(\hat{\xi})T\Phi(\hat{\xi})^{-1}p \\ &\quad + p(T\Phi(\hat{\xi})TF(\hat{\xi})T\Phi(\hat{\xi})^{-1})' \\ &\quad + T\Phi(\hat{\xi})Q_\theta T\Phi(\hat{\xi})' - pC'R_\theta^{-1}Cp \\ &\quad + \overbrace{T\Phi(\hat{\xi})}^{\cdot}PT\Phi(\hat{\xi})' + T\Phi(\hat{\xi})P\overbrace{T\Phi(\hat{\xi})}^{\cdot\,\prime}. \end{aligned} \tag{35}$$

By (30):

$$\begin{aligned} \dot{p} &= (Tf(\hat{\xi}) - D^2\Phi(\hat{\xi})\{F(\hat{\xi})\}T\Phi(\hat{\xi})^{-1})p \\ &\quad + p(Tf(\hat{\xi}) - D^2\Phi(\hat{\xi})\{F(\hat{\xi})\}T\Phi(\hat{\xi})^{-1})' \\ &\quad + q - pC^TR_\theta^{-1}Cp \\ &\quad + \overbrace{T\Phi(\hat{\xi})}^{\cdot}PT\Phi(\hat{\xi})' + T\Phi(\hat{\xi})P\overbrace{T\Phi(\hat{\xi})}^{\cdot\,\prime}, \end{aligned} \tag{36}$$

where $q = T\Phi(\hat{\xi})Q_\theta T\Phi(\hat{\xi})'$, or:

$$\begin{aligned} \dot{p} &= Tf(\hat{x})p + pTF(\hat{x})' + q - pC^TR_\theta^{-1}Cp \\ &\quad - D^2\Phi(\hat{\xi})\{F(\hat{\xi})\}T\Phi(\hat{\xi})^{-1}p \\ &\quad - p(D^2\Phi(\hat{\xi})\{F(\hat{\xi})\}T\Phi(\hat{\xi})^{-1})' \\ &\quad + \overbrace{T\Phi(\hat{\xi})}^{\cdot}T\Phi(\hat{\xi})^{-1}p + p(T\Phi(\hat{\xi})^{-1})'\overbrace{T\Phi(\hat{\xi})}^{\cdot\,\prime}. \end{aligned} \tag{37}$$

$$\begin{cases} \dot{p} &= (I) + (II) + (III), \\ (I) &= Tf(\hat{x})p + p(TF(\hat{x}))' + q - pC^TR_\theta^{-1}Cp, \\ (II) &= (\overbrace{T\Phi(\hat{\xi})}^{\cdot} - D^2\Phi(\hat{\xi})\{F(\hat{\xi})\}) \cdot T\Phi(\hat{\xi})^{-1}p, \\ (III) &= (II)'. \end{cases} \tag{38}$$

Now, let us compute $\overbrace{T\Phi(\hat{\xi})}^{\cdot}$:

$$\begin{aligned} \overbrace{T\Phi(\hat{\xi})}^{\cdot} &= D^2\Phi(\hat{\xi})\{\dot{\hat{\xi}}\} \\ &= D^2\Phi(\hat{\xi})\{F(\hat{\xi}) + PC'R_\theta^{-1}(y - C\hat{\xi})\}. \end{aligned} \tag{39}$$

We use now the following formula, just coming from the fact that $0 = T^2(\Phi^{-1} \circ \Phi(\xi))$:

$$0 = T^2\Phi^{-1} \circ \Phi(\xi)(T\Phi(\xi)u, T\Phi(\xi)v) + T\Phi^{-1}(x)T^2\Phi(\xi)(u, v). \tag{40}$$

Or

$$D^2\Phi(\xi)\{u\} = -T\Phi(x)D^2\Phi^{-1}(x)\{T\Phi^{-1}(x)^{-1}u\}T\Phi(\xi). \tag{41}$$

Putting (39) in (38) gives:

$$(II) = D^2\Phi(\xi)\{PC'R_\theta^{-1}(y - C\hat{\xi})\} \cdot T\Phi(\hat{\xi})^{-1}p, \tag{42}$$

and using (41),

$$(II) = -T\Phi(\hat{x})D^2\Phi^{-1}(\hat{x})\{T\Phi(\hat{\xi})PT\Phi(\hat{\xi})'C'R_\theta^{-1}(y - C\hat{x})\}p, \tag{43}$$

$$(II) = -T\Phi(\hat{x})D^2\Phi^{-1}(\hat{x})\{pC'R_\theta^{-1}(y - C\hat{x})\}p. \tag{44}$$

Going back to (38), we get:

$$\begin{aligned} \dot{p} &= Tf(\hat{x})p + pTf(\hat{x})' + q - pC^TR_\theta^{-1}Cp \\ &\quad + T\Phi(\hat{x})D^2\Phi^{-1}(\hat{x})\{pC'R_\theta^{-1}(C\hat{x} - y)\}p \\ &\quad + p(D^2\Phi^{-1}(\hat{x})\{pC'R_\theta^{-1}(C\hat{x} - y)\})'T\Phi(\hat{x})'. \end{aligned} \tag{45}$$

References

[1] N. Boizot, E. Busvelle, J.-P. Gauthier, An adaptive high-gain observer for nonlinear systems, Automatica 469 (2010) 1483–1488.
[2] V. Alcaraz-Gonzalez, J. Harmand, A. Rapaport, J.-P. Steyer, V. Gonzalez-Alvarez, C. Pelayo-Ortiz, Software sensors for highly uncertain wwtps: a new approach based on interval observers, Water Research 36 (2002) 2515–2524.
[3] A. Assis, R. Filho, Soft sensors development for on-line bioreactor state estimation, Computers and Chemical Engineering 24 (2000) 1099–1103.
[4] G. Bastin, D. Dochain, On-line Estimation and Adaptive Control of Bioreactors, Elsevier Science Publishers, 1990.
[5] G. Bastin, J.V. Impe, Nonlinear and adaptive control in biotechnology: a tutorial, European Journal of Control 1.1 (1995) 37–53.
[6] E. Busvelle, J.-P. Gauthier, On determining unknown functions in differential systems, with an application to biological reactors., ESAIM: Control, Optimisation and Calculus of Variations 9 (2003) 509–551.
[7] E. Busvelle, J.-P. Gauthier, Observation and identification tools for nonlinear systems: application to a fluid catalytic cracker, International Journal of Control 78.3 (2005) 208–234.
[8] B. Chachuat, N. Roche, M. Latifi, Reduction of the ASM1 model for optimal control of small-size activated sludge treatment plants, Journal of Water Science/Revue des Sciences de l'eau 16.1 (2003) 05–26.
[9] D. Dochain, Bioprocess Control, ISTE, 2008, ISBN 9781848210257.
[10] J.-P. Gauthier, H. Hammouri, S. Othman, A simple observer for nonlinear systems. application to bioreactors, IEEE Transactions on Automatic Control 37.6 (1992) 875–880.
[11] F. Nejjari, V. Puig, L. Giancristofaro, S. Koehler, Extended Luenberger observer-based fault detection for an activated sludge process, in: Proceedings of the 17th World Congress The International Federation of Automatic Control, Seoul, Korea, 2008, pp. 9725–9730.
[12] A. Rapaport, D. Dochain, Interval observers for biochemical processes with uncertain kinetics and inputs, Mathematical Biosciences 193 (2005) 235–253.
[13] O. Sotomayor, S. Park, C. Garcia, Software sensor for on-line estimation of the microbial activity in activated sludge systems, ISA Transactions 41 (2002) 127–143.
[14] J. Baras, A. Bensoussan, M. James, Dynamic observers as asymptotic limits of recursive filters: special cases, SIAM Journal on Applied Mathematics 48.5 (1988) 1147–1158.
[15] N. Boizot, E. Busvelle, Adaptive-gain observers and applications in nonlinear observers and applications, in: G. Besanon (Ed.), Lecture Notes in Control and Information Sciences, vol. 363, Springer-Verlag, 2007.
[16] J.-P. Gauthier, I. Kupka, Deterministic Observation Theory and Applications, Cambridge University Press, Cambridge, 2001.
[17] M. Henze, C. Grady, W. Gujer, G. Marais, T. Matsuo, Activated sludge model n1, in: IAWQ (Ed.), Technical Report 1, London, 1987.
[18] I. Takacs, G. Patry, D. Nolasco, A dynamic model of the clarification thickening process, Water Research 25.10 (1991) 1263–1271.

5.7 Observateur à entrées inconnues pour le diagnostic

International Journal of Control
Vol. 84, No. 4, April 2011, 822–833

Taylor & Francis
Taylor & Francis Group

Sensor fault reconstruction and observability for unknown inputs, with an application to wastewater treatment plants

Salowa Methnani[ab]*, Jean-Paul Gauthier[b] and Frederic Lafont[b]

[a]*ENIS, Electrical Engineering Department, Control Unit of Industrial Processes, B.P W, 3038 Sfax, Tunisia;* [b]*Université du Sud-Toulon-Var, LSIS, UMR CNRS 6168, B.P 20132, 83957 La Garde Cedex, France*

(*Received 4 November 2010; final version received 14 April 2011*)

In this article, we propose a general methodology for identifying and reconstructing sensor faults on dynamical processes. This methodology is issued from the general identification theory developed in the previous papers (Busvelle, E., and Gauthier, J.-P. (2003), 'On Determining Unknown Functions in Differential Systems, with an Application to Biological Reactor', *ESAIM: Control, Optimisation and Calculus of Variations*, 9, 509–553; Busvelle, E., and Gauthier, J.-P. (2004), 'New Results on Identifiability of Nonlinear Systems', in *2nd Symposium on Systems, Structure and Control*, Oaxaca, Mexico; Busvelle, E., and Gauthier, J.-P. (2005), 'Observation and Identification Tools for Non Linear Systems. Application to a Fluid Catalytic Cracker', *International Journal of Control*, 78, 208–234): in fact, this identification theory also provides a general framework for the problem of 'observability with unknown inputs'. Indeed, many problems of fault detection can be formulated as such observability problems, the (eventually additive) faults being just considered as unknown inputs. Our application to 'sensor fault detection' for wastewater treatment plants (WWTP) constitutes an ideal academic context to apply the theory: first, in this 3-5 case (3 sensors, 5 states), the theory applies generically and, second, any system is naturally under the 'observability canonical form' required to apply the basic high-gain observer from Gauthier and Kupka (Gauthier, J.-P., and Kupka, I. (1994), 'Observability and Observers for Nonlinear Systems', *SIAM Journal on Control*, 32, 975–994). A simulation study on the Bleesbrük WWTP is proposed to show the effectiveness of this approach.

Keywords: sensor fault detection; high-gain observers; fault reconstruction; observers with unknown inputs; wastewater treatment plants

1. Introduction

State estimation and fault detection and isolation (FDI) constitute the purpose of this article. The main purpose of an FDI scheme is not only to detect the fault when it occurs, by generating an alarm, but also by identifying the nature and the location of the fault. A fault is a malfunction of actuators or sensors, or more generally of internal state variables of the system. These malfunctions occur due to certain abnormal circumstance. If unchecked, such an unallowable deviation of at least one characteristic property or variable from its acceptable range may be devastating (Isermann and Ball 1996; Palade and Bocaniala 2010; Isermann 2011). Various FDI approaches have been proposed (Frank 1990, 1996; Patton and Chen 1993). Others methods based on computational intelligence techniques can be found in Palade and Bocaniala (2010). In Isermann (2011), several model-based methods are defined and developed: fault detection with parameter estimation, with parity equations, with state observers and state estimation.

In general, the FDI methods do not always afford the shape, the magnitude of the time-dependent fault.

Among these approaches, observer-based FDI attract a great deal of attention from the research community (Yang and Saif 1995; Frank and Ding 1997; Chen and Patton 1999; De Persis and Isidori 2000, 2001). In this model-based subcategory, residuals are constructed as the difference between the actual process behaviour and the expected one described by its mathematical model. Using these residuals, a decision is easily achievable whether there is a fault or not. One difficulty is to make a robust observer w.r.t. disturbances which are not faults (De Persis and Isidori 2000, 2001; Besançon 2003).

In this article, where continuous (smooth) nonlinear systems in state-space representation are considered, we propose a systematic methodology dedicated to fault reconstruction with an application to the field of wastewater treatment systems. Via this method, it is possible to detect sensor drift faults and incipient faults, which are not readily detected using

*Corresponding author. Email: salwa.methnani@gmail.com

ISSN 0020–7179 print/ISSN 1366–5820 online

DOI: 10.1080/00207179.2011.582156
http://www.informaworld.com

other methods. Along this article, we make the reasonable assumption that several faults do not occur simultaneously, i.e. we deal with the problem of observability with a single unknown input function.

In the context of observer-based methods, sliding mode observers are applied to reconstruct the faults by an appropriate processing of the so-called '*equivalent output error injection*' concept. Readers may refer to Tan and Edwards (2002, 2003). In other papers (Edwards 2004), unknown-input observers are used in order to reconstruct the fault.

Here, we develop a general theory of observability for unknown inputs, in order to reconstruct simultaneously the states and the graph of the fault. This theory is a by-product of the identification theory developed in Busvelle and Gauthier (2003, 2004, 2005), and it naturally leads to the use of high-gain observers.

The structure of this article is as follows. First (Section 2), we state the main lines of the theory of 'observability for unknown inputs'. In Section 3, we briefly recall the structure of the basic high-gain observer that comes naturally to the rescue. In Section 4, the proposed method is illustrated by an application to the Bleesbrük wastewater treatment plants (WWTP). Finally, Section 5 is devoted to a comparison to another popular method, with a similar geometric flavour (De Persis and Isidori 2000, 2001).

2. Observability for unknown inputs versus identification

2.1 *Generalities*

It turns out that the concept of 'observability for unknown inputs' (or 'unknown-observability') can be seen just as a rephrasing of the concept of identifiability in the sense of Busvelle and Gauthier (2003, 2004, 2005). These three papers contain a complete theory for the case of a single unknown input (or a single function of the state to be identified). In the context of FDI, a single unknown input corresponds to a single fault. If several faults occur simultaneously, one should consider several unknown inputs (the additive faults that could appear simultaneously on different sensors for instance).

The theory is parallel to the 'deterministic observation theory' of Gauthier and Kupka (1994, 1996, 2001). It requires the same mathematical tools and methods to be understood. In this section, we state the main results of the theory. Although these results can be stated in a clear intrinsic way, we limit ourselves to the characterisations in terms of 'normal forms'. Moreover, we ignore certain classical difficulties (such as finite escape-time, analyticity versus smoothness, global-Lipschitzness, etc.). For more details, the reader should refer to Busvelle and Gauthier (2003, 2004, 2005). The concept of genericity under consideration in this article is the usual one from differential topology, i.e. it's genericity w.r.t. the Whitney topology. Since in most cases the problems are located on a compact subset of the state space, it is enough in practice to consider the metric C^∞ topology: a function is close to zero if its values together with the values of all its derivatives are small enough.

A main idea that the reader should keep in mind is the following: the observability property (resp. identifiability, observability for unknown inputs) is the property of injectivity of a certain mapping. Therefore it is a very unstable property: for instance, the function $f(x)=x^3$ is injective, but it does not remain injective under perturbation by a very small function with very small derivatives. Due to this unstability, it is impossible to expect interesting general results. However, the injectivity becomes stable if we require the additional property of 'infinitesimal injectivity', i.e. injectivity of the linearisations (note that the function $f(x)=x^3$ is not infinitesimally injective at $x=0$).

These considerations are the reasons why it is not realistic to avoid considering the concept of 'infinitesimal observability' (resp: identifiability, unknown-observability).

2.2 *Definitions and systems under consideration*

Systems under consideration are smooth (C^ω or C^∞) systems of the form:

$$\Sigma\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=f(x,\varphi(t))\\ y=h(x,\varphi(t))\end{cases}\tag{1}$$

where the state $x=x(t)$ lies in an n-dimensional manifold X, $x(0)=x_0$. The observation y is $\mathbb{R}^{d_y}$-valued and f, h are, respectively, a smooth (parameterised) vector field and a smooth function. The function φ (the unknown input) is a function of time (in the context of identifiability, it is an unknown function of the state). To simplify, each trajectory is assumed to be defined on some interval $[0, T_{x_0,\varphi}[$ depending on both the initial condition and the unknown function φ, but containing a fixed time interval $I=[0, i]$.

The goal is to estimate both the state variable x and the unknown function φ: $\mathbb{R}^+\to\mathbb{R}$. In the applied part of this article (Section 4), the unknown φ will be denoted by d (for 'disturbance').

Let $\Omega=X\times L^\infty[I]$, where $L^\infty[I]$ is the set of $\mathbb{R}$-valued measurable bounded functions defined over I, and by $L^\infty[\mathbb{R}^{d_y}]$ the set of measurable bounded functions from I to $\mathbb{R}^{d_y}$.

Then we can define the input/output mapping P_Σ, mapping the initial state x_0 and the input function $\widehat{\varphi}$ to

the output function y:

$$P_\Sigma : \begin{array}{c} \Omega \to L^\infty[\mathbb{R}^{d_y}] \\ (x_0, \widehat{\varphi}(\cdot)) \to y(\cdot) \end{array} \tag{2}$$

Definition 1: Σ is said to be 'unknown-observable' if P_Σ is injective.

The infinitesimal version of unknown-observability is defined as follows. Let us consider the first variation of the system (1), where T_x denotes the tangent mapping w.r.t. x, and d_φ denotes the differential w.r.t.

$$T\Sigma_{x_0, \widehat{\varphi}, \xi_0, \eta} \begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = f(x, \widehat{\varphi}) \\ \dfrac{\mathrm{d}\xi}{\mathrm{d}t} = T_x f(x, \widehat{\varphi})\xi + d_\varphi f(x, \widehat{\varphi})\eta \\ \widehat{y} = d_x h(x, \widehat{\varphi})\xi + d_\varphi h(x, \widehat{\varphi})\eta \end{cases} \tag{3}$$

and the input/output mapping of $T\Sigma$ is

$$P_{T\Sigma, x_0, \hat{\varphi}} : \begin{array}{c} T_{x_0}X \times L^\infty[\mathbb{R}] \to L^\infty[\mathbb{R}^{d_y}] \\ (\xi_0, \eta(\cdot)) \to \widehat{y}(\cdot) \end{array} \tag{4}$$

Definition 2: Σ is said to be infinitesimally unknown-observable if $P_{T\Sigma, x_0, \hat{\varphi}}$ is injective for any $(x_0, \widehat{\varphi}(\cdot)) \in \Omega$, i.e. $\ker(P_{T\Sigma, x_0, \hat{\varphi}}) = \{0\}$ for any $(x_0, \widehat{\varphi}(\cdot))$.

In other terms, the linearisations along any trajectory of the system are observable linear time-dependent systems.

Remark: Both identifiability and infinitesimal identifiability mean injectivity of certain mapping. Clearly, injectivity depends on the domain (restricting the domain provides a weaker property). Therefore, it could seem that these notions are not well defined, since they depend on the regularity assumed for the inputs (the domain for $\widehat{\varphi}$). In fact, it is not the case: indeed, if an analytic system Σ is not (infinitesimally) unknown-observable for certain L^∞ input function, then there exists another analytic function which makes the system not (infinitesimally) unknown-observable.

2.3 *Main results stated in terms of* '*canonical forms*'

The theory is parallel to the observability theory from Gauthier and Kupka (2001): every unknown-observable system may be put (up to a change of coordinates) into one of the canonical forms presented in the Theorems 1–3.

In order to achieve default reconstruction, it is enough to develop an observer for unknown inputs adapted to each of these canonical forms.

In the previous papers (Busvelle and Gauthier 2003, 2004, 2005), the following results are established:

- Systems are generically unknown-observable if and only if the number of observations is three or more. Generic systems can be put under the canonical form of Theorem 3.
- Contrarily, unknown-observability is not at all generic when the number of observations is only one or two. In this case, infinitesimally unknown-observable systems are exhausted by certain geometric properties that are equivalent to the normal forms presented in Theorems 1 and 2.

Theorem 1: *($d_y = 1$) if Σ is infinitesimally unknown-observable, then, there is a subanalytic closed subset Z of X, of codimension 1 at least, such that for any $x_0 \in X \setminus Z$ there is a coordinate neighbourhood $(x_1, \ldots, x_n, V_{x_0})$, $V_{x_0} \subset X \setminus Z$ in which Σ (restricted to V_{x_0}) can be written as*

$$\Sigma_1 \begin{cases} \dot{x}_1 = x_2 \\ \quad \vdots \\ \dot{x}_{n-1} = x_n \\ \dot{x}_n = \psi(x, \varphi) \\ y = x_1 \end{cases} \quad \text{and} \quad \frac{\partial \psi(x, \varphi)}{\partial \varphi} \neq 0 \tag{5}$$

Theorem 2: *($d_y = 2$) if Σ is infinitesimally unknown-observable, then there is an open-dense subanalytic subset $\widetilde{U}$ of $X \times \mathbb{R}$ such that each point (x_0, φ_0) of $\widetilde{U}$ has a neighbourhood $V_{x_0} \times I_{\varphi_0}$, and coordinates x on V_{x_0} such that the system Σ restricted to $V_{x_0} \times I_{\varphi_0}$, denoted by $\Sigma_{|V_{x_0} \times I_{\varphi_0}}$, has one of the three following normal forms*:

- **Type 1 normal form:**

$$\Sigma_{2,1} \begin{cases} y_1 = x_1 & y_2 = x_2 \\ \dot{x}_1 = x_3 & \dot{x}_2 = x_4 \\ \quad \vdots & \quad \vdots \\ \dot{x}_{2k-3} = x_{2k-1} & \dot{x}_{2k-2} = x_{2k} \\ \dot{x}_{2k-1} = f_{2k-1}(x_1, \ldots, x_{2k+1}) & \\ \dot{x}_{2k} = x_{2k+1} & \\ \quad \vdots & \\ \dot{x}_{n-1} = x_n & \\ \dot{x}_n = f_n(x, \varphi) & \end{cases} \tag{6}$$

with $(\frac{\partial f_n}{\partial \varphi} \neq 0)$.

- **Type 2 normal form:**

$$\Sigma_{2,2}\begin{cases} y_1 = x_1 & y_2 = x_2 \\ \dot{x}_1 = x_3 & \dot{x}_2 = x_4 \\ \vdots & \vdots \\ \dot{x}_{2r-3} = x_{2r-1} & \dot{x}_{2r-2} = x_{2r} \\ \dot{x}_{2r-1} = \psi(x,\varphi) & \dot{x}_{2r} = F_{2r}(x_1,\ldots,x_{2r+1},\psi(x,\varphi)) \\ & \dot{x}_{2r+1} = F_{2r+1}(x_1,\ldots,x_{2r+2},\psi(x,\varphi)) \\ & \vdots \\ & \dot{x}_{n-1} = F_{n-1}(x,\psi(x,\varphi)) \\ & \dot{x}_n = F_n(x,\psi(x,\varphi)) \end{cases} \tag{7}$$

with $\frac{\partial\psi}{\partial\varphi} \neq 0$, $\frac{\partial F_{2r}}{\partial x_{2r+1}} \neq 0, \ldots, \frac{\partial F_{n-1}}{\partial x_n} \neq 0$.

- **Type 3 normal form:**

$$\Sigma_{2,3}\begin{cases} y_1 = x_1 & y_2 = x_2 \\ \dot{x}_1 = x_3 & \dot{x}_2 = x_4 \\ \vdots & \vdots \\ \dot{x}_{n-3} = x_{n-1} & \dot{x}_{n-2} = x_n \\ \dot{x}_{n-1} = f_{n-1}(x,\varphi) & \dot{x}_n = f_n(x,\varphi) \end{cases} \tag{8}$$

with $\frac{\partial(f_{n-1},f_n)}{\partial\varphi} \neq 0$.

Here is the result for the generic case.

Theorem 3: *($d_y = 3$) if Σ is an infinitesimally unknown-observable generic system, then there is a connected open dense subset Z of X such that for any $x_0 \in Z$ there exist a smooth C^∞ function F and a $(\check{y}, \check{y}', \ldots, \check{y}^{(2n)})$-dependent embedding $\Phi_{\check{y},\ldots,\check{y}^{(2n)}}(x)$ such that on Z, trajectories of $\Sigma_{x_0,\varphi}$ are mapped via $\Phi_{\check{y},\ldots,\check{y}^{2n}}$ into trajectories of the following system*:

$$\Sigma_{3+}\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_2 = z_3 \\ \vdots \\ \dot{z}_{2n} = z_{2n+1} \\ \dot{z}_{2n+1} = F(z_1,\ldots,z_{2n+1},\check{y},\ldots,\check{y}^{(2n+1)}) \\ \bar{y} = z_1 \end{cases} \tag{9}$$

where z_i, $i = 1,\ldots,2n+1$ has dimension $p-1$, and with

$$\begin{cases} x = \Phi^{-1}_{\check{y},\ldots,\check{y}^{(2n)}}(z) \\ \varphi = \psi(x,\check{y}) \end{cases} \tag{10}$$

for a certain smooth function ψ.

Here $\check{y}$ is a certain selected output among the outputs y_i, y_1 for instance, and $\bar{y}$ consists of the remaining outputs y_2, y_3.

The proof of this theorem, with detailed results in the generic case, can be found in Busvelle and Gauthier (2004). This is the crucial result for our application.

2.4 *The generic 3-5 case*

The 3-outputs 5-states case is the most simple generic case. It has the additional good property that it is naturally under a useful canonical form, as soon as the outputs are components of the state, which is often the case.

We start with a system of the form:

$$Y = (y_1, y_2, y_3) = (x_1, x_2, x_3),$$
$$x = (x_1,\ldots,x_5), \quad \dot{x}(t) = f(x)$$

We would like to realise Fault Reconstruction for an additive default $d(t)$ on the first output, i.e. in fact, $y_1(t) = x_1(t) + d(t)$. Setting $z_1(t) = x_1(t) + d(t)$, $z_2(t) = x_2(t), \ldots, z_5(t) = x_5(t)$, the system can be rewritten as

$$\begin{aligned} &y_1(t) = z_1(t), y_2(t) = z_2(t), \quad y_3(t) = z_3(t), \\ &\dot{z}_1(t) = f_1(z_1(t) - d(t), \quad z_2(t),\ldots,z_5(t)) + \dot{d} \\ &\dot{z}_i(t) = f_i(z_1(t) - d(t), \quad z_2(t),\ldots,z_5(t)), \quad i = 2,\ldots,5 \end{aligned} \tag{11}$$

or:

$$\dot{z} = g(z, d, \dot{d}) \tag{12}$$

2.4.1 *The most naive strategy*

A simple way to proceed is to assume that $\dot{d} = 0$. We get a 6-state equation of the form

$$\begin{aligned} \dot{z}(t) &= g(z_1(t), z_2(t),\ldots,z_5(t), d) \\ \dot{d} &= 0 \end{aligned} \tag{13}$$

or, setting $Z = (z, d)$,

$$\begin{aligned} \dot{Z} &= G(Z) \\ y &= (Z_1, Z_2, Z_3) \end{aligned} \tag{14}$$

Then, a step change on d corresponds exactly to a (maybe large) jump of the state Z in the model (14).

In that case, a high-gain observer will do the reconstruction job: It has precisely the property to recover arbitrarily fast large changes in the initial conditions.

System (14) is a rather general 6-state 3-output system, but the form (14) is al ready enough for our purposes.

Indeed, in general (for a generic system), the 3×3 matrix formed by the lines

$$\left(\frac{\partial G_i}{\partial z_4}, \frac{\partial G_i}{\partial z_5}, \frac{\partial G_i}{\partial d}\right), \quad i = 1, \ldots, 3 \tag{15}$$

is invertible, which means by the implicit function theorem that, freezing the variables z_1, z_2, z_3, the mapping $\tilde{G} = (G_1(z_4, z_5, d), G_2(z_4, z_5, d), \ G_3(z_4, z_5, d))$ has an inverse $\tilde{G}_1$.

It is then clear that the system is unknown-observable: knowing the output $Y(t) = (z_1(t), \ z_2(t), \ z_3(t))$ and differentiating, we get $(\dot{z}_1(t), \dot{z}_2(t), \dot{z}_3(t)) = \tilde{G}(z_4(t), z_5(t), d(t))$, which we can invert for each value of $z_1(t)$, $z_2(t)$, $z_3(t)$, and we get the knowledge of $z_4(t)$, $z_5(t)$, $d(t)$.

This shows that actually the system is not only unknown-observable (which we know), but also provides a practical way to observe, by using approximate derivators.

2.4.2 *The general strategy*

A more general strategy is to use as in Busvelle and Gauthier (2003, 2004, 2005) a **local model** for the fault $d(t)$. For example, a simple local model is $d^{(k)} = 0$. The question is not that this polynomial models the function d globally as a function of t, but only locally, on reasonable time intervals (reasonable w.r.t. the performances required for input-state reconstruction).

Now, we are in the general situation of a $6 + k$-state, 3-output system. The fact that the original system is infinitesimally unknown-observable implies that the extended $6 + k$-system can be put under certain appropriate observability normal form.

Again, for this normal form, the use of approximate derivators would allow state reconstruction.

2.5 *Observers for unknown inputs*

It is a remarkable fact that, for all the normal forms described above, such a polynomial local model allows the use of the high-gain observers from Gauthier and Kupka (2001).

We leave the details to the reader and we just explain below (Section 3) what happens in the 3-5 case (our application), when we make the naive assumption $\dot{d} = 0$ of Section 2.4.1.

2.6 *The necessity of the theoretical analysis*

One could ask: why is it necessary to perform such a heavy theoretical analysis to get the trivial conclusion that 'high gain observers must be used'?

In fact, the preliminary analysis of the unknown-input observability property is absolutely necessary, as shows the following example. It also shows that 'parametric identification' may be very dangerous without careful analysis.

The example is even linear, therefore it leads more simply to the use of a standard Luenberger observer (not high gain). One can imagine that in the nonlinear case, more catastrophic phenomena may appear.

Consider the linear system on $\mathbb{R}^2$:

$$\Sigma_e \begin{cases} \dot{x}_1 = x_2 - u \\ \dot{x}_2 = u \\ y = x_1 \end{cases} \tag{16}$$

This system is not unknown-observable: actually, setting $X = (x_{10}, x_{20})$, the mapping $(u(\cdot), X) \to y(\cdot)$ is linear, and it is easily seen that it is not injective: its Kernel K is the set of couples of the form $(u = e^t x_{20}, X = (0, x_{20}))$.

However choosing, without observability analysis, a local model of the form $u^{(k)} = 0$, one obtains the extended linear system:

$$\Sigma_{e,1} \begin{cases} \dot{x}_1 = x_2 - u \\ \dot{x}_2 = u \\ \dot{u} = u_1 \\ \vdots \\ \dot{u}_{k-1} = 0 \\ y = x_1 \end{cases} \tag{17}$$

Note that $(\Sigma_{e,1})$ is an observable linear system, and that a standard Luenberger observer will provide 'some result', with arbitrary exponential decay.

However, this result may be a nonsense. In fact, the system Σ_e, although non unknown-observable, is unknown-observable inside the class of polynomial unknown-inputs.

3. Our choice of the high-gain observer in the 3-5 case

3.1 *Preliminary*

Let us go back to the system (14), and consider the 3×3 matrix J defined in formula (15), $J_{ij} = \frac{\partial G_i}{\partial Z}$, $i = 1, \ldots, 3$, $j = 4, \ldots, 6$.

The invertibility of this Jacobian matrix characterises the infinitesimal observability in the sense of Gauthier and Kupka (2001), as was observed above.

In this particular 3×5 case, it provides a generalisation of the basic single-output observability normal form from Gauthier and Kupka (1994) (see also Theorem 2.1, p. 22 in Gauthier and Kupka (2001)).

Actually, in the two-dimensional single output case considered in Gauthier and Kupka (2001), we would have the corresponding normal form:

$$\begin{cases} y = x_1 \\ \dot{x_1} = f_1(x_1, x_2, u) \\ \dot{x_2} = f_2(x_1, x_2, u) \qquad \text{with } \dfrac{\partial f_1}{\partial x_2} \neq 0 \end{cases} \tag{18}$$

The condition $\frac{\partial f_1}{\partial x_2} \neq 0$ for infinitesimal observability is the analogue of our condition that J is invertible.

At this step, we could use (up to a certain additional simple change of coordinates) a high-gain extended Kalman filter. In fact, here is a simpler solution. Due to Hammouri and Farza (2003), a multi-output generalisation of the results in Gauthier and Kupka (1994) shows that we can directly apply the basic version of the (constant gain) high-gain Luenberger observer.

3.2 *The multi-output high-gain Luenberger observer*

We forget about the usual difficulty in high-gain observers of any kind that consists of smoothly prolongating the system out of a compact set (the 'physical' space), in order that it meets certain global-Lipschitz assumptions. In the case of our application, this is more or less trivial.

Physical space will be

$$\text{Ps} = \{M_1 \geq z_1 - d \geq \varepsilon_1, M_i \geq z_i \geq \varepsilon_i > 0\}, \tag{19}$$

for $i = 2, \ldots, 5$.

For this 3-5 case, it is easily seen that the condition from Hammouri and Farza (2003), that allows the use of a constant gain high-gain observer, reduces to the following property (P):

(P) There is a constant 3×3 matrix S such that all (which means for all possible values of the variables in the physical domain) the 3×3 matrices J satisfy: $S^{\mathbb{T}}J + J^{\mathbb{T}}S \leq -aId$, for a certain $a > 0$.

This will be the case in our application, reconstruction of sensor fault for the Bleesbrük WWTP, presented in Section 5.

Let us point out again that, when property (P) holds, it is possible to construct a constant gain, high-gain Luenberger observer that guarantees arbitrarily fast state reconstruction (or fault reconstruction in our case).

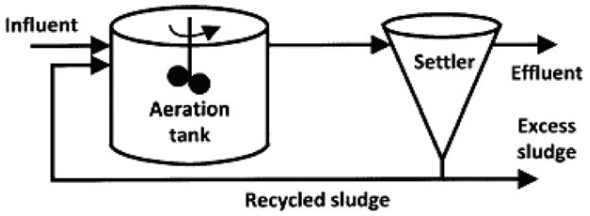


Figure 1. Typical small-size activated sludge treatment plant.

4. Application

4.1 *Activated sludge process*

Due to its efficiency, the activated sludge process (ASP) is the most frequent device for wastewater treatment. An ASP is a chemical–biological process, where a mixed community of microorganisms (called activated sludge), is used to remove pollutant. A basic ASP layout is composed of an aerated tank and a settler (Figure 1).

Wastewater is treated first in the tank, where the level of substrate is degraded by microorganisms. Next, sedimentation takes place in the settler, in order to separate the clean water and the settled solid. A portion of the sludge is recycled with the aim to maintain an appropriate biomass concentration. The remaining amount of sludge is purged.

Nomenclature

S_I	Soluble inert organic matter concentration (mg L^{-1})
S_S	Readily biodegradable substrate concentration (mg L^{-1})
S_O	Dissolved oxygen concentration (mg L^{-1})
S_O^{sat}	Dissolved oxygen saturation concentration (mg L^{-1})
S_{NO}	Nitrate and nitrite nitrogen concentration (mg L^{-1})
S_{NH}	Ammonia nitrogen concentration (mg L^{-1})
S_{ND}	Soluble biodegradable organic nitrogen concentration (mg L^{-1})
X_I	Particulate inert organic matter concentration (mg L^{-1})
X_S	Slowly biodegradable substrate concentration (mg L^{-1})
$X_{B,H}$	Active heterotrophic biomass concentration (mg L^{-1})
$X_{B,A}$	Active autotrophic biomass concentration (mg L^{-1})
X_{ND}	Particulate biodegradable organic nitrogen concentration (mg L^{-1})
b_A	Autotrophic organisms decay rate coefficient (d^{-1})

b_H Heterotrophic organisms decay rate coefficient (d^{-1})
fr_{XI} Fraction of biomass generating the particulate products (−)
i_{NBM} Mass of nitrogen in the biomass (gNg^{-1}_{COD})
i_{NXI} Mass of nitrogen in the inert particulate organic matter (gNg^{-1}_{COD})
K_l a Coefficient of oxygen rate (d^{-1})
$K_{NH,A}$ Half-saturation coefficient of ammonia for autotrophs ($gNHm^{-3}$)
K_{N0} Half-saturation coefficient of nitrate for denitrifying heterotrophs ($gNOm^{-3}$)
$K_{0,A}$ Half-saturation coefficient of oxygen autotrophs ($gO_2\,m^{-3}$)
$K_{0,H}$ Half-saturation coefficient of oxygen heterotrophs ($gO_2\,m^{-3}$)
K_S Half-saturation coefficient for heterotrophic organisms ($gDCOm^{-3}$)
K_X Half-saturation coefficient for hydrolysis of slowly biodegradable substrate $gDCOg^{-1}_{DCO}$
Y_A Yield coefficient for autotrophic organisms (−)
Y_H Yield coefficient for heterotrophic organisms (−)
μ_A Maximum specific growth rate for autotrophic organisms (d^{-1})
μ_H Maximum specific growth rate for heterotrophic organisms (d^{-1})
^{in} influent (d^{-1})
D^{in} Influent flow rate (m^3d^{-1})

Several mathematical models are proposed for the WWTP. The most popular model is the activated sludge model No. 1 (ASM1). However, this nonlinear model is rather complex: 11 state variables and 19 constant parameters. Different kinds of reduced models for the activated sludge plant have been proposed (Jeppsson and Olsson 1993; Steffens, Lant, and Newell 1997; Smets, Haegebaert, Carrette, and Impe 2003; Mulas, Tronci, and Baratti 2007). Here we consider the reduced five-dimensional dynamical model that was developed by Chachuat, Roche, and Latifi (2003).

The following simplifications were applied:

- *Dynamic simplification*: When applying a homotopy method, heterotrophic ($X_{B,H}$), autotrotophic ($X_{B,A}$) biomass and inert particulate organic compounds (X_I) were detected as the slowest state dynamics. Thus, these variables can be assumed constant over a few days. Eliminating these three states with the concentration of soluble inert organic compound (S_I), a seven-dimensional dynamic model was obtained.
- *Organic compounds simplification*: Based on more heuristic considerations, soluble (S_S) and particulate (X_S) concentrations are glued into a single organic compound (denoted by X_{DCO}).
- *Nitrogenised compounds simplification*: Due to a simplification of the mathematical expression that describes the organic nitrogen hydrolysis, the dynamics with respect to soluble and particulate organic nitrogen becomes a separated independent system that we do not consider.

The following standard assumptions are also considered:

- The reactor is well mixed.
- The settler is perfect: No reaction occurs there and the separation between solid and liquid is ideal.

These simplifications lead to the following set of equations:

$$\dot{S_{NO}} = D^{in}\left(S^{in}_{NO} - S_{NO}\right) - \alpha_1 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \cdot \frac{K_{O,H}}{K_{O,H} + S_O} \times \frac{S_{NO}}{K_{NO} + S_{NO}} + \alpha_2 \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_O}{K_{O,A} + S_O} \quad (20)$$

$$\dot{S_{NH}} = D^{in}\left(S^{in}_{NH} - S_{NH}\right) - \alpha_3 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \cdot \frac{S_O}{K_{O,H} + S_O} + \alpha_4 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \frac{K_{O,H}}{K_{O,H} + S_O} \frac{S_{NO}}{K_{NO} + S_{NO}} - \alpha_2 \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_O}{K_{O,A} + S_O} + \alpha_5 S_{ND} \quad (21)$$

$$\dot{S_O} = -D^{in} S_O - \alpha_6 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \frac{S_O}{K_{O,H} + S_O} - \alpha_7 \frac{S_{NH}}{K_{NH,A} + S_{NH}} \frac{S_O}{K_{O,A} + S_O} + k_L a\left(S^{sat}_O - S_O\right) \quad (22)$$

$$\dot{X_{DCO}} = D^{in}\left(X^{in}_{DCO} - \alpha_8 X_{DCO}\right) - \alpha_9 \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \times \frac{S_O}{K_{O,H} + S_O} + \alpha_{10} \frac{X_{DCO}}{K_{DCO} + X_{DCO}} \frac{K_{O,H}}{K_{O,H} + S_O} \times \frac{S_{NO}}{K_{NO} + S_{NO}} + \alpha_{11} \quad (23)$$

$$\dot{S_{ND}} = D^{in}\left(S^{in}_{ND} - S_{ND}\right) - \alpha_5 S_{ND} + \alpha_{12} \frac{X_{DCO}}{K_{ND} + X_{DCO}} \times \frac{S_O}{K_{O,H} + S_O} + \alpha_{13} \frac{X_{DCO}}{K_{ND} + X_{DCO}} \frac{K_{O,H}}{K_{O,H} + S_O} \times \frac{S_{NO}}{K_{NO} + S_{NO}} \quad (24)$$

with: $\alpha_1 = \mu_H \cdot X_{B,H} \cdot \eta_{NO,g} \cdot \frac{1-Y_H}{2.86 Y_H}$, $\alpha_2 = \frac{\mu_A}{Y_A} \cdot X_{B,A}$, $\alpha_3 = \mu_H \cdot X_{B,H} \cdot i_{NBM}$, $\alpha_4 = \mu_H \cdot X_{B,H} \cdot i_{NBM} \cdot \eta_{NO,g}$, $\alpha_5 = k_a \cdot X_{B,H}$, $\alpha_6 = \mu_H \cdot X_{B,H} \cdot \frac{1-Y_H}{Y_H}$, $\alpha_7 = 4.57 \cdot \frac{\mu_A}{Y_A} \cdot X_{B,A}$, $\alpha_8 = \frac{K_S}{K_{DCO}}$, $\alpha_9 = \frac{\mu_H \cdot X_{B,H}}{Y_H}$, $\alpha_{10} = \frac{\mu_H \cdot X_{B,H}}{\cdot} Y_H \cdot \eta_{NO,g}$, $\alpha_{11} = (1 - f_{r_{XI}}) \cdot (b_H \cdot X_{B,H} + b_A \cdot X_{B,A})$, $\alpha_{12} = k_h \cdot \frac{X_{ND}}{X_S} \cdot X_{B,H}$, $\alpha_{13} = k_h \cdot \frac{X_{ND}}{X_S} \cdot X_{B,H} \cdot \eta_{NO,h}$.

In this article, we work in simulation using certain data generated by the team of modelling and simulation of LTI-CRP Henri Tudor in Luxembourg, by using the ASM1 model and SIMBA software (see http://www.enic.impl-nancy-fr/COSTWWTP/Benchmark).

In fact, dry, rain and storm data files are generated from a benchmark simulation of the results for the Bleesbrük wastewater plant (in Luxembourg).

The measured concentrations of this station are: the dissolved oxygen (S_O), nitrate (S_{NO}) and ammonia (S_{NH}).

4.2 *The Luenberger high-gain observer*

The purpose of this study is the reconstruction of the sensor faults. A sensor fault is an unknown function that will be identified on-line. Consider the reduced ASP system described by Equations (20)–(24). The unknown function d will represent the fault signal applied to the S_{NO} sensor. It is assumed to be an additive fault. As explained above, in order to reconstruct the function d, the state vector is extended by making d a state variable, and we just model the fault as a jump of initial conditions: $\dot{d} = 0$.

The vector G is as follows:

$$G(Z) = \begin{pmatrix} D^{in}(S_{NO}^{in} - (z_1 - d)) + \alpha_2 \frac{z_2}{K_{NH,A} + z_2} \frac{z_3}{K_{O,A} + z_3} - \alpha_1 \frac{K_{O,H}}{K_{O,H} + z_3} \frac{(z_1 - d)}{K_{NO} + (z_1 - d)} z_4, \\ D^{in}(S_{NH}^{in} - z_2) - \alpha_2 \frac{z_2}{K_{NH,A} + z_2} \frac{z_3}{K_{O,A} + z_3} - \left(\alpha_3 \frac{z_3}{K_{O,H} + z_3} + \alpha_4 \frac{K_{O,H}}{K_{O,H} + z_3} \frac{(z_1 - d)}{K_{NO} + (z_1 - d)} \right) z_4 + \alpha_5 z_5, \\ - D^{in} z_3 - \alpha_7 \frac{z_2}{K_{NH,A} + z_2} \frac{z_3}{K_{O,A} + z_3} + k_l a (S_O^{sat} - z_3) - \alpha_6 \frac{z_3}{K_{O,H} + z_3} z_4, \\ \frac{(D^{in} XDCO^{in} + \alpha_{11})}{K_{DCO}} (1 - z_4)^2 - D^{in} \alpha_8 (1 - z_4) z_4 - \frac{1}{K_{DCO}} \left(\alpha_9 \frac{z_3}{K_{O,H} + z_3} + \alpha_{10} \frac{K_{O,H}}{K_{O,H} + z_3} \right. \\ \left. \times \frac{(z_1 - d)}{K_{NO} + (z_1 - d)} \right) (1 - z_4)^2 z_4, \\ D^{in}(S_{ND}^{in} - z_5) - \alpha_5 z_5 + \left(\alpha_{12} \frac{z_3}{K_{O,H} + z_3} + \alpha_{13} \frac{K_{O,H}}{K_{O,H} + z_3} \frac{(z_1 - d)}{K_{NO} + (z_1 - d)} \right) \frac{z_4 K_{DCO}}{K_{ND} + z_4 (K_{DCO} - K_{ND})} \\ 0 \end{pmatrix}$$

Here, of course, $z_6 = d$.

Remark: Here, for simplicity in the expressions, we have made the extra change of variables $z_4 = \frac{X_{dco}}{K_{dco} + X_{dco}}$. But this is not absolutely necessary.

The equation of the standard high-gain Luenberger observer is

$$\dot{\hat{X}}(t) = G(\hat{X}) - K_\theta (C\hat{X} - y) \tag{25}$$

where $K_\theta = \Delta_\theta K$ for $\theta > 1$, large enough and:

- Δ_θ is the block diagonal matrix $\Delta_\theta = BD(\theta I_3, \theta^2 I_3)$, where I_3 is the three-dimensional identity matrix,
- K is a certain constant gain, such that: $(\tilde{G}^*(\hat{X}) - KC)'L - L(\tilde{G}^*(\hat{X}) - KC) \leq -aId$, $a > 0$, L constant symmetric positive definite.

Here $\tilde{G}^*(X)$ denotes the Jacobian matrix of $\tilde{G}(X)$ w.r.t. X ($\tilde{G}(X)$ defined in Section 2.4.1).

In the single output case, the existence of such a K comes from Gauthier and Kupka (1994). The multi-output case is much more complicated and has been studied in Hammouri and Farza (2003). The existence of K is guaranteed by the property (P) of Section 3.

To check that property (P) holds in our case, it is enough to observe that the Jacobian matrix J has the following form on the 'physical space' Ps (from (19)):

$$J = \begin{pmatrix} -a & 0 & f \\ -b & \alpha & e \\ -c & 0 & 0 \end{pmatrix},$$

where all the functions a, b, c, f, e, α are strictly positive. The technical lemma in our Appendix provides property (P).

Here, we did not use the explicit construction of the constant gain K provided by Hammouri and Farza (2003), but a heuristic one that works quite well. We have chosen $K = L^{-1}C'$, where L is the solution of the following Riccati equation:

$$-G'L - LG + C'C - LQL = 0 \tag{26}$$

with $Q = \mathrm{diag}(10^{-3}, 10^{-3}, 10^{-3}, 10^{-3}, 10^{-3}, 10^{-1})$, and

$$G = \begin{pmatrix} 0 & 0 & 0 & \frac{\partial g_1(x)}{\partial x_4} & \frac{\partial g_1(x)}{\partial x_5} & \frac{\partial g_1(x)}{\partial d} \\ 0 & 0 & 0 & \frac{\partial g_2(x)}{\partial x_4} & \frac{\partial g_2(x)}{\partial x_5} & \frac{\partial g_2(x)}{\partial d} \\ 0 & 0 & 0 & \frac{\partial g_3(x)}{\partial x_4} & \frac{\partial g_3(x)}{\partial x_5} & \frac{\partial g_3(x)}{\partial d} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

at a typical (or average) point x.

We obtain the constant Luenberger gain:

$$K = \begin{pmatrix} 3.73 & 1.43 \times 10^{-2} & 5 \times 10^{-3} \\ 1.43 \times 10^{-2} & 5.69 \times 10^{-1} & 4.255 \times 10^{-1} \\ 5 \times 10^{-3} & 4.255 \times 10^{-1} & 2.9189 \\ 0 & -2 \times 10^{-4} & -10^{-3} \\ 0 & 10^{-3} & -2 \times 10^{-4} \\ 10^{-1} & 3 \times 10^{-4} & -5 \times 10^{-4} \end{pmatrix}$$

4.3 *Numerical simulations*

The three outputs are corrupted by an additive coloured noise. In a standard way, we have chosen an Ornstein–Uhlenbeck process X_t, simulating the following stochastic equation (Uhlenbeck and Ornstein 1930):

$$\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \delta\sqrt{2a}\,\mathrm{d}W_t, \tag{27}$$

where W_t is a standard Wiener process. The coefficients a, δ have been chosen in order to get the realistic noise level shown in the results below.

The kinetic and stoichiometric parameter values considered are those defined for the ASM1 model (Smets et al. 2003) (Table 1). The complete other parameter values can be found in Table 2.

Table 1. ASM1 kinetic and stoichiometric parameters.

Parameter	Value	Range of variation
Y_H	0.67	0.38–0.75
i_{NBM}	0.08	–
K_S	10	5–225
$K_{0,H}$	0.2	0.01–0.20
K_{NO}	0.5	0.01–0.50
$K_{NH,A}$	1.0	–
$K_{0,A}$	0.40	0.40–2.0
$\eta_{NO,g}$	0.8	0.6–13.2
$\eta_{NO,h}$	0.8	–
Y_A	0.24	0.07–0.28
fr_{XI}	0.08	–
μ_H	4.0	0.60–13.2
μ_A	0.5	0.20–1.0
k_a	0.05	–
k_h	3.0	–
f_{SS}	0.79	–
D^{in} (d^{-1})	69.2 (mean)	62.85–79.52

Table 2. Different parameter values.

Parameter	Value
α_1	3923
α_2	283
α_3	796
α_4	637
α_5	124
α_6	3904
α_7	1293
α_8	0.045
α_9	14,860
α_{10}	11,888
α_{11}	693
α_{12}	480
α_{13}	384
K_{DCO}	220
K_{ND}	258
$X_{B,A}$	$136\,\mathrm{gDCOm}^{-3}$
$X_{B,H}$	$2489\,\mathrm{gDCOm}^{-3}$
X_{ND}	$6\,\mathrm{gNm}^{-3}$
k_la	$240\,\mathrm{d}^{-3}$
V_O	$1333\,\mathrm{m}^3$

4.3.1 *Step fault*

At the second day, a step fault is applied to the S_{NO} sensor (Figure 2). The amplitude equal to $2\,\mathrm{mg\,L}^{-1}$ (compared to an average value of $6\,\mathrm{mg\,L}^{-1}$). The three state variables S_{NO}, S_{NH} and S_O are measured.

Simulations, displayed in Figures 3–5, show the observer outputs: d, X_{DCO}, S_{ND}. They demonstrate the effectiveness of the proposed method to estimate states and simultaneously reconstruct the sensor faults even for systems subject to noisy measurements.

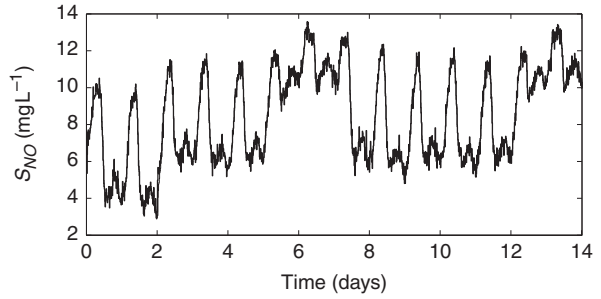


Figure 2. The faulty S_{NO} sensor.

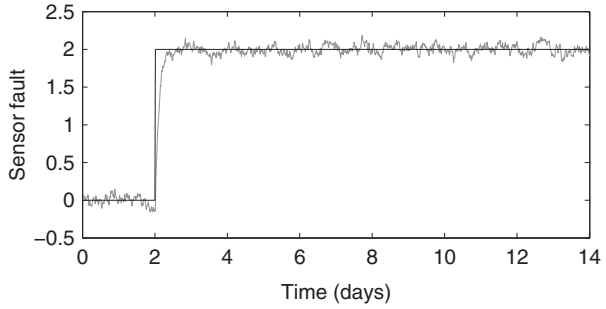


Figure 3. The difference between the applied and the reconstructed step sensor fault.

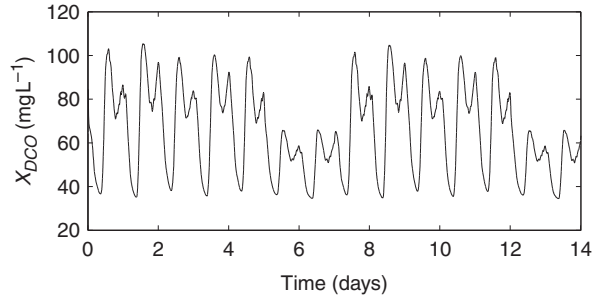


Figure 4. The difference between estimated and real X_{DCO} (unmeasured state) – no visible difference.

Although we reconstruct simultaneously the unknown state variables X_{DCO}, S_{ND}, the main purpose of these simulations is to detect and reconstruct the additive sensor fault d. One readily checks in Figure 3, that the observer's output d is close to zero when there is no fault (before day 2), while it reaches quickly the value $2\,\mathrm{mg\,L^{-1}}$ when the fault occurs.

4.3.2 *Slow drift and intermittent fault*

In order to validate completely the method, it is interesting to consider, besides the step, the most classical types of malfunctions: slow drift and intermittent fault. The corresponding simulation results are shown, respectively, in Figures 6 and 7. In these two

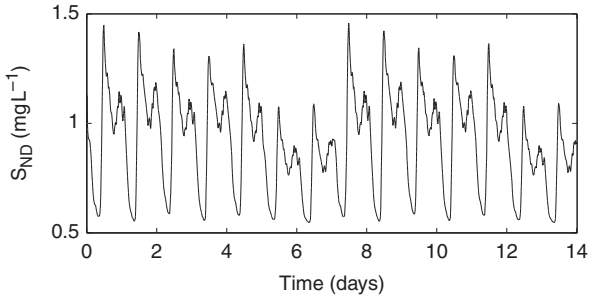


Figure 5. The difference between estimated and real S_{ND} (unmeasured state) – no visible difference.

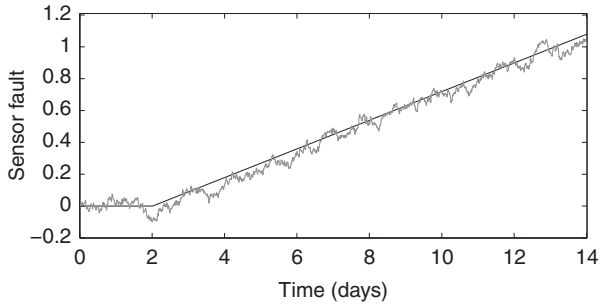


Figure 6. The difference between the applied and the reconstructed slow drift sensor fault.

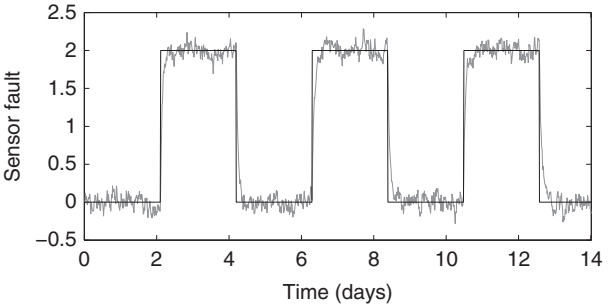


Figure 7. The difference between the applied and the reconstructed intermittent sensor fault.

figures, one can see that the method preserves the shape and amplitude of the fault with high fidelity, despite the noisy measurements.

5. Comparison with other methods

Our method lies in the framework of geometric control theory. Another popular method of this type (referred to as the DPIM) has been developed by De Persis and Isidori (2000, 2001). Let us analyse what is different in our approach.

The DPIM is rather closely related to ours, however the basic problem is different: one wants (1) to detect the

occurrence of the fault and simultaneously (2) to reject perturbations. What we do here is in a sense weaker since we do not ask rejection of any perturbation.

In this case where there is no perturbation, however, the DPIM makes sense, and we feel that our method is stronger, from two points of view:

(a) we not only detect the occurrence of the fault, but also reconstruct the fault.
(b) we do not limit ourselves to control affine systems (w.r.t. the fault in particular), but we consider general nonlinearities.

This last point (b) has to be developed: assume, for instance, an additive sensor fault on the output of the form $y(t) = h(x) + d(t)$ for simplicity. Without loss of generality, we may assume that $y(x) = x_1 + d$. Then, setting $x_1 + d = \tilde{x_1}$, we get $y = \tilde{x_1}$, and the equations for the dynamics are already fully nonlinear w.r.t. $d(t)$, even starting from a system affine w.r.t. $d(t)$. The DPIM simply does not apply.

It is the case in our application. Now, let us have a look to the example in De Persis and Isidori (2001), where the DPIM not only works, but also allows to reconstruct fully the fault (we cite: 'In this particular example, it is even possible to identify the value of m.').

It turns out that, in their case, $l =$ number of controls $= 3$, $k =$ number of 'unknown faults' $= 1$. Assuming the l (=3) controls as known constants, we are in the generic situation of $m = 3$ outputs, $k = 1$ unknown input: the generic case.

Actually, it is easily seen that Theorem 3 applies, and that the change of variables chosen in De Persis and Isidori (2001) leads exactly to our normal form Σ_{3+} of Theorem 3.

Now considering the controls as nonconstant, it is easy to see that we obtain the normal form Σ_{3+}, but with its linear part becoming time-dependent through the 3 controls. Hence, our high-gain observer still applies, and this is more or less what is suggested in De Persis and Isidori (2001) at the end of this article.

Other related works in the same spirit are:

(a) Hou and Patton (1998), but in the linear case,
(b) Kabore and Wang (2001), where conditions are given for observability (detectability) for unknown inputs. This work does not has really a geometric flavour, and moreover, it applies to control affine problems only.

6. Conclusion

An approach for sensor fault identification and reconstruction for a class of nonlinear systems has been proposed based on a theory of observability for unknown inputs. The sensor fault is considered as the unknown input. Our theory naturally leads to the use of a Luenberger-type high-gain observer. The Bleesbrük ASP with ASM^1 model provides an ideal case study. Simulations with ASP have shown the effectiveness of our strategy for fault reconstruction, in the presence of noisy measurements. The Luenberger high-gain observer used for this application is specially simple.

There are several open questions after this work: first, from theoretical point of view, it seems to us that it is now necessary to complete our theory (to the case of simultaneous faults, for instance). Although it is rather clear how to proceed, the task is not technically so obvious. From the point of view of the application, we are starting to apply the method to a real wastewater system. As usual, this is presumably the beginning of a long story.

Acknowledgements

This work is supported by ANR program BLANC GCM (Geometric Control methods, Sub-Riemannian geometry and application) and INRIA project GECO.

References

Besançon, G. (2003), 'High-gain Observation with Disturbance Attenuation and Application to Robust Fault Detection', *Automatica*, 39, 1095–1102.

Busvelle, E., and Gauthier, J.-P. (2003), 'On Determining Unknown Functions in Differential Systems, with an Application to Biological Reactor', *ESAIM: Control, Optimisation and Calculus of Variations*, 9, 509–553.

Busvelle, E., and Gauthier, J.-P. (2004), 'New Results on Identifiability of Nonlinear Systems', in *2nd Symposium on Systems, Structure and Control*, Oaxaca, Mexico.

Busvelle, E., and Gauthier, J.-P. (2005), 'Observation and Identification Tools for Non Linear Systems. Application to a Fluid Catalytic Cracker', *International Journal of Control*, 78, 208–234.

Chachuat, B., Roche, N., and Latifi, M. (2003), 'Reduction of the ASM1 Model for Optimal Control of Small-size Activated Sludge Treatment Plants', *Revue des Sciences de L'eau/Journal of Water Science*, 16, 05–26.

Chen, J., and Patton, R. (1999), *Robust Model-based Fault Diagnosis for Dynamic Systems*, Boston: Kluwer Academic.

De Persis, C., and Isidori, A. (2000), 'On the Observability Codistributions of a Nonlinear System', *Systems & Control Letters*, 40, 297–304.

De Persis, C., and Isidori, A. (2001), 'A Geometric Approach to Nonlinear Fault Detection and Isolation', *IEEE Transactions on Automatic Control*, 46, 853–865.

Edwards, C. (2004), 'A Comparison of Sliding Mode and Unknown Input Observers for Fault Reconstruction', in *IEEE Conference on Decision and Control*, pp. 5279–5284.

Frank, P.-M. (1990), 'Fault Diagnosis in Dynamic Systems using Analytical and Knowledge Based Redundancy:a Survey and Some New Results', *Automatica*, 26, 459–474.
Frank, P.-M. (1996), 'Analytical and Qualitative Model-based Fault Diagnosis: A Survey and Some New Results', *European Journal of Control*, 2, 6–28.
Frank, P.-M., and Ding, S.-X. (1997), 'Survey of Robust Residual Generation and Evaluation Methods in Observer Based Fault Detection Systems', *Journal of Process Control*, 7, 403–424.
Gauthier, J.-P., and Kupka, I. (1994), 'Observability and Observers for Nonlinear Systems', *SIAM Journal on Control*, 32, 975–994.
Gauthier, J.-P., and Kupka, I. (1996), 'Observability for Systems with More Outputs than Inputs', *Mathematische Zeitschrift*, 223, 47–78.
Gauthier, J.-P., and Kupka, I. (2001), *Deterministic Observation Theory and Applications*, Cambridge: Cambridge University Press.
Hammouri, H., and Farza, M. (2003), 'Nonlinear Observers for Locally Uniformly Observable Systems', *ESAIM: Control, Optimization and Calculus of Variations*, 9, 353–370.
Hou, M., and Patton, R. (1998), 'Input Observability and Input Reconstruction', *Automatica*, 34, 789–794.
Isermann, R. (2011), *Fault Diagnosis Applications: Model Based Condition Monitoring, Actuators, Drives, Machinery, Plants, Sensors, and Faulttolerant Systems* (1st ed.), New York: Springer.
Isermann, R., and Ball, P. (1996), 'Trends in the Application of Model Based Fault Detection and Diagnosis of Technical Processes', in *13th IFAC World Congress*, New Jersey, pp. 1–12.
Jeppsson, U., and Olsson, G. (1993), 'Reduced Order Models for Online Parameter Identification of the Activated Sludge Process', *Water Science and Technology*, 28, 173–183.
Kabore, P., and Wang, H. (2001), 'Design of Fault Diagnosis Filters and Fault-tolerant Control for a Class of Nonlinear Systems', *IEEE Transactions on Automatic Control*, 46, 1805–1810.
Mulas, M., Tronci, S., and Baratti, R. (2007), 'Development of a 4-measurable States Activated Sludge Process Model Deduced from the ASM1', in *8th International Symposium on Dynamics and Control of Process Systems*, Cancun, Mexico, pp. 213–218.
Palade, V., and Bocaniala, C. (2010), *Computational Intelligence in Fault Diagnosis* (1st ed.), New York: Springer.
Patton, R.-J., and Chen, J. (1993), 'A Survey of Robustness Problems in Quantitative Model-based Fault Diagnosis', *Applied Maths and Computer Science*, 3, 339–416.
Smets, I., Haegebaert, J., Carrette, R., and Impe, J.V. (2003), 'Linearization of the Activated Sludge Model Asm1 for Fast and Reliable Predictions', *Water Research*, 37, 1831–1851.
Steffens, M.-A., Lant, P.-A., and Newell, R.-B. (1997), 'A Systematic Approach for Reducing Complex Biological Wastewater Treatment Model', *Water Science and Technology*, 31, 590–606.
Tan, C.-P., and Edwards, C. (2002), 'Sliding Mode Observers for Detection and Reconstruction of Sensor Faults', *Automatica*, 38, 1815–1821.
Tan, C.-P., and Edwards, C. (2003), 'Sliding Mode Observers for Robust Detection and Reconstruction of Actuator Sensor Faults', *International Journal of Robust Nonlinear Control*, 13, 443–463.
Uhlenbeck, G.-E., and Ornstein, L.-S. (1930), 'On the Theory of Brownian Motion', *Physical Review*, 36, 823–841.
Yang, H., and Saif, M. (1995), 'Fault Detection in a Class of Nonlinear Systems via Adaptive Sliding Observer', in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 2199–2204.

Appendix: Technical lemma

Let A be a compact subset contained in the set of matrices of the form

$$A = \begin{pmatrix} -a & 0 & d \\ -b & \alpha & e \\ -c & 0 & 0 \end{pmatrix}$$

with $a, b, c, d, e, \alpha > 0$.

Let N be of the form

$$N = \begin{pmatrix} 0 & 0 & -r \\ 0 & -1 & 0 \\ rs & 0 & 0 \end{pmatrix}$$

Then, for $s, r > 0$ large enough

$$N'A + A'N < -\beta Id,\ \beta > 0\ \forall A \in C$$

Proof:

$$N'A = \begin{pmatrix} -crs & 0 & 0 \\ b & -\alpha & -e \\ ar & 0 & -dr \end{pmatrix}$$

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$X'N'AX = -crsx^2 - \alpha y^2 - drz^2 + bxy - eyz + arxz$$

$$-\alpha y^2 + bxy - eyz = -\left(\alpha y^2 - 2\sqrt{\alpha}y\frac{bx - ez}{2\sqrt{\alpha}} + \left(\frac{bx - ez}{2\sqrt{\alpha}}\right)^2\right) + \frac{(bx - ez)^2}{4\alpha^2}$$

$$-\alpha y^2 + bxy - eyz = -\left(\sqrt{\alpha}y - \frac{bx - ez}{2\sqrt{\alpha}}\right)^2 + \frac{b^2x^2}{4\alpha^2} + \frac{e^2z^2}{4\alpha^2} - \frac{bexz}{2\alpha^2}$$

$$X'N'AX = -\left(\sqrt{\alpha}y - \frac{bx - ez}{2\sqrt{\alpha}}\right)^2 + \left(\left(\frac{b^2}{4\alpha^2} - crs\right)x^2 + \left(\frac{e^2}{4\alpha^2} - dr\right)z^2 + \left(ar - \frac{be}{2\alpha^2}\right)xz\right)$$

The result follows QED.