HAL
open science

Identification des systèmes non linéaires par une approche multi-modèle à états découplés

Rodolfo Orjuela, Didier Maquin, José Ragot

To cite this version:

Rodolfo Orjuela, Didier Maquin, José Ragot. Identification des systèmes non linéaires par une approche multi-modèle à états découplés. Journées Identification et Modélisation Expérimentale JIME'2006, Nov 2006, Poitiers, France. pp.CDROM. hal-00115198

HAL Id: hal-00115198

https://hal.science/hal-00115198v1

Submitted on 20 Nov 2006

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

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

Identification des systèmes non linéaires par une approche multi-modèle à états découplés

Rodolfo ORJUELA, Didier MAQUIN et José RAGOT

Centre de Recherche en Automatique de Nancy
UMR 7039, Nancy-Université, CNRS
2, Avenue de la Forêt de Haye, 54516 Vandœuvre-lès-Nancy Cedex France

{Rodolfo.Orjuela, Didier.Maquin, Jose.Ragot}@ensem.inpl-nancy.fr
http://www.cran.uhp-nancy.fr

Résumé— *Les sciences de l'ingénieur font largement appel aux modèles non linéaires pour décrire les comportements dynamiques des systèmes physiques réels. Ces modèles peuvent cependant s'avérer difficiles à obtenir et/ou à manipuler dans un objectif d'identification, de commande ou de diagnostic. Les multi-modèles offrent une alternative intéressante pour contourner ces difficultés car ils permettent de prendre en compte la présence de plusieurs modes de fonctionnement. Cet article traite de l'identification hors ligne des systèmes non linéaires par une approche multi-modèle à états découplés par opposition au multi-modèle à états couplés (Takagi-Sugeno) classiquement utilisé. Une modification de la structure originale du multi-modèle à états découplés est proposée afin d'éviter le phénomène de décrochage qui nuit considérablement à la procédure d'identification et détériore la qualité de l'approximation obtenue.*

Mots-clés— *Approche multi-modèle, identification systèmes non linéaires.*

I. INTRODUCTION

Les sciences de l'ingénieur font largement appel aux modèles non linéaires pour décrire les comportements dynamiques des systèmes physiques réels.

Si les modèles non linéaires sont en mesure de décrire correctement les comportements non linéaires d'un système, ils peuvent néanmoins s'avérer, en fonction de leur complexité mathématique, difficilement exploitables dans un contexte de synthèse d'une loi de commande et/ou de mise en place d'une stratégie de diagnostic du système.

Une hypothèse contraignante, mais largement utilisée, consiste à supposer que le système évolue autour d'un point de fonctionnement. Il est alors possible d'envisager une étape de linéarisation du modèle non linéaire afin de réduire sa complexité mathématique et permettre l'emploi des outils d'analyse de systèmes linéaires. En pratique, cependant, cette hypothèse de linéarité n'est pas toujours respectée et par conséquent le modèle linéarisé n'est pas complètement représentatif du comportement global du système. Une amélioration consiste, lorsque le point de fonctionnement change, à réactualiser le modèle linéarisé.

Dans ce qui suit, on préfère rechercher un modèle apte à donner une bonne caractérisation globale du comportement dynamique du procédé, tout en permettant au concepteur une utilisation aisée, par exemple au moyen de techniques d'analyse de systèmes linéaires.

Afin de répondre à ces attentes, de nouvelles techniques de modélisation ont vu le jour. Parmi elles, figure l'approche multi-modèle.

Nous proposons, dans cet article, une procédure d'identification *hors ligne* des systèmes non linéaires par une approche multi-modèle à états locaux découplés. La section 2 rappelle différentes techniques de modélisation de systèmes non linéaires en rapport avec les multi-modèles. La section 3 présente deux structures de multi-modèles, la première à états couplés, la seconde à états découplés. La procédure d'estimation paramétrique du multi-modèle à états découplés est développée dans la section 4. La section 5 propose un exemple académique d'identification qui met en évidence les limites du multi-modèle à états découplés. Une modification de la structure du multi-modèle est suggérée dans la section 6 afin d'améliorer les performances du multi-modèle identifié. Enfin, la section 7 présente un exemple d'identification du four à gaz proposé par *Box et Jenkins* [2].

II. LES MULTI-MODÈLES

La stratégie *diviser pour régner* est à l'origine de nombreuses techniques de modélisation de systèmes non linéaires (S.N.L.). Le point commun à ces techniques de modélisation est la décomposition du comportement dynamique du système non linéaire en un nombre L de zones de fonctionnement, chaque zone étant caractérisée par un sous-modèle f_i. En fonction de la zone où le système non linéaire évolue, la sortie $\hat{y}_i$ de chaque sous-modèle contribue plus ou moins à l'approximation $\hat{y}$ du comportement global du système non linéaire, soit :

$$\hat{y}(k) = \sum_{i=1}^{L} \mu_i(k)\hat{y}_i(k), \tag{1}$$

où la contribution de chaque sous-modèle est définie par les fonctions de pondération μ_i.

On peut distinguer parmi les différentes techniques de modélisation des systèmes non linéaires basées sur ce même principe (voir [4], [13], [14], pour une discussion plus approfondie) :

Des modèles linéaires par morceaux. Les sous-modèles f_i sont de type linéaire et les fonctions de pondération μ_i sont de type booléen. Le résultat est une approximation discontinue du système non linéaire, dans les phases de commutation, qui peut être indésirable dans certaines applications.

Des réseaux de fonctions à base radiale. Les sous-modèles f_i se réduisent à une constante w_i et les fonctions de pondération sont en général de type gaussien. L'inconvénient principal de cette approche est le nombre important de sous-modèles f_i nécessaires à l'obtention d'une approximation satisfaisante du système non linéaire.

Des modèles flous de Takagi-Sugeno (T.S.) [15]. La modélisation est effectuée à partir de règles de type : **si** *prémisse* **alors** *conséquence*, où les prémisses sont obtenues à partir de propositions linguistiques permettant l'évaluation des pondérations μ_i, les conséquences sont le plus souvent des fonctions affines qui correspondent aux f_i.

Des multi-modèles (M.M.) [9]. Les notions de multi-modèle et de modèle T.S. sont très proches. En effet, si le nombre de règles est égal au nombre de sous-modèles, ces deux approches sont identiques. La différence provient alors du moyen employé pour obtenir les fonctions de pondération μ_i :
- à partir de la connaissance d'experts et des propositions linguistiques pour les modèles T.S.,
- à l'aide de techniques d'optimisation dans le cas des multi-modèles [4], [8], [10].

Les principaux intérêts de l'approche multi-modèle sont les suivants [11] :
- ils constituent des *approximateurs* universels, n'importe quel système non linéaire pouvant être approximé avec une précision imposée en augmentant le nombre de sous-modèles. En pratique cependant, un nombre relativement réduit de sous-modèles suffit à l'obtention d'une approximation satisfaisante ;
- les outils d'analyse des systèmes linéaires peuvent être utilisés, au moins partiellement, sur les multi-modèles si les sous-modèles sont de type linéaire ;
- il est possible de relier le multi-modèle à la physique du système non linéaire afin de donner un sens au multi-modèle et plus précisément d'associer à un sous-modèle un comportement particulier du système non linéaire.

III. Structures des multi-modèles

Plusieurs structures permettent d'interconnecter les différents sous-modèles afin de générer la sortie globale du multi-modèle. Deux structures de multi-modèle peuvent être distinguées, la première à états couplés, la seconde à états découplés.

A. Multi-modèle à états couplés

La structure à états couplés ou modèle de Takagi-Sugeno est la plus répandue en analyse et en synthèse des multi-modèles. Elle se présente sous la forme suivante (figure 1) :

$$\begin{aligned} x_i(k+1) &= A_i\hat{x}(k) + B_i u(k) + D_i, \\ \hat{x}(k+1) &= \sum_{i=1}^{L} \mu_i(\xi(k)) x_i(k+1), \\ \hat{y}(k) &= \sum_{i=1}^{L} \mu_i(\xi(k)) C_i \hat{x}(k), \end{aligned} \tag{2}$$

$$\text{avec} \qquad \sum_{i=1}^{L} \mu_i(\xi(k)) = 1, \qquad \forall k \tag{3}$$

où $\hat{x} \in \mathbb{R}^n$ est le vecteur d'état, $u \in \mathbb{R}^m$ le vecteur de commande, $\hat{y} \in \mathbb{R}^l$ le vecteur de mesures et ξ la variable d'indexation des fonctions de pondération μ_i. Ici le rôle des fonctions de pondération μ_i est d'effectuer **un mélange des paramètres des sous-modèles** en fonction de la zone de fonctionnement du système non linéaire. Le multi-modèle est analogue à un système à paramètres variables dans le temps :

$$\begin{aligned} \hat{x}(k+1) &= \tilde{A}(k)\hat{x}(k) + \tilde{B}(k)u(k) + \tilde{D}(k), \\ \hat{y}(k) &= \tilde{C}(k)\hat{x}(k), \end{aligned} \tag{4}$$

où $\tilde{A}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))A_i$, $\tilde{B}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))B_i$, $\tilde{D}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))D_i$ et $\tilde{C}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))C_i$.

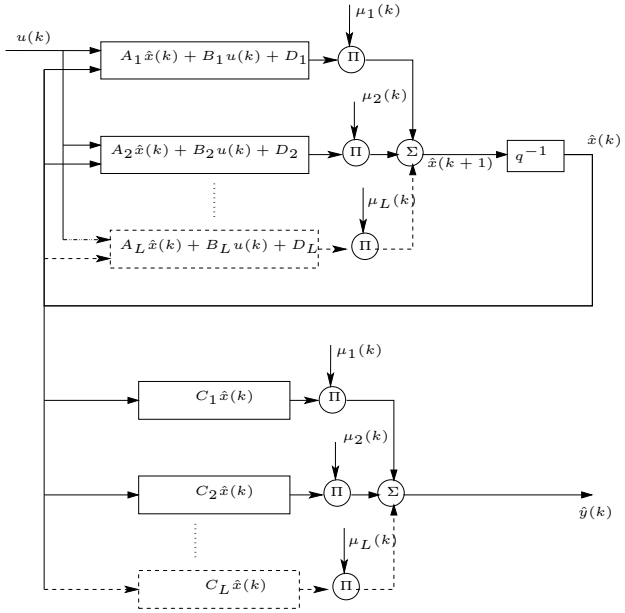

Fig. 1. Architecture d'un multi-modèle à états couplés

Le multi-modèle à états couplés se caractérise par :
- un état global $\hat{x}$ qui couple tous les états x_i des sous-modèles,
- un même ordre de sous-modèles,
- un mélange des paramètres des sous-modèles.

B. Multi-modèle à états découplés

Filev [3] propose un multi-modèle issu de l'agrégation des sous-modèles sous la forme d'une structure à états découplés (figure 2) :

$$\begin{aligned} \hat{x}_i(k+1) &= A_i\hat{x}_i(k) + B_i u(k) + D_i, \\ \hat{y}_i(k) &= C_i\hat{x}_i(k), \\ \hat{y}(k) &= \sum_{i=1}^{L} \mu_i(\xi(k))\hat{y}_i(k), \end{aligned} \tag{5}$$

$$\text{avec} \qquad \sum_{i=1}^{L} \mu_i(\xi(k)) = 1, \qquad \forall k \tag{6}$$

où $\hat{x}_i \in \mathbb{R}^{n_i}$ est le vecteur d'état du i^{ème} sous-modèle, $u \in \mathbb{R}^m$ le vecteur de commande, $\hat{y} \in \mathbb{R}^l$ le vecteur de

mesures. Le rôle des fonctions μ_i est ici de **pondérer la sortie de chaque sous-modèle** sans mélanger les paramètres des sous-modèles A_i, B_i, D_i et C_i.

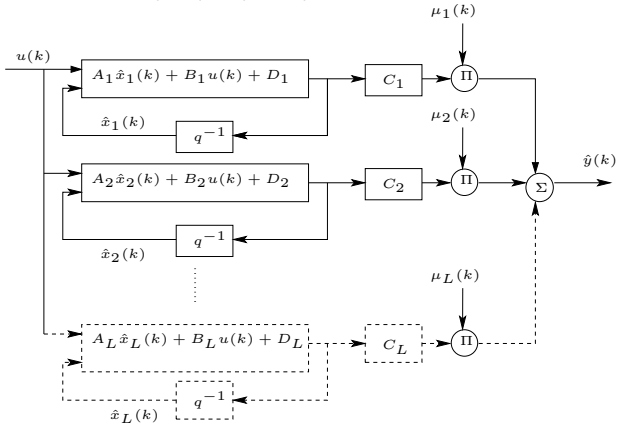

Fig. 2. Architecture d'un multi-modèle à états découplés

Il convient de constater que dans ce cas :
- le multi-modèle est donné par la mise en parallèle de modèles de type Wiener (un sous-modèle linéaire en série avec une non-linéarité),
- la sortie globale du multi-modèle est donnée par la somme pondérée des sorties locales,
- chaque sous-modèle possède son propre espace d'état et y évolue indépendamment en fonction du signal de commande et de son état initial.

C'est le découplage entre les sous-modèles qui motive l'intérêt de cette structure. Il est en effet possible d'envisager :
- de transposer plus facilement les techniques d'analyse de systèmes linéaires au multi-modèle,
- d'exploiter plus aisément le multi-modèle,
- d'utiliser des sous-modèles à structure complètement différente. Par exemple, de type linéaire ou non linéaire comportant un nombre d'états différent (à condition toutefois que les sorties soient sommables).

IV. Estimation paramétrique

Classiquement trois problèmes doivent être résolus pour élaborer un multi-modèle :

1. la décomposition du comportement dynamique du système non linéaire en un nombre L de zones de fonctionnement et le choix de la variable d'indexation ξ des fonctions de pondération,
2. la procédure d'agrégation des sous-modèles, c'est-à-dire le choix de la structure du multi-modèle,
3. la détermination des paramètres de chaque sous-modèle.

Il existe une abondante littérature sur l'identification des systèmes non linéaires par une approche multi-modèle à états couplés. *Johansen et Foss* [10] utilisent un algorithme pour effectuer une décomposition en zones de fonctionnement du système non linéaire suivant une certaine heuristique. A chaque itération, des sous-modèles sont ajoutés en fonction de la complexité du système non linéaire (approche ascendante). *Gasso* [4] propose une technique de décomposition de l'espace de fonctionnement du système non linéaire à travers une partition grille. La complexité du multi-modèle est réduite, d'une part, en éliminant les sous-modèles peu explicatifs et, d'autre part, en fusionnant des sous-modèles redondants (approche descendante). *Gugaliya et al.* [8] suggèrent une modification de l'algorithme CART (Classification And Regression Tree) qui donne à l'origine une approximation par morceaux du système non linéaire afin d'obtenir un multi-modèle.

En revanche, peu de travaux proposent une procédure d'identification basée sur une structure de multi-modèle à états découplés. Quelques travaux dans le domaine de la synthèse des lois de commande pour des systèmes non linéaires ont déjà exploité cette structure et ont démontré son intérêt [5], [6], [7]. Le multi-modèle est obtenu en effectuant des linéarisations successives du modèle non linéaire autour de différents points de fonctionnement. Récemment *Venkat et al.* [16] ont proposé une méthodologie d'identification basée sur cette structure en vue de mettre en place une loi de commande. L'identification des sous-modèles est effectuée individuellement à partir des données d'entrée/sortie du système non linéaire générées dans une zone de fonctionnement. L'idée est d'assurer ainsi la linéarité des sous-modèles à identifier. Il est donc nécessaire de disposer d'autant de jeux de données que de sous-modèles à identifier (chaque jeu de données étant propre à un sous-modèle donc à un mode de fonctionnement).

Nous proposons dans cet article une procédure d'estimation paramétrique du multi-modèle à états découplés. Le problème 1 énoncé au début de la section (décomposition en zones de fonctionnement du système non linéaire) est contourné en se fixant a priori les fonctions de pondération μ_i et la variable d'indexation ξ. La complexité du problème à traiter est ainsi réduite. Un choix simple mais tout à fait raisonnable consiste, par exemple, à effectuer une distribution homogène des fonctions de pondération sur la plage de variation de la variable d'indexation ξ.

Le problème lié à l'identification se pose alors dans les termes suivants : les fonctions de pondération $\mu_i(\xi)$ étant fixées a priori, il s'agit, à partir de la connaissance de l'entrée $u(k)$ et de la sortie $y(k)$ d'un système non linéaire SISO, d'identifier les paramètres de L sous-modèles.

A. Fonctions de pondération μ_i

Les fonctions de pondération $\mu_i(\xi)$ sont construites ici à partir d'une fonction gaussienne :

$$\omega_i(\xi) = \exp\left(-\frac{(\xi - c_i)^2}{\sigma^2}\right), \tag{7}$$

qui est fonction du centre c_i et de la dispersion σ commune à toutes les fonctions de pondération (d'autres choix peuvent être faits, mais cette présentation n'a pas pour objet de comparer ces différentes possibilités). Afin de respecter la contrainte donnée par la relation (3) les fonctions $\omega_i(\xi)$ sont ensuite normalisées. Finalement les fonctions de pondération sont données par :

$$\mu_i(\xi) = \frac{\omega_i(\xi)}{\sum_{j=1}^{L} \omega_j(\xi)}. \tag{8}$$

La variable d'indexation ξ des fonctions de pondération μ_i peut être, par exemple, définie par les variables d'état

mesurables et/ou les valeurs du signal d'entrée ou de sortie du système.

En fonction de la vitesse de transition d'un sous-modèle à l'autre et du recouvrement (zone où deux fonctions de pondération sont significativement non nulles en même temps), on peut introduire les notions de fonctions de pondération *fortement mélangées* et *peu mélangées*. La figure 3 illustre ces notions sur un exemple. Le cas limite, vitesse de transition infinie et recouvrement nul, correspond au cas des modèles linéaires par morceaux où les fonctions de pondération sont booléennes.

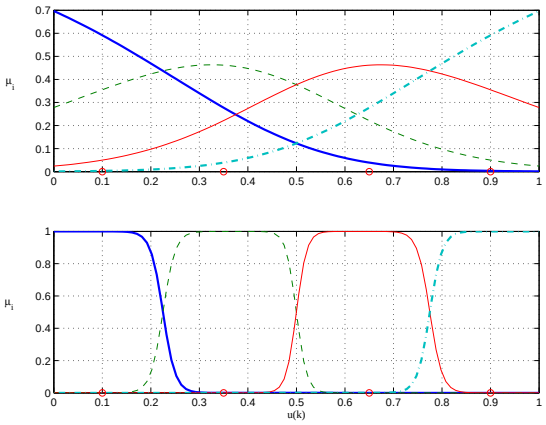

Fig. 3. Fonctions de pondération fortement mélangées $\sigma = 0.35$ (haut), fonctions de pondération peu mélangées $\sigma = 0.08$ (bas)

B. Critères d'estimation

Le multi-modèle doit répondre à un certain nombre de contraintes en fonction de son contexte d'exploitation. A cet effet, trois critères : global, local ou combiné, sont utilisés pour effectuer l'optimisation paramétrique.

B.1 Le critère global

Le critère global est défini par :

$$J_G = \frac{1}{2}\sum_{k=1}^{N}(\hat{y}(k) - y(k))^2, \tag{9}$$

où $\hat{y}(k)$ est la sortie du multi-modèle, $y(k)$ la sortie du système non linéaire et N le nombre de mesures. Ce critère favorise une bonne caractérisation du comportement global du système non linéaire par le multi-modèle. L'adéquation entre le comportement local des différents sous-modèles et le comportement local du système non linéaire n'est pas prise en considération. Il est intéressant, par exemple, dans un contexte d'identification type boîte noire où l'on souhaite avoir une très bonne adéquation entre les données d'entrée/sortie du système sans chercher une interprétation quelconque du modèle obtenu. On obtient *un multi-modèle comportemental ou/et de prédiction*. La forme (9) peut être assortie de pondérations afin de privilégier éventuellement le poids de certaines données.

B.2 Le critère local

Le critère local donné par :

$$J_L = \frac{1}{2}\sum_{i=1}^{L}\sum_{k=1}^{N}\mu_i(\xi(k))(\hat{y}_i(k) - y(k))^2, \tag{10}$$

favorise une bonne adéquation entre le comportement local des sous-modèles et le comportement local du système non linéaire à condition toutefois que les μ_i soient *peu mélangées*. Il est très bien adapté à l'obtention *d'un multi-modèle phénoménologique et/ou explicatif*. Il est possible en effet de donner une interprétation des sous-modèles obtenus, par exemple, comme étant des linéarisés tangents au système non linéaire. Cependant, par rapport au critère global, un nombre de sous-modèles supérieur est en général nécessaire à l'obtention d'une bonne caractérisation du comportement global du système [4].

B.3 Le critère mixte

Finalement, le critère combiné ou mixte défini par [18] :

$$J_C = \alpha J_G + (1-\alpha)J_L \qquad 0 \leq \alpha \leq 1, \tag{11}$$

représente un compromis entre les deux critères précédents. Il est ainsi possible de profiter des avantages offerts par chacun des critères en fonction de la valeur donnée au paramètre α.

C. Procédure d'estimation paramétrique

Dans cette section sont présentés les points remarquables de la procédure d'identification mise en place pour effectuer l'estimation paramétrique du multi-modèle à états locaux découplés. Pour davantage de renseignements sur les techniques classiques d'identification, le lecteur peut consulter par exemple [17].

On définit le vecteur θ comme étant le vecteur colonne des paramètres du multi-modèle à estimer. Le vecteur θ est représenté sous la forme d'un vecteur partitionné en L *blocs colonne* θ_p, soit :

$$\theta = [\theta_1 \ldots \theta_p \ldots \theta_L]^T, \tag{12}$$

où chaque bloc colonne θ_p est relatif aux q_p paramètres inconnus d'un sous-modèle particulier p, soit :

$$\theta_p = [\theta_{p,1} \ldots \theta_{p,q} \ldots \theta_{p,q_p}]^T, \tag{13}$$

où $\theta_{p,q}$ est un paramètre (scalaire) à estimer appartenant au sous-modèle p.

L'estimation paramétrique des $\theta_{p,q}$ des sous-modèles, utilisée ici, est basée sur une procédure itérative de minimisation d'un critère J (global, local ou combiné) à l'aide de l'algorithme de Gauss-Newton :

$$\theta^+ = \theta - \mathrm{H}^{-1}\mathrm{G}, \tag{14}$$

où θ représente la valeur du vecteur de paramètres à une itération particulière, θ^+ cette même valeur à l'itération suivante,

$\mathrm{H} = \frac{\partial^2 \mathrm{J}}{\partial\theta\partial\theta^{\mathrm{T}}}$ la matrice hessienne et $\mathrm{G} = \frac{\partial \mathrm{J}}{\partial\theta}$ le vecteur gradient. Le calcul du vecteur gradient G et de la matrice hessienne H est effectué à partir du calcul des *fonctions de sensibilité* de la sortie du multi-modèle par rapport aux paramètres de chaque sous-modèle.

Quelques améliorations peuvent être apportées à la procédure initiale d'identification afin d'assurer la convergence de l'algorithme :

1. Un *coefficient de relaxation* Δ est introduit au niveau de l'équation (14) :

$$\theta^+ = \theta - \Delta \mathrm{H}^{-1}\mathrm{G}. \tag{15}$$

   Δ est le pas qui minimise le critère dans la direction du vecteur $\mathrm{H}^{-1}\mathrm{G}$. Idéalement et pour un critère de type quadratique, $\Delta = 1$. En pratique, on peut être amené à modifier sa valeur afin d'améliorer la vitesse de convergence de l'algorithme. En effet, si H est défini positif alors il existe toujours un $\Delta \leq 1$ qui minimise le critère dans la direction du vecteur $\mathrm{H}^{-1}\mathrm{G}$.

2. L'algorithme de Marquardt [12] est utilisé afin de surmonter les problèmes liés à l'inversion de la matrice H (provenant d'un mauvais conditionnement par exemple) lors de la mise à jour des paramètres. L'équation (15) est remplacée par :

$$\theta^+ = \theta - \Delta(\mathrm{H} + \lambda \mathrm{I})^{-1}\mathrm{G}, \tag{16}$$

   où I est la matrice identité de dimension appropriée et λ un scalaire (paramètre de régularisation). A chaque itération la valeur de λ est réglée, le plus souvent, au moyen d'une heuristique (voir [1], [17] pour plus de détails). Si la valeur de λ est proche de zéro alors l'algorithme est proche de celui de Gauss-Newton. Si la valeur de $\lambda >> 0$ tend vers l'infini alors l'algorithme est proche de celui du gradient.

3. Un filtrage passe-bas des nouveaux paramètres θ^+ est introduit :

$$\theta^+ = p\theta^+ + (1-p)\theta \text{ avec } 0 \leq p \leq 1,$$

   afin de limiter d'éventuelles oscillations des paramètres.

Les trois degrés de liberté, donnés par les paramètres Δ, λ et p, permettent d'améliorer la convergence de la procédure d'identification.

Une des principales difficultés de cette approche porte sur la sensibilité de l'algorithme par rapport au choix des paramètres initiaux. En effet, rien n'assure que le problème soit de type convexe et que l'algorithme converge vers un minimum global. Le succès de l'optimisation dépend d'un choix judicieux des paramètres initiaux ; dans les cas difficiles, il peut être nécessaire de réaliser la procédure d'estimation à partir de différents choix de paramètres initiaux.

D. Estimation paramétrique avec un critère global

Le critère global est défini par :

$$J_G = \frac{1}{2}\sum_{k=1}^{N} \varepsilon(k)^2,$$

où $\varepsilon(k)$ est l'erreur entre la sortie complète du M.M. et la sortie du S.N.L., donnée par :

$$\varepsilon(k) = \hat{y}(k) - y(k). \tag{17}$$

D.1 Le vecteur gradient $\mathrm{G_G}$

Le vecteur gradient $\mathrm{G_G}$ s'obtient en dérivant le critère global par rapport aux paramètres θ, soit :

$$\mathrm{G}_G = \frac{\partial J_G}{\partial\theta} = \sum_{k=1}^{N} \varepsilon(k)\frac{\partial \hat{y}(k)}{\partial\theta}, \tag{18}$$

avec avec

$$\frac{\partial \hat{y}(k)}{\partial\theta} = \sum_{i=1}^{L} \mu_i(\xi(k))\frac{\partial \hat{y}_i(k)}{\partial\theta}. \tag{19}$$

Les $\frac{\partial \hat{y}_i(k)}{\partial\theta}$ sont les fonctions de sensibilité de la i^{ème} sortie par rapport aux paramètres inconnus du multi-modèle. Elles sont définies par :

$$\frac{\partial \hat{y}_i(k)}{\partial\theta_{p,q}} = \frac{\partial C_i}{\partial\theta_{p,q}}\hat{x}_i(k) + C_i\frac{\partial \hat{x}_i(k)}{\partial\theta_{p,q}}. \quad \begin{array}{l} p = 1,2,...,L \\ q = 1,2,...,q_p \end{array} \tag{20}$$

L'état étant inconnu à l'instant présent, les fonctions de sensibilité sont calculées à l'instant suivant en dérivant la relation (5) par rapport à chaque paramètre $\theta_{p,q}$, c'est-à-dire :

$$\frac{\partial \hat{x}_i(k+1)}{\partial\theta_{p,q}} = \frac{\partial A_i}{\partial\theta_{p,q}}\hat{x}_i(k) + A_i\frac{\partial \hat{x}_i(k)}{\partial\theta_{p,q}} + \frac{\partial B_i}{\partial\theta_{p,q}}u(k) + \frac{\partial D_i}{\partial\theta_{p,q}}. \tag{21}$$

D.2 La matrice hessienne $\mathrm{H_G}$

Le matrice hessienne $\mathrm{H_G}$ s'obtient en dérivant deux fois le critère global par rapport aux paramètres θ, soit :

$$\mathrm{H}_G = \frac{\partial^2 J_G}{\partial\theta\partial\theta^T} = \sum_{k=1}^{N} \underbrace{\varepsilon(k)}_{\to 0}\frac{\partial^2 \hat{y}(k)}{\partial\theta\partial\theta^T} + \sum_{k=1}^{N}\frac{\partial \hat{y}(k)}{\partial\theta}\frac{\partial \hat{y}(k)}{\partial\theta^T}. \tag{22}$$

L'utilisation de la *méthode de Gauss-Newton* évite le calcul explicite de la matrice $\mathrm{H_G}$ réduisant ainsi le temps de calcul. En effet, les dérivées du deuxième ordre dans l'expression (22) sont négligées sous l'hypothèse que l'erreur $\varepsilon(k)$ tende vers zéro. Le calcul la matrice $\mathrm{H_G}$ se fait alors à l'aide des fonctions de sensibilité du premier ordre nécessaires au calcul du gradient.

La matrice $\mathrm{H_G}$ approchée est donnée par :

$$\mathrm{H}_G \approx \sum_{k=1}^{N}\frac{\partial \hat{y}(k)}{\partial\theta}\frac{\partial \hat{y}(k)}{\partial\theta^T}. \tag{23}$$

E. Estimation paramétrique avec un critère local

Le critère local est de la forme :

$$J_L = \frac{1}{2}\sum_{i=1}^{L}\sum_{k=1}^{N} \mu_i(\xi(k))\varepsilon_i(k)^2,$$

où $\varepsilon_i(k)$ est l'erreur entre la sortie du i^{ème} sous-modèle et la sortie du S.N.L., donnée par :

$$\varepsilon_i(k) = \hat{y}_i(k) - y(k). \tag{24}$$

E.1 Le vecteur gradient $\mathrm{G_L}$

Le vecteur gradient $\mathrm{G_L}$ est défini par :

$$\mathrm{G}_L = \frac{\partial J_L}{\partial \theta} = \sum_{i=1}^{L}\sum_{k=1}^{N} \mu_i(\xi(k))\varepsilon_i(k)\frac{\partial \hat{y}_i(k)}{\partial \theta}. \qquad (25)$$

Il convient de remarquer que les fonctions de sensibilité $\frac{\partial \hat{y}_i(k)}{\partial \theta}$ sont les mêmes que dans le cas d'un critère global (relation (20)).

E.2 La matrice hessienne $\mathrm{H_L}$

La matrice hessienne $\mathrm{H_L}$ s'obtient en dérivant deux fois le critère local par rapport aux paramètres θ, soit :

$$\mathrm{H}_L \approx \sum_{i=1}^{L}\sum_{k=1}^{N} \mu_i(\xi(k))\frac{\partial \hat{y}_i(k)}{\partial \theta}\frac{\partial \hat{y}_i(k)}{\partial \theta^T}, \qquad (26)$$

sous l'hypothèse que localement l'erreur $\varepsilon_i(k)$ tende vers zéro (méthode de Gauss-Newton).

F. Comparaison entre l'approche globale et locale

Le critère de comparaison est la ressemblance entre les vecteurs gradients et les matrices hessiennes obtenus dans chaque approche.

F.1 Comparaison des vecteurs $\mathrm{G_G}$ et $\mathrm{G_L}$

La différence entre le gradient global (18) et le gradient local (25) est le type d'erreur utilisée : dans le premier cas, il s'agit de l'erreur globale $\varepsilon(k)$ (erreur entre la sortie complète du M.M. et la sortie du S.N.L.), dans le deuxième cas de l'erreur locale $\varepsilon_i(k)$ (erreur entre la i^{ème} sortie locale du M.M. et la sortie du S.N.L.).

Si les fonctions de pondération sont peu mélangées (hypothèse légitime si l'on souhaite conserver un sens local à chaque sous-modèle dans J_L), alors : $\varepsilon_i(k) \approx \varepsilon(k)$. Il est possible de conclure que :

$$\mathrm{G_L} \approx \mathrm{G_G}. \qquad (27)$$

F.2 Comparaison des matrices $\mathrm{H_G}$ et $\mathrm{H_L}$

En utilisant (19), l'expression de la matrice Hessienne $\mathrm{H_G}$ (23) devient :

$$\mathrm{H}_G \approx \sum_{k=1}^{N}\sum_{i=1}^{L}\sum_{j=1}^{L} \mu_i(\xi(k))\mu_j(\xi(k))\frac{\partial \hat{y}_i(k)}{\partial \theta}\frac{\partial \hat{y}_j(k)}{\partial \theta^T}. \qquad (28)$$

Si les fonctions de pondération μ_i sont peu mélangées alors :

$$\mu_i(\xi(k))\mu_j(\xi(k)) \approx \begin{cases} 1 & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j \end{cases}.$$

On peut alors conclure que :

$$\mathrm{H_L} \approx \mathrm{H_G}. \qquad (29)$$

En conclusion, si l'on choisit des fonctions de pondération μ_i peu mélangées, la distinction entre critère global et critère local disparaît ; les deux approches d'identification fournissent alors des résultats similaires.

En revanche, si les fonctions de pondération sont fortement mélangées, on pourra choisir le critère mixte (11) afin de pondérer l'importance donnée à l'interprétation des sous-modèles en regard de la qualité du modèle global.

Remarque

Les expressions (19), (20) et (21) relatives au calcul des fonctions de sensibilité sont des formes génériques qui peuvent être simplifiées lors de la mise en œuvre de l'algorithme. En effet,

$$\frac{\partial \hat{y}_i(k)}{\partial \theta_p} = 0 \quad \text{pour} \quad p \neq i \begin{array}{l} i = 1, 2, ..., L \\ p = 1, 2, ..., L \end{array},$$

du fait que les paramètres de chaque sous-modèle sont complètement indépendants des paramètres des autres sous-modèles.

V. Exemple d'identification

Soit à approcher, par une structure multi-modèle, le modèle non linéaire :

$$\begin{aligned} x(k+1) &= Ax(k) + \sin(\gamma u(k))(\beta - u(k)), \\ y(k) &= x(k), \end{aligned} \qquad (30)$$

avec $A = 0.95$, $\gamma = 0.8\pi$ et $\beta = 1.5$.

Nous avons choisi ici d'utiliser un critère global pour effectuer l'identification du multi-modèle. On souhaite obtenir un multi-modèle de prédiction et non pas d'explication de comportements locaux. L'estimation paramétrique est effectuée à partir de la relation (16) en considérant le vecteur gradient $\mathrm{G_G}$ et la matrice hessienne $\mathrm{H_G}$ définis par les relations (18) et (23).

Le multi-modèle est constitué arbitrairement par $L = 6$ sous-modèles (cette valeur pouvant être optimisée). Les paramètres A_i, B_i, D_i et C_i des sous-modèles sont de type scalaire. Les fonctions de pondération μ_i dépendent du signal d'entrée $u(k)$, les centres sont donnés par : $c_i = [0, 0.2, 0.4, 0.6, 0.8, 1]$ et la dispersion $\sigma = 0.2$.

A. Multi-modèle obtenu par linéarisation

Un premier multi-modèle peut être obtenu en effectuant des linéarisations successives du modèle non linéaire (30) autour des points de fonctionnement c_i. Le comportement dynamique du multi-modèle est illustré sur la figure 4.

L'erreur en régime statique du multi-modèle provient en particulier du fait que les fonctions de pondération sont fortement mélangées. Le caractère local de chaque sous-modèle n'est pas conservé. On peut constater un phénomène dit de *décrochage* qui se traduit par l'apparition de "pics" dans les zones de commutation. Cependant les paramètres obtenus par linéarisation peuvent servir de valeurs initiales à une autre procédure d'identification.

B. Multi-modèle obtenu par identification

L'entrée $u(k)$ du modèle est constituée par la concaténation de créneaux d'amplitudes variables ($u \in [0, 1]$). Un jeu de données entrée/sortie de 5000 points a servi à l'identification. Un jeu de données entrée/sortie de même taille a servi à

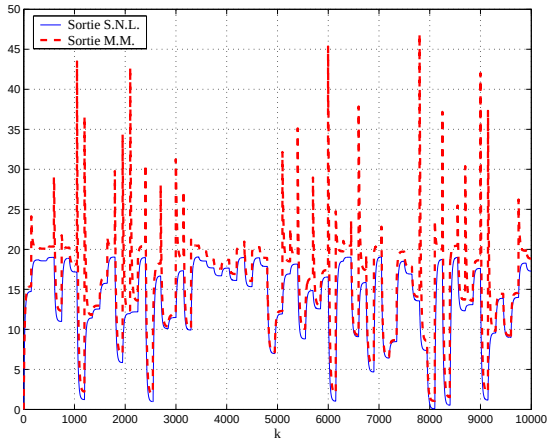

Fig. 4. Sortie du modèle non linéaire et du multi-modèle à états découplés obtenu par linéarisation

la validation. Les paramètres initiaux du multi-modèle sont les paramètres du multi-modèle obtenu par linéarisation. Le comportement dynamique du multi-modèle identifié est illustré sur la figure 5.

Le comportement dynamique du modèle non linéaire (30) est assez mal approché par le multi-modèle identifié. L'apparition de "pics", dans certaines zones de commutation du signal de commande, détériore considérablement la qualité de l'approximation obtenue. Ces "pics" sont une conséquence du phénomène dit de *décrochage*. Ce phénomène de décrochage agit comme un élément perturbateur dans la procédure d'identification. On présente dans la section suivante quelques éléments expliquant ce phénomène et une solution pour y remédier.

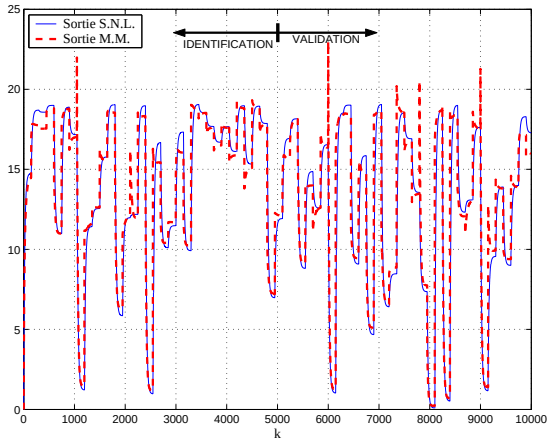

Fig. 5. Sortie du modèle non linéaire et du multi-modèle à états découplés obtenu par identification

VI. Modification du Multi-Modèle

Dans le cas du multi-modèle à états découplés rien n'assure que la sortie de chaque sous-modèle soit proche des autres quand elle est mise à contribution (ce problème ne se pose pas pour le multi-modèle à états couplés). C'est cet écart entre les sorties qui est à l'origine du phénomène de décrochage, se traduisant quelquefois par des variations brutales de la sortie.

En effet, un changement de la valeur de la variable d'indexation ξ (ici le signal de commande) peut entraîner un changement de la pondération de la sortie de chaque sous-modèle. Un décrochage apparaît si les sorties mises à contribution sont éloignées les unes des autres. Si les sorties des sous-modèles sont proches, au moment où elles sont mises à contribution, le décrochage diminue et disparaît complètement si elles sont identiques.

Le problème de décrochage peut être posé comme étant un problème de conditions initiales sur les sorties des sous-modèles quand elles sont mises à contribution.

En présence de ce phénomène, la procédure d'identification doit calculer les paramètres du multi-modèle qui approchent au mieux le comportement du S.N.L. tout en minimisant l'écart entre les sorties des sous-modèles. Le décrochage agit comme un élément perturbateur dans la procédure d'identification dans la mesure où quelques-uns des sous-modèles estimés peuvent présenter une faible variation de leur sortie vis-à-vis du signal de commande (comportement très gênant pour un modèle explicatif). Il est en effet nécessaire de limiter la variation de la sortie des sous-modèles pour obtenir une diminution du phénomène de décrochage. En revanche, le multi-modèle obtenu offre une mauvaise approximation du système non linéaire dans les zones de décrochage.

Une première solution pour éliminer le phénomène de décrochage consiste simultanément à :

– mélanger fortement les fonctions de pondération,
– augmenter le nombre des sous-modèles.

Nous proposons une deuxième solution qui consiste à introduire dans la structure du multi-modèle trois filtres passe-bas agissant respectivement sur le signal de commande, le signal d'indexation de μ_i et la sortie du multi-modèle (figure 6). L'objectif est de prendre progressivement en compte la contribution de chaque sous-modèle et d'adoucir la commutation entre les sorties des sous-modèles. Ces trois filtres font alors partie du modèle du système.

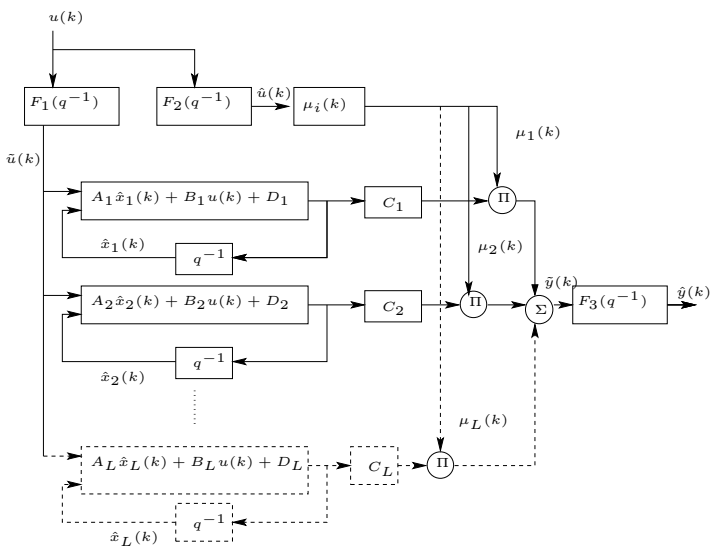

Fig. 6. Architecture modifiée du multi-modèle

Le nouveau critère à minimiser s'écrit :

$$J_G = \frac{1}{2}\sum_{k=1}^{N}(\hat{y}(k) - y(k))^2, \tag{31}$$

le modèle du système étant défini par :

$$\begin{aligned}
\hat{x}_i(k+1) &= A_i\hat{x}_i(k) + B_i\tilde{u}(k) + D_i, \\
\hat{y}_i(k) &= C_i\hat{x}_i(k), && (32)\\
\tilde{y}(k) &= \sum_{i=1}^{L}\mu_i(\xi(k))\hat{y}_i(k), \text{ avec } \xi(k) = \hat{u}(k), \\
\tilde{u}(k) &= F_1(q^{-1})u(k), && (33)\\
\hat{u}(k) &= F_2(q^{-1})u(k), && (34)\\
\hat{y}(k) &= F_3(q^{-1})\tilde{y}(k), && (35)
\end{aligned}$$

où $\hat{y}$ est la nouvelle sortie du multi-modèle.
On opte ici pour des filtres F_1, F_2 et F_3 du premier ordre :

$$\begin{aligned}
\tilde{u}(k+1) &= \alpha_1\tilde{u}(k) + (1-\alpha_1)u(k+1), && (36)\\
\hat{u}(k+1) &= \alpha_2\hat{u}(k) + (1-\alpha_2)u(k+1), && (37)\\
\hat{y}(k+1) &= \alpha_3\hat{y}(k) + (1-\alpha_3)\tilde{y}(k+1). && (38)
\end{aligned}$$

Le calcul des nouvelles fonctions de sensibilité s'effectue comme dans la section IV-D en prenant en considération les nouveaux paramètres α_1, α_2 et α_3 des filtres passe-bas. Les paramètres des filtres font alors partie de l'ensemble des paramètres inconnus du multi-modèle qui sont estimés par la procédure d'optimisation paramétrique.
La figure 7 illustre le comportement dynamique du multi-modèle identifié en introduisant les trois filtres passe-bas. Les paramètres des filtres obtenus par la procédure d'optimisation sont donnés par :

$$\alpha_1 = 0.0018, \quad \alpha_2 = 0.1364 \quad \text{et} \quad \alpha_3 = 0.9516.$$

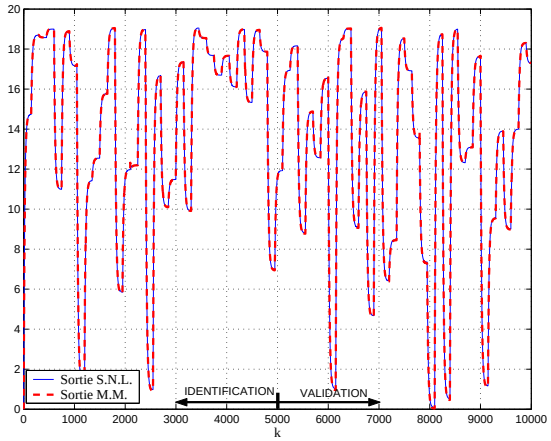

Fig. 7. Sortie du modèle non linéaire et du multi-modèle à états découplés obtenu par identification

Il est possible de constater la bonne adéquation entre les données du modèle non linéaire et celles du multi-modèle. Le phénomène de décrochage est rejeté. Pour cet exemple particulier les valeurs que prend le critère J_G, en fonction de l'approche utilisée pour obtenir le multi-modèle, sont présentées dans le tableau I.

Méthode	J_G
Linéarisation	7.99×10^4
Estimation sans filtrage	1.46×10^4
Estimation avec filtrage	21.72

TABLE I
ÉVOLUTION DU CRITÈRE EN FONCTION DE L'APPROCHE UTILISÉE

VII. EXEMPLE D'IDENTIFICATION

En guise d'exemple d'identification, nous utiliserons les données du four à gaz de type SISO proposées par *Box et Jenkins* [2]. Dans ce four à gaz, de l'air et du méthane ont été combinés de façon à obtenir un mélange de gaz contenant du CO_2. L'entrée du four correspond au débit du méthane et la sortie à la concentration en pourcentage de CO_2 (figure 8). Le jeu des données d'entrée/sortie du four est constitué de 296 points.

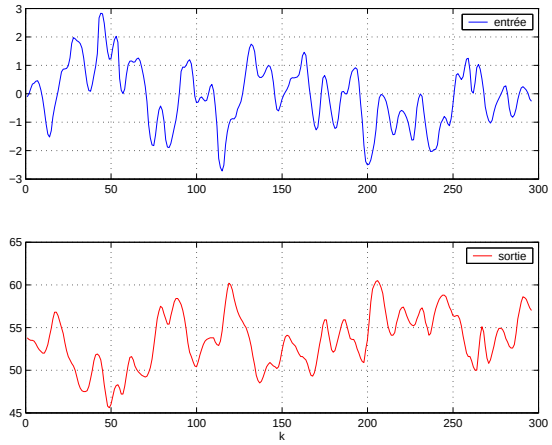

Fig. 8. Données d'entrée et de sortie du four

La procédure d'estimation a été effectuée avec un critère global en considérant un nombre différent de sous-modèles. Les paramètres A_i, B_i, D_i et C_i des sous-modèles sont de type scalaire. L'évolution du critère en fonction du nombre des sous-modèles ainsi que les différents paramètres des fonctions de pondération sont présentés dans le tableau II. La figure 9 montre les fonctions de pondération pour $L = 3$.

L	J_G	σ	c_i
1	952.28	- -	- -
2	126.77	3	$[-2.2, 2.2]$
3	108.25	1.5	$[-2.2, 0, 2.2]$
4	107.35	1	$[-2.2, -0.73, 0.73, 2.2]$

TABLE II
ÉVOLUTION DU CRITÈRE EN FONCTION DU NOMBRE DE SOUS-MODÈLES

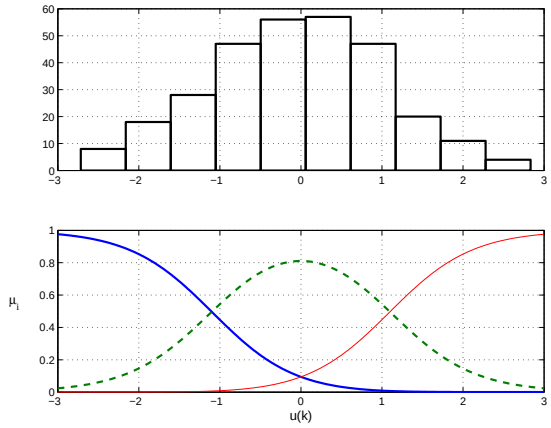

Fig. 9. Histogramme des données d'entrée du four (haut). Fonctions de pondération μ_i pour L=3 (bas)

La figure 10 illustre le comportement dynamique du système et du multi-modèle comportant $L = 3$ sous-modèles. Ce multi-modèle parvient à représenter globalement le comportement dynamique du système.

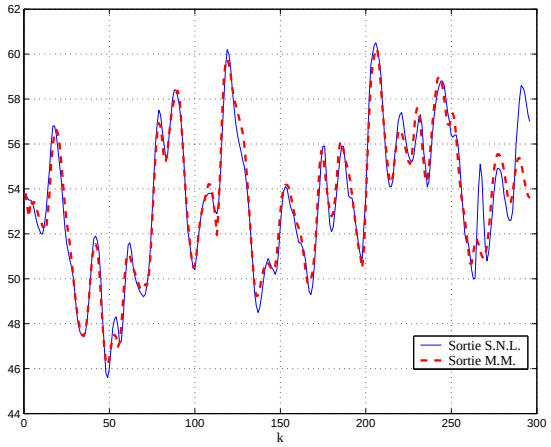

Fig. 10. Sortie du four et sortie du multi-modèle avec $L = 3$

Il y a cependant des régions où l'adéquation entre le multi-modèle et le système n'est pas satisfaisante.

En effet deux facteurs rendent l'identification difficile :

– un nombre relativement réduit de données,
– un signal d'entrée peu riche qui n'excite pas uniformément tous les modes du système (voir histogramme figure 9).

Il est possible d'améliorer l'approximation obtenue avec le multi-modèle en augmentant, par exemple, la dimension des sous-modèles.

VIII. Conclusion

Plusieurs structures permettent d'agréger des sous-modèles afin de construire un multi-modèle. Notre attention s'est ici portée sur le multi-modèle à états découplés par opposition au multi-modèle à états couplés (ou de Takagi-Sugeno) classiquement utilisé. Une procédure d'identification des systèmes non linéaires par une approche multi-modèle à états locaux découplés a été proposée dans cet article. L'intérêt du multi-modèle à états locaux découplés (par rapport au multi-modèle à états couplés) est d'avoir des sous-modèles complètement indépendants. Ce type de structure est a priori plus simple à manipuler (dans un contexte d'analyse du multi-modèle, de synthèse d'une loi de commande ou de diagnostic) avec des techniques d'analyse de systèmes linéaires. Un phénomène de décrochage qui diminue la qualité de l'approximation du multi-modèle a été mis en évidence. Une solution a été suggérée afin d'éliminer ce phénomène de décrochage. Elle consiste à introduire trois filtres passe-bas dans la procédure d'identification. Les résultats de l'identification ont été illustrés à travers un exemple d'identification d'un four à gaz.

Parmi les points pouvant constituer l'objet de travaux ultérieurs, on peut mentionner la recherche de la dimension du multi-modèle à états découplés, c'est-à-dire la détermination du nombre de sous-modèles et la dimension de ces sous-modèles. Il est possible également d'envisager une procédure d'optimisation des fonctions de pondération μ_i.

Références

[1] Y. Bard. *Nonlinear parameters estimation*. Academic Press, 1974.

[2] G.E.P. Box and G.M. Jenkins. *Time Series Analysis : Forecasting and Control*. San Francisco : Holden-Day, 1976.

[3] D. Filev. Fuzzy modeling of complex systems. *International Journal of Approximate Reasoning*, 5(3), pages 281–290, 1991.

[4] K. Gasso. *Identification des systèmes dynamiques non-linéaires : approche multi-modèle*. Thèse de doctorat, Institut National Polytechnique de Lorraine, 2000.

[5] E. P. Gatzke and F. J. Doyle III. Multiple model approach for CSTR control. In *IFAC World Congress*, pages 343–348, Beijing, P. R. China, 1999.

[6] P. J. Gawthrop. Continuous-time local state local model networks. In *IEEE Conference on Systems, Man & Cybernetics*, pages 852–857, Vancouver, Canada, 1995.

[7] G. Gregorcic and G. Lightbody. Control of highly nonlinear processes using self-tuning control and multiple/local model approaches. In *2000 IEEE International Conference on Intelligent Engineering Systems, INES 2000*, pages 167–171, Sep 2000.

[8] J. K. Gugaliya, R. D. Gudi, and S. Lakshminarayanan. Multi-model decomposition of nonlinear dynamics using a fuzzy-CART approach. *Journal of Process Control*, 15(4), pages 417–434, 2005.

[9] T. A. Johansen and A. B. Foss. Constructing narmax using armax. *Int. Journal of Control*, 58(5), pages 1125–1153, 1993.

[10] T. A. Johansen and B. A. Foss. Identification of non-linear system structure and parameters using regime decomposition. *Automatica*, 31(2), pages 321–326, 1995.

[11] T. A. Johansen, R. Shorten, and R. Murray-Smith. On the interpretation and identification of dynamic Takagi-Sugeno fuzzy models. *IEEE Trans. on Fuzzy Systems*, 8(3), pages 297–313, 2000.

[12] D.W. Marquardt. An algorithm for least-squares estimation of non-linear parameters. *J. Soc. Indust. Appl. Math*, 11(2), pages 431–441, 1963.

[13] R. Murray-Smith and T. A. Johansen. *Multiple model approaches to modelling and control*. Taylor & Francis, 1997.

[14] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification : a unified overview. *Automatica*, 31(12), pages 1691–1724, 1995.

[15] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Trans. on Systems Man and Cybernetics*, 15, pages 116–132, 1985.

[16] A. N. Venkat, P. Vijaysai, and R. D. Gudi. Identification of complex nonlinear processes based on fuzzy decomposition of the steady state space. *Journal of Process Control*, 13(6), pages 473–488, 2003.

[17] E. Walter and L. Pronzato. *Identification de modèles paramétriques à partir de données expérimentales*, chapitre 4, pages 123–196. MASSON, 1994.

[18] J. Yen, L. Wang, and Gillespie C. W. Improving the interpretability of Takagi-Sugeno fuzzy models by combining global learning and local learning. *IEEE Trans. on Fuzzy Systems*, 6(4), pages 530–537, 1998.