Contributions à l'identification de modèles paramétriques non linéaires. Application à la modélisation de bassins versants ruraux.

Vincent Laurain

Nancy-Université
Université Henri Poincaré

UFR – Faculté des Sciences et Technologies
École Doctorale IAEM Lorraine

CRAN

Contributions à l'identification de modèles paramétriques non linéaires. Application à la modélisation de bassins versants ruraux.

Thèse présentée pour l'obtention du

Doctorat de l'Université Henri Poincaré, Nancy 1

Spécialité Automatique, Traitement du Signal et Génie Informatique

par

Vincent Laurain

Soutenance publique le 19 octobre 2010

Président :	Luc Dugard	Directeur de Recherche CNRS au GIPSA de Grenoble
Rapporteurs :	Michel Gevers	Professeur à l'Université catholique de Louvain, Belgique
	Qinghua Zhang	Directeur de Recherche INRIA à l'INRIA de Rennes
Examinateurs :	Lennart Ljung	Professeur à l'Université de Linköping, Suède
	Marion Gilson	Maître de Conférences à l'Université Henri Poincaré, Nancy 1
	Hugues Garnier	Professeur à l'Université Henri Poincaré, Nancy 1

Centre de Recherche en Automatique de Nancy
CRAN – UMR 7039

Remerciements

Je tiens tout d'abord à remercier les membres du jury, qui m'ont fait l'honneur d'accepter de participer à l'examen de ce travail.
Mes remerciements à Michel Gevers, professeur à l'Université catholique de Louvain, et à Qinghua Zhang, directeur de recherche INRIA à L'INRIA de Rennes, pour avoir accepté d'être rapporteurs. Je leur suis reconnaissant d'avoir examiné avec attention mon travail, et d'avoir contribué à significativement améliorer le présent manuscrit.
I am very grateful to Lennart Ljung, professor at Linköping University in Sweden, for accepting the task of examiner for my thesis.
Merci enfin à Luc Dugard, directeur de recherche au CNRS au GIPSA de Grenoble, d'avoir accepté de présider ce jury.
It is hard to express with words, how a boundless honor it is for me, who was a complete neophyte in system identification a few years ago, to have my work reviewed and commented by such a prestigious jury composed of the most world-widely recognized researchers.

Avant de commencer cette thèse, j'étais ingénieur en traitement d'image médicale en Autriche. La décision de retourner dans un milieu universitaire, et donc de couper mes revenus en deux, fût en partie motivée par la confiance que m'ont inspirée mes directeurs de thèse Marion et Hugues lors de nos premiers entretiens. De leur coté, et avant même que commence la thèse, ils ont donné beaucoup de leur temps afin de concrétiser une entreprise de trois ans avec un associé pour qui le mot identification ne faisait même pas partie d'un quelconque répertoire linguistique. Depuis, cette confiance mutuelle ne s'est jamais démentie et je suis vraiment reconnaissant que cette confiance sur le plan humain se soit exprimée également professionnellement. Marion et Hugues m'ont toujours laissé la liberté dans la direction donnée à mes recherches et m'ont toujours supporté dans ces choix. De plus, ils n'ont pas hésité à me confier l'image de l'équipe lors de nombreuses présentations orales. Je n'ai jusqu'à ce jour pas regretté le choix de mes encadrants ni celui d'effectuer une thèse et les travaux présentés dans ce manuscrit sont l'accomplissement d'un travail d'équipe remarquable. Merci encore !

I would like to thank Roland Tóth, whose input about LPV models was indubitably essential for this thesis. We met together at the ERNSI in 2008 and this successful collaboration (which slowly turned into friendship) was indeed one of the best outcome one can expect from a conference. köszönöm Roland.

Pour leur accueil enjoué et leur bonne humeur, je tiens également à remercier tous les membres du

département Réseaux et Télécommunications de l'I.U.T. Nancy-Brabois, au sein duquel j'ai effectué mes enseignements.

Alors qu'on peut s'attendre à rencontrer au sein d'un laboratoire de recherche des gens éduqués et ouverts, nul ne pouvait prédire que je trouverais au sein de mon environnement de travail des personnes qui ~~m'appréeient malgré mes paires de Crocs~~ sont devenues de si proches amis. Merci de m'avoir supporté durant toutes les phases de cette thèse jusqu'en dehors du labo et d'avoir partagé autant d'expériences avec moi.

Pour finir, mes remerciements les plus chaleureux pour Karoline et Tristan qui ont mangé des pommes de terre pendant trois ans pour m'autoriser à m'épanouir dans ma vie professionnelle, qui n'ont jamais douté de moi et qui m'ont toujours épaulé. Merci à mes parents (pour avoir partiellement financé les pommes de terre) et à Audrey pour s'être si bien occupé de ma petite famille pendant tout ce temps. Merci à Renée pour sa présence inoubliable.

Table des matières

Table des figures **v**

Liste des tableaux **vii**

Liste des abréviations **ix**

Liste des symboles **xi**

1 Introduction **1**
1.1 Identification des sytèmes . . . 1
1.2 Cadre d'application considéré . . . 2
1.3 Modèles non-linéaires . . . 4
1.4 Minimisation de l'erreur de prédiction pour les modèles linéaires . . . 6
1.5 Méthode d'estimation : la variable instrumentale optimale . . . 7
1.5.1 Problème d'identification . . . 7
1.5.2 Variable instrumentale basique . . . 10
1.5.2.1 Analyse de convergence . . . 10
1.5.2.2 Analyse de précision . . . 10
1.5.3 Variable instrumentale étendue . . . 11
1.5.3.1 Propriétés de convergence . . . 11
1.5.3.2 Analyse de précision . . . 11
1.5.4 Variable instrumentale optimale . . . 12
1.5.5 Variable instrumentale optimale itérative . . . 12
1.6 Estimation directe de modèles à temps continu . . . 13
1.6.1 Modèles à temps continu . . . 14
1.6.2 Gérer les dérivées non-mesurées . . . 15
1.6.3 Modèles hybrides . . . 16
1.7 Organisation de la thèse et principales contributions . . . 17
1.8 Travaux publiés . . . 18

2 Variable instrumentale itérative pour les modèles Hammerstein BJ **21**
2.1 Introduction . . . 21
2.2 Identification de modèles Hammerstein à temps discret . . . 22
2.2.1 Systèmes considérés . . . 22

2.2.2 Modèle considéré . . . 23
2.2.2.1 Modèle de la partie linéaire . . . 23
2.2.2.2 Modèle de la non-linéarité . . . 23
2.2.2.3 Modèle de bruit . . . 24
2.2.2.4 Modèle complet . . . 24
2.2.2.5 Reformulation du modèle . . . 25
2.2.3 Formulation du problème d'identification . . . 26
2.2.4 VI optimale pour les modèles Hammerstein . . . 27
2.2.5 L'algorithme Hammerstein RIV (`HRIV`) pour les modèles BJ . . . 28
2.2.6 L'algorithme `HSRIV` pour les modèles OE . . . 30
2.2.7 Évaluation des méthodes `HRIV` et `HSRIV` . . . 30
2.3 Identification de modèles Hammerstein en boucle fermée . . . 31
2.3.1 Systèmes considérés . . . 32
2.3.2 Modèle considéré . . . 33
2.3.3 Position du problème d'identification . . . 33
2.3.4 VI optimale pour les modèles Hammerstein en BF . . . 34
2.3.5 L'algorithme `HRIV` en BF (`HCLRIV`) pour les modèles BJ . . . 35
2.3.6 L'algorithme `HCLSRIV` pour les modèles OE . . . 37
2.3.7 Évaluation des méthodes `HCLRIV` et `HCLSRIV` . . . 37
2.3.7.1 Performance des algorithmes proposés . . . 37
2.3.7.2 Robustesse face à l'erreur de modélisation . . . 39
2.4 Identification de modèles Hammerstein à temps continu . . . 42
2.4.1 Systèmes considérés . . . 42
2.4.2 Modèle considéré . . . 44
2.4.2.1 Modèle du processus . . . 44
2.4.2.2 modèle de bruit . . . 45
2.4.2.3 Modèle complet . . . 45
2.4.3 Erreur de prédiction . . . 46
2.4.3.1 System reformulation and prediction error . . . 46
2.4.3.2 Modèle pour l'erreur de prédiction . . . 47
2.4.3.3 Minimisation de l'erreur de prédiction . . . 48
2.4.4 Position du problème d'identification . . . 48
2.4.5 VI itérative pour les modèles Hammerstein à temps continu . . . 48
2.4.6 L'algorithme Hammerstein `RIVC` (`HRIVC`) pour les modèles BJ . . . 50
2.4.7 L'algorithme `HSRIVC` pour les modèles OE . . . 52
2.4.8 Évaluation des méthodes `HRIVC` et `HSRIVC` . . . 52
2.4.8.1 Performances des méthodes présentées . . . 52
2.4.8.2 Les avantages de l'identification directe . . . 53
2.5 Conclusion . . . 55

3 Variable instrumentale itérative pour les Modèles LPV BJ 57
3.1 Introduction . . . 57
3.2 Identification de modèles LPV à temps discret . . . 60
3.2.1 Systèmes considérés . . . 60
3.2.2 Modèle considéré . . . 61
3.2.2.1 Modèle du processus . . . 61
3.2.2.2 Modèle de bruit . . . 62
3.2.2.3 Modèle complet . . . 62
3.2.3 Prédiction et erreur de prédiction . . . 63
3.2.4 Persistence d'excitation . . . 64
3.2.5 Position du problème d'identification . . . 65
3.2.6 De l'utilisation de la régression linéaire . . . 65
3.2.6.1 Les conclusions tirées de la littérature . . . 66
3.2.6.2 Méthodes existantes et estimées itératives . . . 66
3.2.6.3 Le cas LPV . . . 67
3.2.7 Reformulation du modèle . . . 67
3.2.8 Variable instrumentale optimale pour les modèles LPV . . . 68
3.2.9 L'algorithme `LPV-RIV` pour les modèles BJ . . . 70
3.2.10 L'algorithme `LPV-SRIV` pour les modèles OE . . . 71
3.2.11 Évaluation des méthodes `LPV-RIV` et `LPV-SRIV` . . . 71
3.2.11.1 Comparaison avec les méthodes existantes . . . 71
3.2.11.2 Robustesse à l'erreur de modélisation . . . 73
3.2.11.3 Modèle de bruit LPV . . . 77
3.3 Identification de modèles LPV en boucle fermée . . . 78
3.3.1 Systèmes considérés . . . 78
3.3.2 Modèle considéré . . . 79
3.3.3 Position du problème d'identification . . . 80
3.3.4 VI optimale pour les modèles LPV en BF . . . 81
3.3.5 L'algorithme LPV-RIV en BF (`LPV-CLRIV`) pour les modèles BJ . . . 82
3.3.6 L'algorithme `LPV-CLSRIV` pour les modèles OE . . . 83
3.3.7 Évaluation des méthodes `LPV-CLRIV` et `LPV-CLSRIV` . . . 83
3.4 Identification de modèles LPV à temps continu . . . 85
3.4.1 Systèmes considérés . . . 85
3.4.2 Modèles considérés . . . 86
3.4.2.1 Modèle du processus . . . 86
3.4.2.2 Modèle de bruit . . . 87
3.4.2.3 Modèle complet . . . 87
3.4.3 Prédiction et erreur de prédiction . . . 88
3.4.3.1 Reformulation du système . . . 88
3.4.3.2 Modèle de l'erreur de prédiction . . . 90
3.4.3.3 Minimisation de l'erreur de prédiction . . . 90
3.4.4 Position du problème d'identification . . . 90

3.4.5 VI optimale pour les modèles LPV à temps continu . . . 91
3.4.6 L'algorithme `LPV-RIVC` pour les modèles BJ . . . 92
3.4.7 L'algorithme `LPV-SRIVC` pour les modèles OE . . . 94
3.4.8 Évaluation des méthodes `LPV-RIVC` et `LPV-SRIVC` . . . 94
3.5 Conclusion . . . 101

4 Identification de la relation pluie/débit d'un bassin versant viticole 103
4.1 Modélisation et hydrologie . . . 103
4.2 Le bassin versant concerné . . . 105
4.3 Les modèles linéaires . . . 108
4.4 Les modèles Hammerstein . . . 109
4.5 Les modèles LPV . . . 110
4.5.1 Choix des variables de séquencement . . . 111
4.5.2 Détermination du paramètre Δ . . . 112
4.5.3 Détermination de la dépendance aux variables de séquencement . . . 112
4.5.4 Modèle LPV obtenu . . . 113
4.6 Conclusion . . . 116

5 Conclusion et perspectives 119
5.1 Conclusion . . . 119
5.2 Perspectives . . . 121

Bibliographie 123

Table des figures

1.1 Procédure d'identification 2
1.2 Représentation d'un bassin versant rural 3
1.3 Modèles non-linéaires en bloc 5
1.4 Modèle Hammerstein pour la relation pluie/débit 5

2.1 Système Hammerstein 22
2.2 Modèle Hammerstein augmenté 26
2.3 Système Hammerstein en boucle fermée 32
2.4 Système Hammerstein non bruité en boucle fermée 35
2.5 Modèle Auxiliaire en boucle fermée 35
2.6 Exemple 2.3.7.2 : Résultats avec erreur de modélisation 41
2.7 Système Hammerstein à temps continu 42
2.8 Système hybrid Hammerstein 43
2.9 Modèle Hammerstein à temps continu augmenté 46
2.10 Exemple 2.4.8.2 : Estimation directe à temps continu 56

3.1 Exemple d'un ressort avec masse variant dans le temps 58
3.2 Interprétation LTI du modèle LPV 68
3.3 Exemple 3.2.11.1 : Signaux d'entrée sortie à RSB = 0dB 73
3.4 Exemple 3.2.11.2 : Résultats de l'erreur de modélisation (simulation) 76
3.5 Système LPV en BF 78
3.6 Système non bruité 81
3.7 Modèle auxiliaire estimé 83
3.8 Exemple 3.4.8 : Diagramme de Bode du vrai système 96
3.9 Exemple 3.4.8 : Sorties des modèles estimés 98
3.10 Example 3.4.8 : Diagramme de Bode des modèles estimés 99
3.11 Exemple 3.4.8 : Pôles des modèles estimés 100

4.1 Bassin du Hohrain 105
4.2 Photos du bassin versant 106
4.3 Données de l'année 2008 107
4.4 Effets de l'accumulation de sédiments sur les données 108
4.5 Exemple 4.3 : Résultat de la modélisation linéaire 109

4.6 Exemple 4.4 : Résultat de la modélisation Hammerstein 111
4.7 Processus d'identification de la relation pluie/débit 113
4.8 Exemple 4.5 :Résultat de la modélisation LPV-ARX 114
4.9 Exemple 4.5 : Résultat de la modélisation LPV-OE 115
4.10 Exemple 4.5 : Comparaison des modèles linéaires et LPV 115

Liste des tableaux

2.1 Exemple 2.2.7 : Estimatation de modèles Hammerstein à temps discret 31
2.2 Exemple 2.3.7 : Estimation en boucle fermée 39
2.3 Exemple 2.3.7.2 : Résultats avec erreur de modélisation 41
2.4 Exemple 2.4.8 : Robustesse au bruit 54
2.5 Exemple 2.4.8 : Estimation du modèle de bruit 54

3.1 Exemple 3.2.11.1 : LPV en BO, estimation du dénominateur 74
3.2 Exemple 3.2.11.1 : LPV en BO, estimation du numérateur 74
3.3 Exemple 3.2.11.1 : Comparaison des méthodes d'estimation 75
3.4 Exemple 3.2.11.2 : Résultats de l'erreur de modélisation 76
3.5 Exemple 3.2.11.2 : Résultats de l'erreur de modélisation 76
3.6 Exemple 3.2.11.3 : Estimation avec modèle de bruit LPV 77
3.7 Exemple 3.3.7 : Modèles LPV en boucle fermée 85
3.8 Exemple 3.4.8 : Estimation de modèles LPV à temps continu 97

4.1 Exemple 4 : Coefficient NASH pour les différents modèles 116

Liste des Abréviations

ARMA	Auto-Regressive with Moving Average
ARX	Auto-Regressive with eXogenous input
BIBO	Bounded Input Bounded Output
BJ	Box–Jenkins
CLRIV	Closed-Loop Refined Instrumental Variable
CLSRIV	Closed-Loop Simplified Refined Instrumental Variable
CT	Continuous-Time
DT	Discrete-Time
EIV	Errors-In-Variables
IO	Input/Output
IRR	Impulse Response Representation
IV	Instrumental Variable
LPV	Linear Parameter Varying
LS	Least Squares
LTI	Linear Time Invariant
MCS	Monte Carlo Simulation
MIMO	Multi Input Multi Output
MISO	Multi Input Single Output
OE	Output Error
OSIV	One-Step Instrumental Variable method
PEM	Prediction Error Minimization
RIV	Refined Instrumental Variable
RIVC	Refined Instrumental Variable for Continuous-time systems
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SRIV	Simplified Refined Instrumental Variable
SRIVC	Simplified Refined Instrumental Variable for Continuous-time systems

Liste des Symboles

u	Input
$\mathring{u}$	Input of the noise–free closed–loop system
y	Measured output
χ	Noise-free output (process output in closed loop)
$\mathring{\chi}$	Output of the noise–free closed–loop system
p	Scheduling variable
$y^{(n)}(t_k)$	n^{th} time-derivative of signal y sampled at time t_k
N	Number of samples
t_k	Sampling time
$\mathtt{T}_\mathrm{s}$	Sampling period
$\mathcal{D}$	Data set
z	Z-transform variable
s	Laplace variable
ρ	process model parameter vector
η	Noise model parameter vector
θ	Full model parameter vector
$\hat{\theta}^{(\tau)}$	Estimate of the parameter vector at iteration τ
θ_o	Parameter vector corresponding to the true system
ρ_o	Parameter vector corresponding to the true process
η_o	Parameter vector corresponding to the true noise system
ρ^*	Limiting estimate of ρ as $N \to \infty$
e_o	True white noise added to the system
v_o	True colored noise added to the output
ζ	Instrument
φ	regressor
$\mathcal{G}$	Process model set
$\mathcal{H}$	Noise model set
$\mathcal{M}$	Full model set
d	Differential operator

q	Forward time shift operator
q^{-1}	Delay time shift operator
$\mathbb{E}$	Generalized expectation
$\|.\|$	Euclidean norm
$(.)^\top$	Transpose of a matrix
$\mathcal{G}_\mathrm{o}$	True process
$\mathcal{H}_\mathrm{o}$	True noise system
$\mathcal{G}_\rho$	Process model associated to the parameter vector ρ
$\mathcal{H}_\eta$	Noise model associated to the noise parameter vector η
G_o	True process transfer function
H_o	True noise system transfer function
$\mathcal{S}_\mathrm{o}$	Data generating system

Chapitre 1

Introduction

1.1 Identification des sytèmes

L'identification de systèmes est la spécialité visant à modéliser les systèmes dynamiques à partir de données expérimentales et un historique de cette discipline peut être trouvé par exemple dans [Gev06]. Contrairement à la modélisation mathématique, l'identification de systèmes est une approche expérimentale, lors de laquelle un modèle issu de connaissances *a priori* est optimisé à partir des données mesurées afin d'approcher le plus fidèlement possible le comportement du système. Le processus d'identification est représenté dans la figure 1.1 et les trois principales étapes sont [Lju99] :

- l'acquisition de données échantillonnées à partir du système réel ;
- la définition d'une structure de modèle qui puisse approximer le fonctionnement du système ;
- la définition d'un critère de sélection pour choisir le meilleur modèle au sein de la structure définie ainsi qu'une méthode d'optimisation permettant la minimisation de ce critère.

La figure 1.1 montre que la conception d'une méthode d'estimation s'inscrit dans un processus expérimental plus général. En conséquence, la structure de modèle, le critère à minimiser ainsi que la méthode d'estimation associée doivent être choisis en fonction de l'application visée et des connaissances *a priori* disponibles. La plupart des méthodes d'estimation développées lors des dernières décennies sont principalement destinées à des applications industrielles pour la commande. Néanmoins, les systèmes dynamiques ne se limitent pas aux systèmes automatisés et sont souvent des systèmes naturels qu'ils soient chimiques, biologiques ou environnementaux par exemple. Ces systèmes naturels présentent intrinsèquement de nombreuses difficultés :

- comportements non-linéaires ;
- les modèles conceptuels sont souvent trop complexes pour être identifiables ;
- ces systèmes ne peuvent pas être controlés ;
- l'acquisition de certaines variables est parfois impossible.

En règle générale, les modèles issus des lois physiques décrivent avec précision les comportements locaux mais deviennent trop complexes lorsqu'étendus à la globalité de ces systèmes. De plus, le nombre trop élevé de paramètres pose de nombreux problèmes d'identifiabilité. En contrepartie, la simplification du comportement général de ces systèmes est elle aussi un problème non résolu. Ainsi, l'identification des systèmes naturels considérés dans leur ensemble est un problème stimulant tant au niveau du choix de la structure de modèle que du développement de la méthode d'estimation.

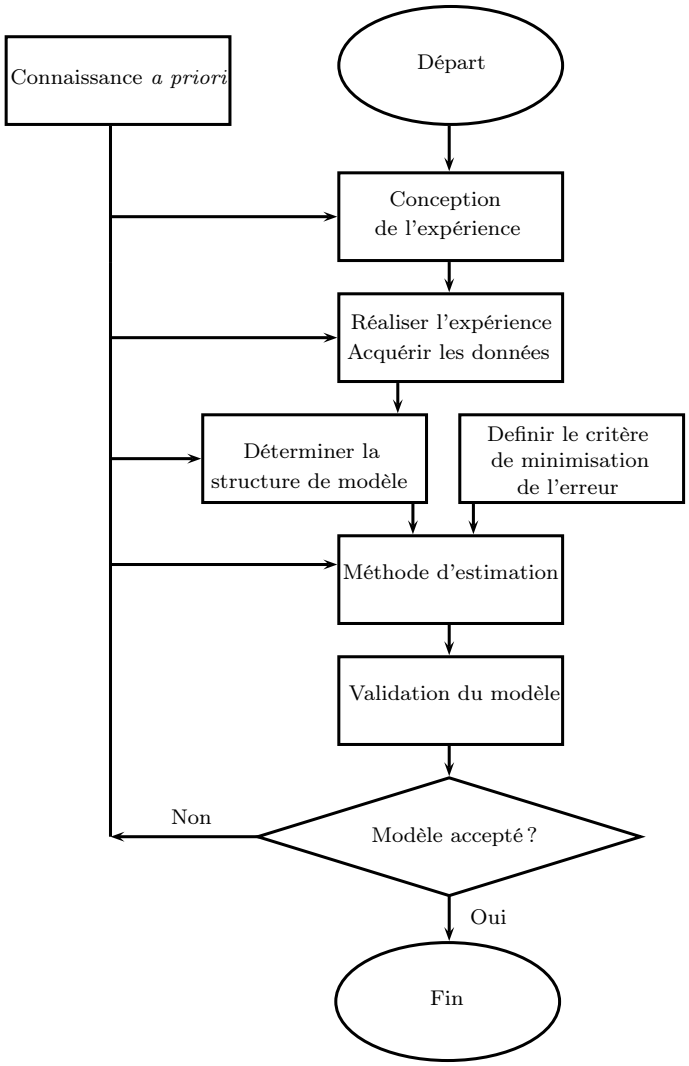


FIG. 1.1 – Procédure d'identification

1.2 Cadre d'application considéré

Parmi les systèmes environnementaux, il y a dernièrement un regain d'intérêt pour ceux liés au traitement de l'eau de par l'augmentation de la pollution en zone rurale ou la taille des infrastructures en milieu urbain.

Un bassin versant est défini comme la portion de territoire délimitée par des lignes de crête, dont les eaux alimentent un exutoire commun : cours d'eau, lac, mer, *etc.* Les bassins versant varient très largement en taille : de quelques milliers de kilomètres carrés pour certaines rivières à seulement quelques hectares pour de petits exutoires. Leurs caractéristiques (physiques, chimiques, biologiques) sont déterminées par la nature du bassin et des activités, humaines ou naturelles, développées sur le bassin. La figure 1.2 représente le fonctionnement d'un bassin versant.

L'identification de la relation pluie/débit est un problème stimulant, de par la complexité à trouver une structure de modèle définissant le comportement du bassin dans son ensemble ([Bev00a]). Les modèles conceptuels sont souvent trop limités pour la simulation à long terme et ne peuvent pas modéliser correctement les changement de dynamiques du système face à des événements pluvieux

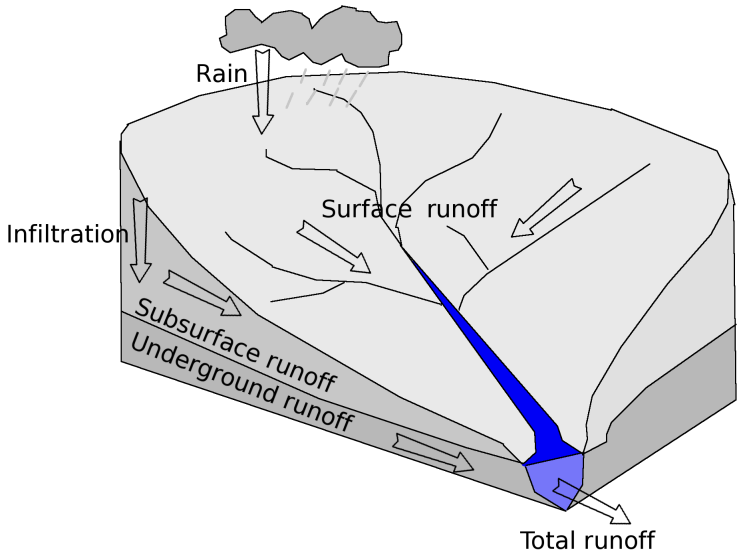


FIG. 1.2 – Représentation d'un bassin versant rural

différents en longueur ou en intensité ([MP96, PLM99]). Les difficultés rencontrées sont cependant différentes qu'il s'agisse d'un bassin rural ou urbain. Au sein d'un bassin urbain, les non-linéarités résultent des infrastructures humaines dédiées au traitement de l'eau ([BGM01, PL09]) , alors que la plupart des eaux tombées rejoignent l'exutoire. Dans ce cas, les modèles linéaires, bien que moins performants que leurs homologues non-linéaires, proposent une capacité de simulation acceptable ([PL09, KLG$^+$09]).

A contrario, dans les bassins ruraux,une grande variabilité spatio-temporelle des propriétés du sol est observée, tant au niveau de la végétation que du type de sol ou de l'évapotranspiration ([YG06]). Par ailleurs, seulement une partie de la pluie totale (pluie tombée) ruisselle et contribue au débit à l'exutoire (pluie efficace), l'autre partie étant absorbée par le sol ou évaporée. Dans ce cas, les modèles linéaires ne sont pas adaptés et ne peuvent pas délivrer de modèle correct pour la relation pluie/débit.

En plus de constituer un défi conséquent en terme d'identification, la modélisation de bassin versant à partir de données expérimentales présente une véritable innovation dans la communauté hydrologue. Dans ce contexte, l'identification de modèles parcimonieux représente un réel besoin et vise à offrir une structure simple ayant à la fois de bonnes capacités de prédiction et d'interprétation physique. Les propriétés nécessaires pour la structure de modèle choisie sont les suivantes :

- le modèle recherché doit pouvoir gérer le comportement non-linéaire du système ;
- le modèle doit pouvoir offrir une possibilité de validation physique ("Data-based mechanistic modelling " [YB94]) : ainsi, le modèle doit être paramétrique et il doit être le plus parcimonieux possible afin de faciliter cette interprétation ;
- les systèmes naturels sont le plus souvent exprimés sous forme d'équations différentielles. Par conséquent, l'interprétation du modèle recherché sera facilitée si ce modèle est exprimé sous forme entrée/sortie (IO). De plus, en terme d'identification, les modèles entrée/sortie sont très largement utilisés pour l'identification car la signification stochastique de l'estimation est mieux comprise, par exemple grâce à l'erreur de prédiction, que pour d'autres structures de modèles.

Un autre problème majeur lié à l'identification de systèmes environnementaux est l'impossibilité de les contrôler. Cela résulte en un problème d'identification avec *erreur en-les-variables*. Cependant, les méthodes dédiées à ce contexte nécessitent de fortes hypothèses sur le bruit considéré ([SÖ7], [TZGG09]). Nonobstant, l'hypothèse gaussienne habituellement utilisée pour modéliser le bruit sur les données mesurées n'est pas forcément vérifiée dans l'application considérée. Par exemple, la variance du bruit augmente avec le débit alors qu'elle devient nulle lorsque qu'aucun débit n'est présent. De plus, lors de forts événements pluvieux, des alluvions s'accumulent dans le capteur utilisé pour mesurer le débit à l'exutoire, résultant dans la mesure de débits non-existants. L'hypothèse d'une expression stochastique du bruit est clairement violée dans ce contexte. Néanmoins, la plupart des méthodes d'estimation reposent sur cette représentation stochastique du bruit. Par conséquent, la méthode d'estimation proposée devra être robuste face à ces conditions de bruit.
Enfin, les systèmes naturels s'expriment sous forme de systèmes continus alors que l'identification ne peut se aire qu'à partir de données échantillonnées. Cette dualité est largement maîtrisée pour les modèles linéaires invariants dans le temps. En revanche, la transformation *temps discret/temps continu* (TD/TC) implique plus de complexité pour les modèles variants dans le temps (cf section 3.1). Par conséquent, l'estimation directe de modèles non-linéaires à temps continu est très attractive pour faire face au défi d'une modélisation interprétable physiquement pour les systèmes dynamiques naturels.

1.3 Modèles non-linéaires

Le besoin croissant de commande performante pour les applications industrielles actuelles ainsi que de la représentation de systèmes naturels poussent l'identification de systèmes à faire face à une modélisation plus précise des systèmes physiques. L'identification des systèmes linéaires a été très largement étudiée dans les précédentes décennies et l'identification paramétrique basée sur la minimisation de l'erreur de prédiction comme définie dans [Lju99] prédomine clairement. Cependant, les systèmes rencontrés en pratique sont souvent de nature non-linéaire ou variant dans le temps [Nel01], [GS01]. Les modèles non-linéaires peuvent être classés en deux classes : les modèles non-paramétriques d'une part et les modèles paramétriques d'autre part. L'identification des systèmes non-linéaires par des modèles non-paramétriques est très alléchante car ces derniers ne requièrent aucune connaissance (ou très peu) *a priori* du système étudié. Nonobstant, ces modèles sont souvent inutilisables en pratique et ne donnent aucune information quant au fonctionnement du système étudié. Le choix d'une structure appropriée est l'une des plus importantes étapes du processus d'identification. Elle doit contenir au moins un modèle approchant à la précision souhaitée le fonctionnement du système réel. Il est donc important de choisir une structure paramétrique pouvant modéliser le plus grand nombre de comportements tout en gardant une structure simple facilitant l'optimisation et l'interprétation. Parmi les modèles non-linéaires paramétriques, les modèles par blocs sont très largement étudiés et les plus utilisés sont les types : Wiener, Hammerstein, Hammerstein-Wiener, Wiener-Hammerstein représentés dans la figure 1.3.
Parmi ces modèles, le modèle le plus communément utilisé pour la modélisation de la relation pluie/débit est la structure Hammerstein représentée dans la figure 1.4 [YB94]. L'hypothèse sur le comportement du système peut être résumée ainsi : le bassin est considéré comme une éponge qui

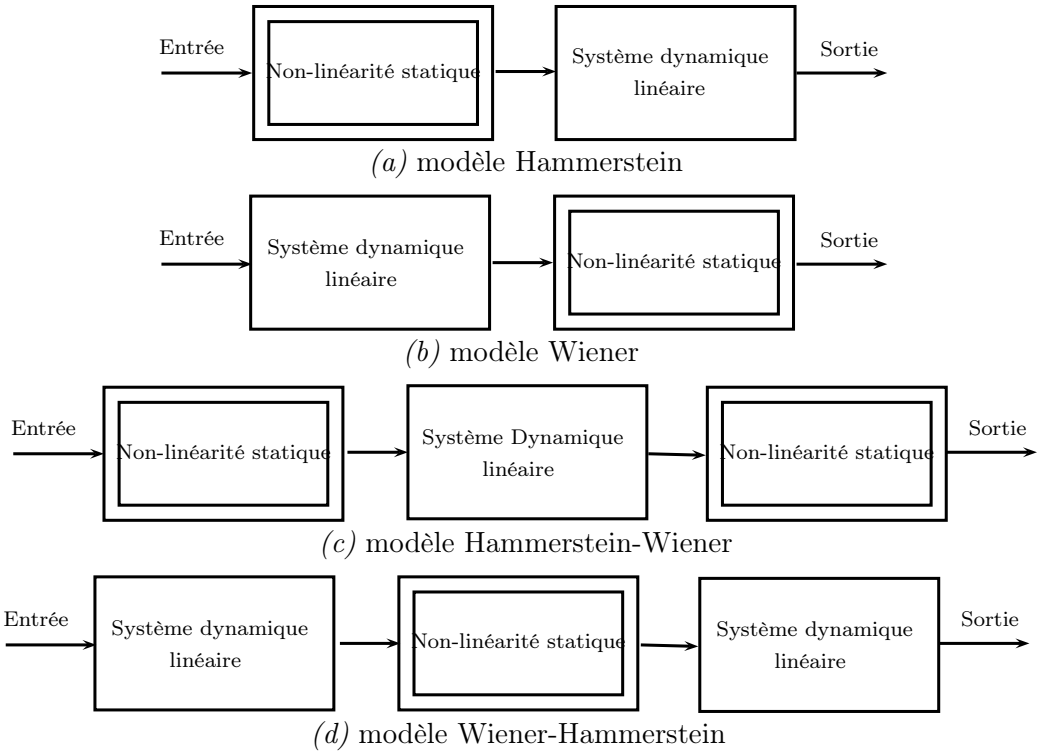


FIG. 1.3 – Modèles non-linéaires en bloc

absorbe l'eau jusqu'à un certain seuil au dessus duquel le ruissellement commence et où le système réagit comme une cuve percée. Ceci n'est bien sûr qu'une approximation et dans ce cas précis, la non-linéarité est la fonction liant la pluie totale et la pluie efficace contribuant au ruissellement.

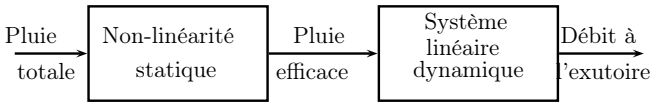


FIG. 1.4 – Modèle Hammerstein pour la relation pluie/débit

Cependant, il est aussi possible de traduire ce modèle de type "éponge" comme un changement qui s'exerce autant sur le gain du système que sur sa dynamique. En conséquence, parallèlement aux modèles Hammerstein, un autre type de modèle non-linéaire paramétrique attire l'attention de la communauté scientifique depuis peu pour faire face aux problèmes de modélisation, tout en créant une étape intermédiaire entre les modèles linéaires invariants dans le temps et les modèles non-linéaires : les modèles linéaires à paramètres variants (LPV) qui furent introduits au travers du séquencement de gain [SA92]. Pour les modèles LPV, les relations entre les signaux sont linéaires, mais les paramètres du modèles linéaires peuvent varier dans le temps en fonction d'une grandeur mesurable appelée *variable de séquencement* $p : \mathbb{Z} \mapsto \mathbb{P}$. L'ensemble compact $\mathbb{P} \subseteq \mathbb{R}^{n_\mathbb{P}}$ définit l'espace de séquencement. La structure des modèles LPV peut représenter un large panel de comportements de systèmes physiques et est déjà très largement utilisée dans la communauté de la commande. En outre, ces modèles sont de plus en plus utilisés dans la modélisation du traitement de l'eau à partir de données mesurées [PL09] principalement car :

– ils peuvent représenter des systèmes non-linéaires ;

– ils sont intuitifs, faciles à formuler et à interpréter ;
– ils peuvent incorporer les connaissances *a priori* grâce aux choix de variables de séquencement.

Malgré les avances en commande des modèles LPV, l'identification de ces systèmes est très largement en retrait. Une liste de références exhaustive sur l'utilisation des modèles LPV est donnée dans [Tót10]. En conclusion, en plus du fait que ces modèles sont de très bons candidats pour la modélisation de systèmes naturels, les défis relatifs à cette classe de modèles restent nombreux.
Après avoir défini la classe de modèles, il reste à définir :
– un critère à minimiser ;
– un algorithme d'optimisation minimisant ce critère.

La plupart des méthodes dédiées à l'identification des modèles précités sont fondées sur l'extension de la minimisation de l'erreur de prédiction pour les systèmes linéaires définie dans [Lju99] qui est brièvement décrite dans la section suivante.

1.4 Minimisation de l'erreur de prédiction pour les modèles linéaires invariants dans le temps

Un critère doit être défini afin de justifier le choix d'un modèle en particulier au sein de la structure de modèle définie. La plupart des méthodes d'identification utilisent la minimisation de l'erreur de prédiction. Cette section décrit brièvement ce critère de choix pour les modèles linéaires comme défini dans [Lju99] qui sera étendu dans les prochains chapitres à différents types de modèles non-linéaires.
Le système générant les données est défini par :

$$\mathcal{S}_{\mathrm{o}} : y(t_k) = G_{\mathrm{o}}(q)u(t_k) + v_{\mathrm{o}}(t_k), \tag{1.1}$$

où le signal d'entrée u est quasi-stationnaire[1], y est le signal de sortie, q est l'opérateur retard ($q^d u(t_k) = u(t_{k+d})$), $G_{\mathrm{o}}(q)$ est la fonction de transfert du système étudié et v_{o} est le bruit. En pratique, une bonne approximation du cas le plus réaliste correspond à un bruit de sortie coloré à densité spectrale finie qui n'a aucune relation avec la dynamique du système étudié.
Par conséquent, tout au long de cette thèse, pour tous les modèles considérés, deux types de structure seront étudiées :
– le modèle Box-Jenkins (BJ) : v_{o} est modélisé par un modèle *auto-régressif à moyenne ajustée* (ARMA) :

$$v_{\mathrm{o}}(t_k) = H_{\mathrm{o}}(q)e_{\mathrm{o}}(t_k) = \frac{C_{\mathrm{o}}(q^{-1})}{D_{\mathrm{o}}(q^{-1})}e_{\mathrm{o}}(t_k), \tag{1.2}$$

où e_{o} est un bruit blanc de moyenne nulle et de variance σ_e^2, $C_{\mathrm{o}}(q^{-1})$ et $D_{\mathrm{o}}(q^{-1})$ sont des polynômes moniques de degré n_{c} et n_{d}. De plus les racines de $z^{n_{\mathrm{d}}}D_{\mathrm{o}}(z^{-1})$ se situent à l'intérieur du disque unitaire et z est la variable de la transformée en Z,
– le modèle Output-Error (OE) : $v_{\mathrm{o}}(t_k)$ est un bruit blanc. Ce modèle est équivalent au modèle BJ lorsque $C_{\mathrm{o}}(q^{-1}) = D_{\mathrm{o}}(q^{-1}) = 1$ dans (1.2) .

[1] Les propriétés statistiques du signal ne varient pas significativement durant la période d'observation

Soit $\mathcal{D}_N = \{y(t_k), u(t_k)\}_{k=1}^N$ une séquence de données générée par (1.1). Il est possible de définir la prédiction à un pas de $y(t_k)$ fondée sur les échantillons précédents[2] $\{y(t_{k-1}), y(t_{k-2}), \ldots\}$ et $\{u(t_{k-1}), u(t_{k-2}), \ldots\}$ comme étant :

$$\hat{y}(t_k) = \Big(1 - H_{\mathrm{o}}^{-1}(q)\Big)y(t_k) + H_{\mathrm{o}}^{-1}(q)G_{\mathrm{o}}(q)u(t_k). \tag{1.3}$$

La seconde étape consiste à définir une classe de modèles $\mathcal{M} = \{(G(q,\rho), H(q,\eta) \mid \theta = col(\rho,\eta) \in \mathbb{R}^{n_\theta}\}$ au sein de laquelle se trouve le système vrai (dans les cas BJ et OE, le processus et le bruit peuvent être modélisés séparément). Le prédicateur à un pas est alors défini comme :

$$\hat{y}_\theta(t_k) = \Big(1 - H^{-1}(q,\eta)\Big)y(t_k) + H^{-1}(q,\eta)G(q,\rho)u(t_k). \tag{1.4}$$

Enfin, dans le cadre de l'erreur de prédiction, le choix du meilleur vecteur de paramètres θ est motivé par la minimisation de l'erreur de prédiction $\varepsilon_\theta(t_k) = y(t_k) - \hat{y}_\theta(t_k)$. Habituellement, un critère d'identification $W(\mathcal{D}_N, \theta)$ est introduit, comme le critère des moindres carrés (bien qu'une autre norme puisse être utilisée) :

$$W(\mathcal{D}_N, \theta) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon_\theta^2(t_k), \tag{1.5}$$

tel que le paramètre estimé soit donné par :

$$\hat{\theta}_N = \arg\min_{\theta \in \mathbb{R}^{n_\theta}} W(\mathcal{D}_N, \theta). \tag{1.6}$$

Une fois le critère de minimisation défini, la dernière difficulté est le développement d'un algorithme minimisant ce critère.

Parmi les différentes méthodes d'optimisation, les plus connues dans le domaine temporel sont fondées sur les moindres carrés, sur le maximum de vraisemblance [Lju99] ou sur la méthode de variable instrumentale optimale [YJ80]. Bien que les moindres carrés proposent une solution analytique lorsque $G(q)$ est paramétré de façon linéaire, les modèles obtenus sont biaisés lorsque le modèle de bruit est erroné. Par opposition, d'autres algorithmes d'optimisation reposent sur des méthodes d'optimisation non-linéaires, pour lesquelles l'obtention du minimum global est très sensible aux conditions d'initialisation. La méthode de variable instrumentale est une méthode itérative intermédiaire puisqu'elle est robuste à l'initialisation, asymptotiquement non biaisée dans le cas d'une fausse hypothèse sur le bruit, et propose une solution analytique d'une itération à l'autre. En contrepartie, le principal désavantage de cette méthode est que sa convergence n'a pas été prouvée théoriquement mais est acceptée empiriquement. La prochaine section décrit l'approche par variable instrumentale optimale introduite dans [YJ80].

1.5 Méthode d'estimation : la variable instrumentale optimale

1.5.1 Problème d'identification

Le problème d'estimation pour les modèles linéaires peut se définir comme suit en se fondant sur les considérations précédemment décrites.

[2] On peut noter que le retard zéro ou l'utilisation d'échantillons futurs ne change pas significativement le résultats

Problem 1 *Soit le système $\mathcal{S}_\mathrm{o}$ générant les données et l'ensemble de données $\mathcal{D}_N$ issues de $\mathcal{S}_\mathrm{o}$. Au sein de la structure de modèles $\mathcal{M}_\theta$ definie, il s'agit d'estimer le vecteur de paramètres θ en utilisant $\mathcal{D}_N$ sous les hypothèses :*

LTI1 *Le système vrai appartient à la classe de modèless définie : $\mathcal{S}_\mathrm{o} \in \mathcal{M}$.*

LTI2 *$u(t_k)$ n'est pas corrélé à $e_\mathrm{o}(t_k)$.*

LTI3 *$\mathcal{D}_N$ est assez informatif[3] par rapport à $\mathcal{M}$.*

LTI4 *$\mathcal{G}_\mathrm{o}$ est stable BIBO [4], i.e. pour tout signal borné u, la sortie de $\mathcal{S}_\mathrm{o}$ est bornée.*

Le problème d'optimisation posé par (1.6) n'est en général pas convexe. Néanmoins, si la modélisation est linéaire en les paramètres, il peut s'écrire sous forme de régression linéaire et se réduit à un problème d'optimisation convexe. Soit le modèle paramétrique :

$$\mathcal{M}_\theta : y_\theta(t_k) = \frac{B(q^{-1},\rho)}{A(q^{-1},\rho)}u(t_k) + H(q,\eta)e(t_k). \tag{1.7}$$

avec $B(q^{-1},\rho)$ et $A(q^{-1},\rho)$ des polynômes en q^{-1} d'odre n_b et n_a respectivement. Si le système étudié se trouve au sein de la structure de modèle définie ($\mathcal{S}_\mathrm{o} \in \mathcal{M}$) , soit si il existe $\theta_\mathrm{o} = col(\rho_\mathrm{o}, \eta_\mathrm{o})$ tel que $y_{\theta_\mathrm{o}}(t_k) = y(t_k)$ alors l'équation du système peut s'écrire sour la forme :

$$y(t_k) = \varphi^\top(t_k)\rho_\mathrm{o} + \tilde{v}(t_k), \tag{1.8}$$

où ρ_o le vecteur de paramètres lié au processus et φ le vecteur de regression donné par :

$$\varphi^\top(t_k) = \left[y(t_{k-1}), y(t_{k-2}), \ldots, y(t_{k-n_a}), u(t_{k-1}), u(t_{k-2}), \ldots, u(t_{k-n_b})\right]. \tag{1.9}$$

L'erreur d'équation $\tilde{v}(t_k)$ est donnée par :

$$\tilde{v}(t_k) = A_\mathrm{o}(q^{-1})H_\mathrm{o}(q)e_\mathrm{o}(t_k). \tag{1.10}$$

En considérant l'approche de minimization de l'erreur de prédiction pour les modèles linéaires, l'erreur de prédiction $\varepsilon_\theta(t_k)$ de (1.8) par rapport à (1.7) est donnée par :

$$\varepsilon_\theta(t_k) = \frac{1}{H(q,\eta_\mathrm{o})A_\mathrm{o}(q^{-1})}\Big(A_\mathrm{o}(q^{-1})y(t_k) - B_\mathrm{o}(q^{-1})u(t_k)\Big). \tag{1.11}$$

Les opérations polynomiales commutent et par conséquent, $\varepsilon_\theta(t_k)$ est équivalent à l'erreur $\varepsilon_*(t_k)$ définie par :

$$\varepsilon_*(t_k) = A_\mathrm{o}(q^{-1})y_\mathrm{f}(t_k) - B_\mathrm{o}(q^{-1})u_\mathrm{f}(t_k), \tag{1.12}$$

où $y_\mathrm{f}(t_k) = Q(q,\theta_\mathrm{o})y(t_k)$ et $u_\mathrm{f}(t_k) = Q(q,\theta_\mathrm{o})u(t_k)$ représentent les sorties de l'opération de préfiltrage avec

$$Q(q,\theta_\mathrm{o}) = \frac{1}{H_\mathrm{o}(q)A_\mathrm{o}(q^{-1})}. \tag{1.13}$$

[3] Deux modèles différents au sein de la classe de modèles $\mathcal{M}$ produisent une erreur de prédiction différente lorsque $\mathcal{D}_N$ est utilisé

[4] Bounded Input Bounded Output

Ainsi, (1.8) équivaut à :

$$y_{\mathrm{f}}(t_k) = \varphi_{\mathrm{f}}^{\top}(t_k)\rho_{\mathrm{o}} + \tilde{v}_{\mathrm{f}}(t_k), \tag{1.14}$$

avec

$$\tilde{v}_{\mathrm{f}}(t_k) = A_{\mathrm{o}}(q^{-1})v_{\mathrm{f}}(t_k) = e_{\mathrm{o}}(t_k). \tag{1.15}$$

Il est possible de voir dans (1.15) que le préfiltrage de l'entrée et de la sortie par $Q(q, \theta_{\mathrm{o}})$ transforme (1.7) en un modèle ARX de la forme :

$$y_{\theta_{\mathrm{f}}}(t_k) = \frac{B(q^{-1}, \eta)}{A(q^{-1}, \rho)}u_{\mathrm{f}}(t_k) + \frac{1}{A(q^{-1}, \rho)}e(t_k). \tag{1.16}$$

Ergo, en émettant momentanément l'hypothèse que $Q(q, \theta_{\mathrm{o}})$ est connu *a priori*, l'estimée statistiquement optimale (non biaisée à variance minimale) peut être obtenue en utilisant les moindres carrés :

$$\hat{\rho}_{\mathtt{LS}}(N) = \arg\min_{\rho \in \mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N} \varphi_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k) \right] \rho - \left[\frac{1}{N}\sum_{k=1}^{N} \varphi_{\mathrm{f}}(t_k)y_{\mathrm{f}}(t_k) \right] \right\|, \tag{1.17}$$

qui peut être résolue analytiquement :

$$\hat{\rho}_{\mathtt{LS}}(N) = \left[\sum_{k=1}^{N} \varphi_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k) \right]^{-1} \left[\sum_{k=1}^{N} \varphi_{\mathrm{f}}(t_k)y_{\mathrm{f}}(t_k) \right]. \tag{1.18}$$

En pratique, s'il est acceptable de considérer que le système recherché soit dans la classe de modèles définie, il est hautement improbable que le bruit de mesure corresponde exactement au modèle de bruit considéré. Par conséquent, il se peut qu'il n'existe pas de préfiltre optimal exact $Q(q, \theta_{\mathrm{o}})$ et ainsi, aucun filtre qui puisse mener à l'estimation d'un modèle ARX comme défini dans (1.16). Dans ce cas, l'estimée des moindres carrés est biaisée (bien que la variance reste faible) de par la corrélation entre le vecteur de régression et le bruit telle que $\mathbb{E}\varphi_{\mathrm{f}}(t_k)\tilde{v}_{\mathrm{f}}(t_k) \neq 0$. Il est donc important d'utiliser une méthode d'estimation capable de générer des estimées robustes dans le cas où le modèle de bruit n'est pas correctement défini (LTI1 est alors relaxée). Néanmoins, il existe un dilemme biais/variance bien connu et qui dépend de la méthode d'estimation utilisée.

Parmi les méthodes d'identification disponibles pour les modèles entrée/sortie, les approches par variable instrumentale connaissent un intérêt croissant depuis quelques années. La principale raison pour cet attrait est que les approches par VI offrent des performances similaires aux moindres carrés étendus ou aux autres méthodes de minimisation de l'erreur de prédiction [RG04, Lju09] tout en offrant des résultats satisfaisants lorsque la structure du bruit est inconnue ; ce qui est le cas de la plupart des applications.

Le principe de base de toute méthode par variable instrumentale est d'introduire un instrument $\zeta(t_k)$ choisi pour répondre aux conditions suivantes : l'instrument $\zeta(t_k)$ doit être le plus possible corrélé avec le vecteur de régression $\varphi(t_k)$ tout en étant décorrélé du bruit.

1.5.2 Variable instrumentale basique

Le vecteur de paramètres ρ peut être estimé en utilisant une variable instrumentale basique. Soient les échantillons passés des entrées-sorties $\{u(t_{k-1}), u(t_{k-2}), \ldots\}$ et $\{y(t_{k-1}), y(t_{k-2}), \ldots\}$, l'estimée par variable instrumentale basique de ρ est calculée comme étant [SS83] :

$$\hat{\rho}_{iv} = sol\left\{\frac{1}{N}\sum_{k=1}^{N}\zeta(t_k)[y(t_k) - \varphi^\top(t_k)\rho] = 0\right\} \tag{1.19}$$

où N dénote le nombre d'échantillons et $\zeta(t_k)$ est appelé *l'instrument*.
Il y a un nombre considérable de possibilités quant au choix de l'instrument. Par exemple, l'instrument peut être composé des signaux retardés dans le cas d'un bruit balnc ajouté en sortie. Le choix $\zeta(t_k) = \varphi(t_k)$ correspond à la solution des moindres carrés. Néanmoins le choix de l'instrument joue un rôle important sur la variance des estimées comme montré dans le paragraphe suivant.

1.5.2.1 Analyse de convergence

En incluant (1.8) dans (1.19), l'équation suivante est obtenue

$$\hat{\rho}_{iv} = \rho_o + \left[\sum_{k=1}^{N}\zeta(t_k)\varphi^\top(t_k)\right]^{-1}\left[\sum_{k=1}^{N}\zeta(t_k)\tilde{v}(t_k)\right] \tag{1.20}$$

où $\varphi^\top(t_k)$ et $\tilde{v}(t_k)$ sont donnés par (1.9) et (1.10) respectivement. On peut déduire de (1.20) que $\hat{\rho}_{iv}$ est une estimée asymptotiquement convergente (ou non biaisée dans le cas de données finies) de ρ si [5] :

$$\begin{cases}\bar{\mathbb{E}}[\zeta(t_k)\varphi^\top(t_k)] \text{ est non singulière}\\ \bar{\mathbb{E}}[\zeta(t_k)\tilde{v}(t_k)] = 0\end{cases} \tag{1.21}$$

Plusieurs méthodes de variable instrumentale peuvent être obtenues suivant le choix de l'instrument $\zeta(t_k)$ dans (1.19), tout en respectant les conditions (1.21).

1.5.2.2 Analyse de précision

La distribution asymptotique du vecteur de paramètres estimé $\hat{\rho}_{iv}$ dans (1.19) a très largement été étudiée pour les modèles en boucle ouverte (e.g. [SS83, Lju99]). Plus récemment, cette étude a été réalisée pour les systèmes opérant en boucle fermée [GVdH05]. Selon les hypothèses formulées (LTI1 à LTI4), $\hat{\rho}_{iv}$ suit asymptotiquement une distribution Gaussienne

$$\sqrt{N}(\hat{\rho}_{iv} - \rho^*) \overset{dist}{\rightarrow} \mathcal{N}(0, \mathbf{P}_{iv}) \tag{1.22}$$

où ρ^* représente la limite $\hat{\rho}_{iv}$ quand $N \rightarrow \infty$ et où la matrice de covariance est définie comme :

$$\mathbf{P}_{iv} = \sigma_{e_o}^2\left[\bar{\mathbb{E}}\zeta(t_k)\varphi^\top(t_k)\right]^{-1}\left[\bar{\mathbb{E}}\tilde{\zeta}(t_k)\tilde{\zeta}^\top(t_k)\right]\left[\left(\bar{\mathbb{E}}\zeta(t_k)\varphi^\top(t_k)\right)^{-1}\right]^T \tag{1.23}$$

avec $\tilde{\zeta}(t_k) = H_o(p)A_o(p)\zeta(t_k)$ et $\sigma_{e_o}^2$ est la variance de $\{e_o(t_k)\}$.

[5] La notation $\bar{\mathbb{E}}\{.\} = \lim_{N\rightarrow\infty}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\{.\}$ est adoptée de [Lju99].

1.5.3 Variable instrumentale étendue

L'estimée par VI étendue de ρ est obtenue en préfiltrant les données d'entrée-sortie figurant dans (1.19) et en généralisant l'estimée par variable instrumentale basique $\hat{\rho}_{iv}$ par l'augmentation de l'instrument $\zeta(t_k) \in \mathbb{R}^{n_\zeta}$ ($n_\zeta \geq n_a+n_b+1$) de façon à obtenir les équations du système surparamétré de la forme [SS83],

$$\hat{\rho}_{\mathtt{IV}}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N} L(q)\zeta(t_k)L(q)\varphi^\top(t_k)\right]\rho - \left[\frac{1}{N}\sum_{k=1}^{N} L(q)\zeta(t_k)L(q)y(t_k)\right]\right\|_W^2, \quad (1.24)$$

où $\|x\|_W^2 = x^\top W x$, avec W définie positive et $L(q)$ un filtre stable. Cette VI étendue demande plus de calcul que la VI basique. Néanmoins, l'augmentation de l'instrument peut améliorer la précision des estimées. Il est à noter que lorsque $L(q) = 1$ et $n_\zeta = n_a + n_b + 1$ ($W = I$), la VI basique (1.19) est obtenue.

1.5.3.1 Propriétés de convergence

Les propriétés de convergence sont trivialement obtenues en généralisant (1.21) à l'estimateur (1.24) :

C1 $\bar{\mathbb{E}}\{L(q)\zeta(t_k)L(q)\varphi^\top(t_k)\}$ est de rang plein (corrélation de ζ et φ).

C2 $\bar{\mathbb{E}}\{L(q)\zeta(t_k)L(q)\tilde{v}(t_k)\} = 0$ (décorrélation de ζ et $\tilde{v}$).

Une discussion sur l'obtention de C1 peut être trouvée dans [SS83]. C2 peut facilement être remplie mais le choix de l'instrument a un rôle majeur sur la qualité des estimées.

1.5.3.2 Analyse de précision

La distribution asymptotique du vecteur de paramètres (1.24) est obtenue en suivant le même raisonnement que dans le paragraphe 1.5.2.2. Ainsi, sous les hypothèses LTI1 à LTI4, $\hat{\rho}_{xiv}$ suit asymptotiquement une distribution Gaussienne,

$$\sqrt{N}(\hat{\rho}_{xiv} - \rho^*) \overset{dist}{\rightarrow} \mathcal{N}(0, \mathbf{P}_{xiv}) \quad (1.25)$$

où la matrice de covariance est donnée par

$$\mathbf{P}_{xiv} = \sigma_{e_{of}}^2 \left[\mathbf{R}^\top\mathbf{W}\mathbf{R}\right]^{-1}\mathbf{R}^\top\mathbf{W}\left[\bar{\mathbb{E}}\tilde{\zeta}(t_k)\tilde{\zeta}^\top(t_k)\right]\mathbf{W}\mathbf{R}\left[\mathbf{R}^\top\mathbf{W}\mathbf{R}\right]^{-1}$$

avec $\tilde{\zeta}(t_k) = L(q)H_o(q)A_o(q^{-1})\zeta(t_k)$ et $\mathbf{R} = \bar{\mathbb{E}}L(q)\zeta(t_k)L(q)\mathring{\varphi}^\top(t_k)$. $\mathring{\varphi}(t_k)$ est la version non bruitée du vecteur de régression $\varphi(t_k)$ (1.9), défini comme

$$\mathring{\varphi}^\top(t_k) = [-\chi(t_{k-1}) \ldots - \chi(t_{k-n_a})u(t_{k-1}) \ldots u(t_{k-n_b})], \quad (1.26)$$

où χ est la sortie non-bruitée :

$$\chi(t_k) = G_{\mathrm{o}}(q)u(t_k). \quad (1.27)$$

1.5.4 Variable instrumentale optimale

Le choix de l'instrument $\zeta(t_k)$, sa taille n_ζ, la matrice de pondération W ainsi que le préfiltre $L(q)$ ont une contribution considérable sur la valeur de la matrice de covariance $\mathbf{P}_{xiv}$. Dans le cas de la boucle ouverte à temps discret, la borne inférieure de $\mathbf{P}_{xiv}$ pour des méthodes d'estimation non biaisées est donnée par la borne de Cramér-Rao [Lju99], [SS83]. Le choix optimal des variables précitées est définie par l'égalité entre $\mathbf{P}_{xiv}$ et la borne de Cramér-Rao. La matrice de covariance optimale (la borne de Cramér-Rao) pour le système (1.1) est alors :

$$\begin{aligned} &\mathbf{P}_{xiv} \geq \mathbf{P}_{xiv}^{opt} \text{ and} \\ &\mathbf{P}_{xiv}^{opt} = \sigma^2_{e_{ofopt}} \left\{ \bar{\mathbb{E}} \Big[\big[A_o(q^{-1}) H_o(q)\big]^{-1} \mathring{\varphi}^\top(t_k) \Big]^\top \Big[\big[A_o(q^{-1}) H_o(q)\big]^{-1} \mathring{\varphi}^\top(t_k) \Big] \right\}^{-1}. \end{aligned} \tag{1.28}$$

$\mathbf{P}_{xiv}^{opt}$ est alors obtenue en choisissant :

C3 $W = I$;

C4 ζ est la version non bruitée du vecteur de régression φ ;

C5 n_ρ est égal au nombre minimal de paramètres nécessaires à la représentation de G_o (le modèle est minimal).

C6 $L(q)$ est choisi comme étant $Q(q, \theta_\mathrm{o})$ (voir 1.13) ;

Afin de remplir C4 et C6, un algorithme de relaxation et habituellement utilisé, où à chaque itération, l'instrument et les préfiltres associés sont mis à jour dépendamment des itérations précédentes. Une méthode itérative de l'implémentation de la VI optimale est présentée dans le paragraphe suivant.

1.5.5 Variable instrumentale optimale itérative

Au sein des méthodes de VI, la plus reconnue est sans doute la méthode nommée *Refined Instrumental Variable* (RIV) développée pour la première fois dans [YJ80]. Cette approche a été utilisée avec succès dans de nombreux cadres comme l'identification directe de modèles à temps continu [YJ80, RG04, GW08], l'identification en boucle fermée [You08, GGYdH09, GdH05] et atteint l'estimée statistiquement optimale sous la condition LTI1. De plus, il a été empiriquement démontré que l'estimée asymptotiquement non biaisée garde une variance relativement faible lorsque l'hypothèse sur le modèle de bruit est erronée [You08, Lju09]. Pour l'algorithme RIV, à chaque étape de l'algorithme de relaxation, un modèle auxiliaire est construit et simulé pour calculer l'instrument d'une part (ce qui garanti C2), et le préfiltre d'autre part. Ce modèle auxiliaire est construit à partir des estimées obtenues à l'itération précédente. Par conséquent, s'il y a convergence, C4 and C6 sont remplies.
L'algorithme RIV peut être résumé de la façon suivante :

Algorithm 1 (RIV)

Étape 1 *Générer une estimée initiale du vecteur de paramètres* $\hat{\rho}^{(0)}$ *(*e.g. *en utilisant la méthode des moindres carrés). Forcer* $D(q^{-1}, \hat{\eta}^{(0)}) = C(q^{-1}, \hat{\eta}^{(0)}) = 1$. *Initialiser* $\tau = 0$.

Étape 2 *Calculer une estimée de la version non bruitée* $y(t_k, \hat{\rho}^{(\tau)})$ *de la sortie* $y(t_k)$.

Étape 3 *Calculer le préfiltre :*

$$L(q,\hat{\theta}^{(\tau)}) = \frac{1}{H(q,\eta^{(\tau)})A(q,\rho^{(\tau)})} \tag{1.29}$$

Il est à noter que si l'un des filtres devient instable lors du processus itératif, il est nécessaire le stabiliser.

Étape 4 *Construire le vecteur de régression préfiltré $\varphi_{\mathrm{f}}(t_k)$ et l'instrument préfiltré $\zeta_{\mathrm{f}}(t_k)$. Dans un contexte boucle ouverte, il est suffisant de remplacer les valeurs retardées de y par les valeurs retardées de son estimée $y(t_k,\hat{\rho}^{(\tau)})$ (remplissant C2).*

Étape 5 *Résoudre le problème d'optimisation de la VI explicité dans* (1.24) *dont la solution est*

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k)\right]^{-1} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}(t_k). \tag{1.30}$$

Le résultat $\hat{\rho}^{(\tau+1)}(N)$ est l'estimée RIV à l'itération $\tau+1$.

Étape 6 *Une estimée du bruit v_{o} peut être obtenue comme étant*

$$\hat{v}(t_k) = y(t_k) - y(t_k,\hat{\rho}^{(\tau)}). \tag{1.31}$$

À partir de $\hat{v}$,l'estimation du vecteur de paramètres liés au modèle de bruit $\hat{\eta}^{(\tau+1)}$ peut être obtenue en utilisant par exemple l'algorithme `ARMA` *de la boîte à outils identification de* `MATLAB` *(une approche par VI peut également être utilisée [You08]).*

Étape 7 *Si $\theta^{(\tau+1)}$ a convergé ou si le nombre maximum d'itérations a été dépassé, arrêter, sinon, incrémenter τ de 1 et retourner à l'étape 2.*

Il est à noter qu'actuellement, même dans le cas linéaire, la preuve formelle de la convergence n'a pas été montrée. Néanmoins, elle est communément et empiriquement acceptée car elle est utilisée dans de nombreux exemples de simulation tout autant que de cas pratiques depuis plus de trente ans [You08].
L'erreur de prédiction et l'algorithme RIV viennent d'être présentés dans un context à temps discret. Cependant, afin d'appréhender le cas du temps continu, une discussion préliminaire doit être menée comme décrit dans la section suivante.

1.6 Estimation directe de modèles à temps continu

En pratique, les systèmes s'expriment en temps continu alors que leur identification ne peut se faire qu'à partir de données mesurées, donc échantillonnées. C'est pourquoi les approches d'identification se scindent en deux catégories :

- **les approches indirectes :** ces méthodes identifient un modèle à temps discret avant de transformer le modèle estimé en un modèle à temps continu,
- **les approches directes :** ces méthodes identifient directement la valeur des paramètres du modèle à temps continu à partir des données échantillonnées issues des signaux à temps continu.

Grâce aux développements récents en terme d'instrumentation et des vitesses d'échantillonnage croissantes, l'utilisation des approches directes connaissent un gain d'intérêt non négligeable et

montrent sous certaines conditions de meilleures performances que les méthodes indirectes pour les modèles linéaires ou non-linéaires (*cf e.g.* [RH06, LGGY08, You08, GW08]). Une bibliographie exhaustive des méthodes directes peut être trouvée dans [RH06].

1.6.1 Modèles à temps continu

Soit le système générant les données exprimé sous sa forme à temps continu :

$$\mathcal{S}_{\mathrm{o}} \begin{cases} \chi(t) = G_{\mathrm{o}}(\not{d})u(t) \\ y(t) = \chi(t) + H_{\mathrm{o}}(\not{d})e(t), \end{cases} \tag{1.32}$$

où $\not{d}$ est l'opérateur de différentiation. Les discussions présentées concernant la minimisation de l'erreur de prédiction pour les modèles à temps discret peuvent être menées de façon similaire dans le cas continu. Afin d'éviter une certaine redondance, ces résultats ne seront pas exposés dans le cas linéaire mais peuvent être trouvés dans [GW08]. Ils seront néanmoins présentés pour chaque type de modèle non-linéaire présenté dans la suite de ce document. En considérant le modèle à temps continu approprié,

$$\mathcal{M}_\theta : y_\theta(t) = \frac{B(\not{d}, \rho)}{A(\not{d}, \rho)}u(t) + H(\not{d}, \eta)e(t), \tag{1.33}$$

où $B(\not{d}, \rho)$ et $A(\not{d}, \rho)$ sont des polynômes fonction de $\not{d}$ d'ordre n_b et n_a respectivement. Les équations du système (1.32) peuvent être exprimées sous la forme de régression linéaire :

$$y^{(n_a)}(t) = \varphi^\top(t)\rho_{\mathrm{o}} + \tilde{v}(t), \tag{1.34}$$

avec ρ_{o} le vecteur de paramètre du processus et φ le vecteur de régression défini comme

$$\varphi^\top(t) = \left[y^{(n_a-1)}(t), \ldots, y(t), u^{(n_b)}(t), \ldots, u(t)\right]. \tag{1.35}$$

Le bruit $\tilde{v}(t)$ est égal à

$$\tilde{v}(t) = A_{\mathrm{o}}(\not{d})H_{\mathrm{o}}(\not{d})e_{\mathrm{o}}(t). \tag{1.36}$$

On peut constater que seul le cas réaliste de données échantillonnées peut être considéré en identification de systèmes. Ainsi, (1.34) est verifiée pour tout temps t et donc particulièrement aux temps d'échantillonnage t_k :

$$y^{(n_a)}(t_k) = \varphi^\top(t_k)\rho_{\mathrm{o}} + \tilde{v}(t_k), \tag{1.37}$$

avec

$$\varphi^\top(t_k) = \left[y^{(n_a-1)}(t_k), \ldots, y(t_k), u^{(n_b)}(t_k), \ldots, u(t_k)\right], \tag{1.38}$$

où $w^{(n)}(t_k)$ est la dérivée temporelle n^{ieme} du signal $w(t)$ au temps t_k.
En utilisant (1.37) et de la même façon que pour les modèles à temps discret, la VI étendue peut être définie comme :

$$\hat{\rho}_{\mathtt{IV}}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N} L(\not{d})\zeta(t_k)L(\not{d})\varphi^\top(t_k)\right]\rho - \left[\frac{1}{N}\sum_{k=1}^{N} L(\not{d})\zeta(t_k)L(\not{d})y^{(n_a)}(t_k)\right]\right\|_W^2, \tag{1.39}$$

avec $\|x\|_W^2 = x^\top W x$, W une matrice définie positive et $L(\not{d})$ un pré-filtre stable.
De même, le choix optimal des différentes variables peut être trivialement appliqué au cas continu :

$$\zeta^\top(t_k) = \left[\chi^{(n_a-1)}(t_k), \ldots, \chi(t_k), u^{(n_b)}(t_k), \ldots, u(t_k)\right], \tag{1.40}$$

$$L(\not{d}) = \frac{1}{H_o(\not{d})A_o(\not{d})}, \tag{1.41}$$

$$W = I, \tag{1.42}$$

$$n_\zeta = n_a + n_b + 1. \tag{1.43}$$

Néanmoins, ces résultats soulèvent deux questions importantes liées à la représentation à temps continu du modèle :

– l'instrument et le vecteur de régression contiennent des dérivées temporelles des signaux qui sont implicitement considérées comme connues dans le développement précédent. Naturellement, dans le contexte de signaux échantillonnés, cette hypothèse est violée et ainsi, seules les estimées sont accessibles ;
– l'identification directe de modèles à temps continu, même dans le cas LTI, implique la gestion de sources de bruit à temps continu et des problèmes mathématique liés [PSR00].

1.6.2 Gérer les dérivées non-mesurées

En comparaison de son homologue à temps discret, l'identification de modèles à temps continu comporte quelques problèmes intrinsèques. Le premier est lié à l'implémentation. Contrairement aux équations aux différences finies, les équations différentielles contiennent des dérivées temporelles qui étaient considérées comme déterministes jusqu'ici. Cependant, les dérivées de ces signaux sont généralement inaccessibles. Les approches standard à temps continu introduisent généralement un filtre passe-bas

$$y_{f_c}(t_k) = f_c(\not{d})y(t_k), \tag{1.44}$$

où l'indice f_c dénote la forme préfiltrée du signal concerné. Les dérivées temporelles peuvent alors être obtenues en filtrant les signaux d'entrée sortie avec un banc de filtres de la forme $f_c(\not{d})p^i$

$$y_{f_c}^{(i)}(t_k) = (f_c(\not{d})p^i y)(t_k), \quad i \leq n_a \tag{1.45}$$

$$u_{f_c}^{(i)}(t_k) = (f_c(\not{d})p^i u)(t_k), \quad i \leq n_b. \tag{1.46}$$

La motivation invoquée est que les signaux filtrés $u_{f_c}(t)$ et $y_{f_c}(t)$ satisfont la relation

$$y_{f_c}(t) = G_o(\not{d})u_{f_c}(t) + f_c(\not{d})H_o(\not{d})e_o(t) \tag{1.47}$$

i.e. la fonction de transfert ne change pas mais la fonction de transfert du bruit est modifiée par l'introduction du filtre. L'équation (1.47) peut alors être récrite sous une forme de régression linéaire

$$y_{f_c}^{(n_a)}(t_k) = \varphi_{f_c}^\top(t_k)\rho_o + \tilde{v}_{f_c}(t_k) \tag{1.48}$$

avec

$$\varphi_{f_c}^\top(t) = [-y_{f_c}^{(n_a-1)}(t) \ \cdots \ -y_{f_c}(t) \ u_{f_c}^{(n_b)}(t) \ \cdots \ u_{f_c}(t)] \tag{1.49}$$

$$\tilde{v}_{f_c}(t) = f_c(\not{d})\tilde{v}(t). \tag{1.50}$$

Différents types de filtres ont été utilisés pour solutionner le problème de reconstruction des dérivées [GMR03] et la boîte à outils *CONtinuous-Time System IDentification* (CONTSID) a été développée sur la base de ces méthodes (voir [GW08]).
Un filtre usuel souvent utilisé pour les méthodes de VI basiques est [GMR03]

$$f_c(\not{d}) = \left(\frac{\beta}{\not{d} + \lambda}\right)^{n_a} \tag{1.51a}$$

où $f_c(\not{d})$ représente le filtre d'ordre minimum lors de filtrages multiples. Le désavantage de ces filtres est le choix difficile d'une variable définissant une fréquence de coupure appropriée pour le système considéré. En conséquence, un avantage important de la méthode RIV est l'utilisation du filtre $A(\not{d}, \rho)^{-1}$ pour deux usages : la minimisation de l'erreur de prédiction et l'approximation des dérivées temporelles.

1.6.3 Modèles hybrides

Les méthodes directes impliquent la gestion de signaux de bruit à temps continu et des problèmes mathématiques associés [PSR00]. Par conséquent, l'idée pour palier ce problème est de considérer le bruit à temps continu échantillonné comme des échantillons d'un bruit à temps discret filtré par un filtre à temps discret [Joh94]. Cela résulte en des modèles hybrides ou le processus est exprimé à temps continu alors que le bruit est exprimé à temps discret :

$$\begin{cases} \chi(t) = G(\not{d})u(t), \\ y(t_k) = \chi(t_k) + H(q)e(t_k). \end{cases} \tag{1.52}$$

Néanmoins, cette représentation soulève un autre problème : le mélange de filtres à temps continu et à temps discret. Une discussion préliminaire doit être conduite afin de définir comment une telle coexistence est possible dans le contexte présenté.
En terme d'identification, la seule hypothèse valide est de considérer des échantillons de signaux continus. Afin d'appliquer un filtre à temps continu sur un signal échantillonné, seules deux solutions sont disponibles. La première est d'interpoler les échantillons pour obtenir un signal à temps continu et appliquer le filtre à temps continu. La seconde est d'utiliser une approximation à temps discret du filtre à temps continu considéré. Ceci est un problème connu pour la simulation de modèles à temps continu. Dans une optique de simulation, les approximations numériques sont gérées efficacement en utilisant des algorithmes numériques performants [Atk89]. Il est habituellement suffisant de considérer que les signaux d'entrée à temps continu sont constants entre deux échantillons [GGS00]. Ceci a également été montré dans le cas des systèmes LPV [Tót10]. Cette hypothèse est également utilisée dans [PSR00], [Joh94] pour les signaux (y, p, u). Cette hypothèse est renforcée par les nouvelles technologies d'acquisition, toujours plus rapides. Sous ces conditions, il peut être considéré que [GW08] :

$$\left(F(\not{d})y\right)(t_k) = F(\not{d})y(t_k). \tag{1.53}$$

En d'autres termes, en considérant qu'un filtre à temps continu ne peut être appliqué à des données échantillonnées qu'au travers d'une approximation numérique, les propriétés usuelles telles que la commutativité restent vraies entre un filtre à temps discret et l'approximation numérique d'un filtre

à temps continu. Cependant, il est important de noter que l'approximation numérique n'est utilisée que pour la simulation et que par conséquent, les coefficients estimés restent ceux du modèle à temps continu quelle que soit la méthode de discrétisation choisie.

1.7 Organisation de la thèse et principales contributions

Dans le contexte d'application précédemment décrit, deux structures de modèles présentent un intérêt particulier : la structure de type Hammerstein et la structure de type LPV. Ces deux structures sont étudiées dans le présent document dans le cas *mono-entrée-mono-sortie* (SISO). Alors que la structure Hammerstein a été très largement étudiée en temps discret, peu de méthodes ont été proposées pour l'identification de modèles Hammerstein à temps continu. En ce qui concerne les modèles LPV, un large écart est observé entre les besoins du domaine de la commande et les solutions inadaptées des méthodes d'identification actuelles. De plus, aucune méthode à notre connaissance n'a été proposée pour répondre au problème d'identification de modèles LPV à temps continu. De plus, les méthodes fondées sur l'exploitation de la variable instrumentale optimale semblent être prometteuses pour l'identification de systèmes naturels de par i) leurs performances statistiques lorsque l'hypothèse émise sur le bruit est fausse ii) leur application avec succès dans un contexte linéaire, à la fois à temps discret et à temps continu.

Par conséquent, le chapitre 2 se concentre sur des méthodes d'estimation pour les modèles Hammerstein et se divise en trois parties : i) le cas discret en boucle ouverte, ii) le cas discret en boucle fermée, iii) le cas continu qui constitue la principale contribution de cette thèse aux modèles Hammerstein. Bien que l'erreur de prédiction ait déjà été définie pour les modèles Hammerstein à temps discret dans [Lju99], ce chapitre donne une bonne idée de la possible transformation de modèles non-linéaires en modèles linéaires grâce à un paramétrage linéaire en les paramètres ainsi que de l'extension de la variable instrumentale optimale pour les modèles non-linéaires. Ce chapitre expose également à travers des exemples représentatifs, les avantages et inconvénients d'utiliser une régression linéaire pour ce type de modèles, les avantages de l'estimation directe de modèles à temps continu ainsi que les performances des méthodes de variable instrumentale.
De plus, le chapitre 3 introduit les résultats obtenus pour l'estimation des modèles LPV. Une fois encore les travaux se divisent entre i) le cas discret en boucle ouverte, ii) le cas discret en boucle fermée, iii) le cas continu en boucle ouverte. La principale contribution de ce chapitre réside dans le développement d'une des premières méthodes visant à minimiser l'erreur de prédiction dans le cas de modèles LPV Box-Jenkins, tant à temps discret qu'à temps continu. Les résultats obtenus par la méthode développée dans cette thèse sont comparés aux résultats des méthodes existant dans la littérature au travers d'exemples représentatifs.
Les principales contributions théoriques de cette thèse sont l'introduction de méthodes RIV appliquées aux types de modèles non-linéaires définis, à temps continu, à temps discret, en boucle ouverte ou fermée. En conséquence les sections de ces chapitres théoriques suivent des structures similaires :
– définition du système d'intérêt ;
– définition du modèle associé ;
– définition de l'erreur de prédiction ;

– définition de l'instrument optimal et du préfiltre optimal ;
– algorithme détaillé de la méthode.

Afin de clore cette thèse, les résultats théoriques exposés sont appliqués dans le chapitre 4 pour la modélisation de la relation pluie/débit d'un bassin versant viticole alsacien en utilisant des données réelles. Un processus d'identification ainsi qu'un modèle innovant sont proposés et détaillés. Les performances de la méthode fondée sur la RIV sont comparés à celles de la méthode des moindres carrés appliquée au même modèle LPV.

Enfin, afin de ne pas surcharger l'introduction principale de cette thèse, les tenants et aboutissants ainsi que l'état de l'art liés à chaque type de modèles sont exposés en détail en début de chaque chapitre. De plus, les papiers soumis par l'auteur sont cités tout au long du document et mélangés à d'autres références bibliographiques. Par conséquent, une liste détaillée des travaux publiés au cours de la thèse est donnée ici.

1.8 Travaux publiés

Revue internationale (1) :

V. Laurain, M. Gilson, R. Tóth and H. Garnier. Refined Instrumental Variable Methods for Identification of LPV Box-Jenkins Models. *Automatica, Vol. 46, Issue 6 : 959–967, June 2010.*

Article soumis à une revue internationale (1) :

V. Laurain, R. Tóth, M. Gilson and H. Garnier. Direct Identification of Continuous-time LPV Input/Ouput Models. *Submitted to special issue "Continuous-time model Identification", IET Control Theory & Applications, April 2010.*

Communication invitée dans une conférence internationale (1) :

V. Laurain, M. Gilson, R. Tóth and H. Garnier. Identification of LPV Box-Jenkins Models via Optimal Refined Instrumental Variable Methods. *Proceedings of the American Control Conference 2010*, invited session "Identification of LPV models", Baltimore, Maryland, USA, June 2010.

Communication invitée soumise dans une conférence internationale (1) :

V. Laurain, M. Gilson, R. Tóth and H. Garnier. Direct Identification of continuous-time LPV models. *Submitted to the American Control Conference 2011*, invited session "LPV system identification and application".

Chapitres de livres invités (2) :

V. Laurain, M. Gilson and H. Garnier. Refined Instrumental variables for Hammerstein models. *In System Identification, Environmetric Modelling and Control Systems Design*, L. Wang, H. Garnier, T. Jakeman (Eds), Springer Verlag, to appear, 2011.

V. Laurain, M. Gilson, R. Tóth and H. Garnier. Chapter Refined Instrumental variables for LPV

models. *In Identification of LPV systems*, C. Novara, D. Rivera, T. Perdicoúlis, J. Ramos (Eds), World Scientific, to appear, 2011.

Communications dans des conférences internationales (6) :

V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire and H. Garnier. A new data-based modelling method for identifying parsimonious nonlinear rainfall/flow models. *International Congress on Environmental Modelling and Software (IEMSS 2010)*, Ottawa, Ontario, Canada, July 2010.

V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire and H. Garnier. Identification de modèles LPV : application à la modélisation pluie/débit d'un bassin versant viticole. *Conference Internationale Francophone d'Automatique (CIFA 2010)*, Nancy, France, June 2010.

V. Laurain, M. Gilson and H. Garnier. Refined Instrumental Variable Methods for Identifying Hammerstein Models Operating in Closed Loop. *Proceedings of the 48th IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, December 2009.

D. Kuss, **V. Laurain**, H. Garnier, M. Zug and J. Vasquez. Data-based mechanistic rainfall-runoff continuous-time modelling in urban context. *15th IFAC Symposium on System Identification (SYSID 2009)*, Saint-Malo, France, July 2009.

H. Garnier, M. Gilson, **V. Laurain**. The CONTSID toolbox for Matlab : extensions and latest developments. *15th IFAC Symposium on System Identification (SYSID 2009)*, Saint-Malo, France, July 2009.

V. Laurain, M. Gilson, H. Garnier and P. C. Young. Refined Instrumental Variable Methods for Identification of Hammerstein Continuous-time Box-Jenkins Models. *Proceedings of the 47th IEEE Conference on Decision and Control (CDC 2008)*, Cancún, Mexico, December 2008.

Communication soumise en conférence internationale (1) :

R. Tóth, **V. Laurain**, M. Gilson and H. Garnier. On the closed loop identification of LPV models using instrumental variables. *Submitted to IFAC World Congress 2011.*

Communications (4) :

V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire and H. Garnier. Data-based modelling of rainfall/runoff relationship in an agricultural catchment. *European Geosciences Union General Assembly 2010*, Vienna, Austria, May 2010.

V. Laurain, M. Gilson, R. Tóth and H. Garnier. Identification de modèles LPV Box-Jenkins : application à la modélisation de la relation pluie/débit dans un bassin versant rural, Réunion du Groupe de Travail "Identification des Systèmes" du GdR MACS, Paris, March 2010

V. Laurain, M. Gilson, R. Tóth and H. Garnier. Refined Instrumental Variable Methods for Identification of LPV Output-Error and Box-Jenkins Models. *Annual Workshop of the European Research Network on System Identification (ERNSI 2009)*, Vorau, Austria, September 2009.

V. Laurain, M. Gilson, H. Garnier and P. C. Young. An IV-based method for non-linear continuous-

time Hammerstein model identification. Application to rainfall-flow modelling. *Annual Workshop of the European Research Network on System Identification (ERNSI 2008)*, Sigtuna, Sweden, September 2008.

Chapitre 2

Refined Instrumental Variable Methods for Hammerstein Box–Jenkins Models

2.1 Introduction

Hammerstein block diagram model is widely represented for modelling nonlinear systems [FE10], [DC05], [BC08], [SWGP07]. The nonlinear block can be represented as a piecewise linear function [Bai02b] or as a sum of basis functions [PMEB$^+$07], [DSC07].

Among the very recent work on DT Hammerstein models in the time domain, the most exposed methods are the extended least squares for Hammerstein ARMAX models [DC05] which were further extended to Hammerstein OE models [DSC07]. E.R Bai exposed a two stage algorithm involving least squares and single value decomposition used in different configurations [Bai98], [LB07], [BC08] and was very recently analysed for Hammerstein Box–Jenkins models [WZL09]. Nonetheless, the convergence properties of the algorithm are studied but there was no study driven in case of noise modelling error. Suboptimal Hammerstein model estimation in case of a bounded noise was studied in [CR03]. A blind maximum likelihood method is derived in [VPS08] but the output signal is considered to be errorless.

The problem of nonlinear plants operating in closed loop have not received a lot of attention so far. Indeed, this problem is handled only in a few papers as in *e.g* [FAD01], [LAD01], [DAL02], and [WJC07] and most methods are designed for OE models.

Concerning the CT case, an exhaustive survey by Rao and Unbehauen [RH06] shows that CT model identification methods applied to Hammerstein models are poorly studied in the literature. In [PU95], the authors focus on the time-derivative approximation problems while solving the optimization problem using least squares. A non-parametric method can be found in [Gre00] while an approach dedicated to periodic input signals can be found in [ZRBW06]. To the best of the author's knowledge, the parametric estimation problem has not been addressed yet for CT Hammerstein models which focus with some colored added noise.

Consequently very few methods were developed in order to cope with the issues exposed in 1.2 and therefore the contributions of this thesis are exposed in the following manner.

Section 2.2 shows how the LTI-IV method can be extended in order to deal with Hammerstein BJ models. Moreover, the development of instrumental variable techniques able to cope with the

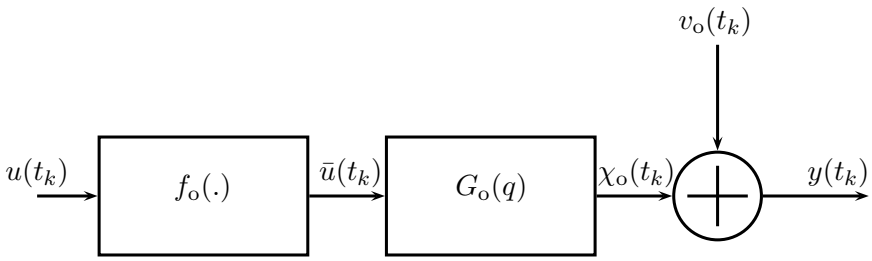


FIG. 2.1 – Hammerstein block representation

closed-loop case and the direct continuous-time model estimation in colored noise conditions and are exposed in Section 2.3 and Section 2.4 respectively. All presented method estimates are statistically analyzed through relevant Monte Carlo simulations and the features of the proposed method are studied in the different pre-cited contexts.

2.2 Discrete-time open-loop Hammerstein model identification

2.2.1 System description

Consider the Hammerstein system represented in Figure 2.1 and assume that both input and output signals, $u(t_k)$ and $y(t_k)$ are uniformly sampled at a constant sampling time $\mathtt{T_s}$ over N samples. The Hammerstein system $\mathcal{S}_o$, is described by the following input-output relationship :

$$\mathcal{S}_o \begin{cases} \bar{u}(t_k) = f(u(t_k)) \\ \chi_o(t_k) = G_o(q)\bar{u}(t_k) \\ y(t_k) = \chi_o(t_k) + v_o(t_k) \end{cases} \tag{2.1}$$

where u and y are the deterministic input and noisy output respectively, χ_o is the noise-free output and v_o the additive noise with bounded spectral density. $G_o(q)$ is the linear transfer function which can be written as

$$G_o(q) = \frac{B_o(q^{-1})}{A_o(q^{-1})}, \tag{2.2}$$

where $B_o(q^{-1})$ and $A_o(q^{-1})$ are polynomial in q^{-1} of degree n_b and n_a respectively :

$$A_o(q^{-1}) = 1 + \sum_{i=1}^{n_a} a_i^o q^{-i}, \tag{2.3a}$$

$$B_o(q^{-1}) = \sum_{j=0}^{n_b} b_j^o q^{-j}, \tag{2.3b}$$

where the coefficients a_i^o and $b_j^o \in \mathbb{R}$. In this thesis, the most general case is considered where the colored noise associated with the sampled output measurement $y(t_k)$ is assumed to have a rational spectral density which might have no relation to the actual process dynamics of $\mathcal{S}_o$. Therefore, v_o is represented by a discrete-time *autoregressive moving average* (ARMA) model :

$$v_o(t_k) = H_o(q)e_o(t_k) = \frac{C_o(q^{-1})}{D_o(q^{-1})}e_o(t_k), \tag{2.4}$$

where $C_{\mathrm{o}}(q^{-1})$ and $D_{\mathrm{o}}(q^{-1})$ are monic polynomials with constant coefficients and with respective degree n_{c} and n_{d}. Furthermore, all roots of $z^{n_{\mathrm{d}}}D_{\mathrm{o}}(z^{-1})$ and $z^{n_{\mathrm{c}}}C_{\mathrm{o}}(z^{-1})$ are inside the unit disc. It can be noticed that in case $C_{\mathrm{o}}(q^{-1}) = D_{\mathrm{o}}(q^{-1}) = 1$, (2.4) defines an OE noise model. It can be noticed that the same theory could be straightforwardly used if some *a priori* known pure delay was present on the input but this case is not exposed here for clarity's sake.

2.2.2 Model considered

Next we introduce a discrete-time Hammerstein Box-Jenkins (BJ) type of model structure that we propose for the identification of the data-generating system (2.1) with noise model (2.4). In the chosen model structure, the noise model and the process model are parameterized separately.

2.2.2.1 Linear part of the Hammerstein model

The linear process model is denoted by $\mathcal{L}_{\rho_{\mathrm{L}}}$ and is defined in a linear representation form as :

$$\mathcal{L}_{\rho_{\mathrm{L}}} : \big(A(q^{-1}, \rho_{\mathrm{L}}), B(q^{-1}, \rho_{\mathrm{L}})\big) \tag{2.5}$$

where the polynomials A and B are parameterized as

$$\mathcal{L}_{\rho_{\mathrm{L}}} \quad \begin{cases} A(q^{-1}, \rho_{\mathrm{L}}) = 1 + \sum_{i=1}^{n_{\mathrm{a}}} a_i q^{-i}, \\ B(q^{-1}, \rho_{\mathrm{L}}) = \sum_{j=0}^{n_{\mathrm{b}}} b_j q^{-j}, \end{cases}$$

The associated model parameters ρ_{L} are stacked columnwise :

$$\rho_{\mathrm{L}} = \begin{bmatrix} a_1 & \dots & a_{n_a} & b_0 & \dots & b_{n_b} \end{bmatrix}^\top \in \mathbb{R}^{n_a+n_b+1} \tag{2.7}$$

Introduce also $\mathfrak{L} = \{\mathcal{L}_{\rho_{\mathrm{L}}} \mid \rho \in \mathbb{R}^{n_{\rho_{\mathrm{L}}}}\}$, as the collection of all process models in the form of (2.5).

2.2.2.2 Nonlinear part of the Hammerstein model

The static nonlinearity model is denoted by $\mathcal{F}_{\rho_{\mathrm{NL}}}$ and defined :

$$\mathcal{F}_{\rho_{\mathrm{NL}}} : (f(u, \rho_{\mathrm{NL}})) \tag{2.8}$$

where $f(u, \rho_{\mathrm{NL}})$ is parameterized as a sum of basis functions

$$f(u(t_k), \rho_{\mathrm{NL}}) = \sum_{i=1}^{l} \alpha_i(\rho_{\mathrm{NL}})\gamma_i(u(t_k)). \tag{2.9}$$

In this parametrization, $\{\gamma_i\}_{i=1}^{l}$ are meromorphic functions [1] of $u(t_k)$ which are assumed to be *a priori* known. Furthermore, they have a static dependence on u, and are chosen such that they allow the identifiability of the model (pairwise orthogonal functions on $\mathbb{R}$ for example). The associated model parameters ρ_{NL} is stacked columnwise :

$$\rho_{\mathrm{NL}} = \begin{bmatrix} \alpha_1 & \dots & \alpha_l \end{bmatrix}^\top \in \mathbb{R}^l, \tag{2.10}$$

Introduce also $\mathcal{F} = \{\mathcal{F}_{\rho_{\mathrm{NL}}} \mid \rho_{\mathrm{NL}} \in \mathbb{R}^l\}$, as the collection of all process models in the form of (2.8).

[1] $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a real meromorphic function if $f = g/h$ with g, h analytic and $h \neq 0$

Remark

Note that the Hammerstein model $(\beta f(u,\rho_{\mathrm{NL}}), \frac{G(q,\rho_{\mathrm{L}})}{\beta})$ produces the same input-output data for any β. Therefore, to get a unique parametrization, the gain of $(\beta f(u,\rho_{\mathrm{NL}})$ or $G(q,\rho_{\mathrm{L}})/\beta$ has to be fixed [DC05], [Bai02a]. Hence, the first coefficient of the function $f(.)$ is fixed to 1, *i.e.* $\alpha_1 = 1$ in (2.10).

2.2.2.3 Noise model

The noise model denoted by $\mathcal{H}$ is defined as an LTI transfer function :

$$\mathcal{H}_\eta : (H(q,\eta)) \tag{2.11}$$

where H is a monic rational function given in the form of

$$H(q,\eta) = \frac{C(q^{-1},\eta)}{D(q^{-1},\eta)} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_{\mathrm{c}}} q^{-n_{\mathrm{c}}}}{1 + d_1 q^{-1} + \ldots + d_{n_{\mathrm{d}}} q^{-n_{\mathrm{d}}}}. \tag{2.12}$$

The associated model parameters η are stacked columnwise in the parameter vector,

$$\eta = \begin{bmatrix} c_1 & \ldots & c_{n_{\mathrm{c}}} & d_1 & \ldots & d_{n_{\mathrm{d}}} \end{bmatrix}^\top \in \mathbb{R}^{n_\eta}, \tag{2.13}$$

where $n_\eta = n_{\mathrm{c}} + n_{\mathrm{d}}$. Additionally, denote $\mathcal{H} = \{\mathcal{H}_\eta \mid \eta \in \mathbb{R}^{n_\eta}\}$, the collection of all noise models in the form of (2.11).

2.2.2.4 Whole Hammerstein model

With respect to a given nonlinear, linear process and noise part $(\mathcal{F}_{\rho_{\mathrm{NL}}}, \mathcal{L}_{\rho_{\mathrm{L}}}, \mathcal{H}_\eta)$, the parameters can be collected as

$$\theta_H = \begin{bmatrix} \rho_{\mathrm{L}}^\top & \rho_{\mathrm{NL}}^\top & \eta^\top \end{bmatrix}, \tag{2.14}$$

and the signal relations of the Hammerstein BJ model, denoted in the sequel as $\mathcal{M}_\theta$, are defined as :

$$\mathcal{M}_{\theta_{\mathrm{H}}} \begin{cases} \bar{u}(t_k) = \sum_{i=1}^{l} \alpha_i(\rho_{\mathrm{NL}})\gamma_i(u(t_k)) \\ A(q^{-1},\rho_{\mathrm{L}})\chi(t_k) = B(q^{-1},\rho_{\mathrm{L}})\bar{u}(t_k) \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{2.15}$$

Based on this model structure, the model set, denoted as $\mathcal{M}$, with the linear process $(\mathcal{L}_{\rho_{\mathrm{L}}})$, the nonlinearity $(\mathcal{F}_{\rho_{\mathrm{NL}}})$ and noise $(\mathcal{H}_\eta)$ models parameterized independently, takes the form

$$\mathcal{M} = \left\{ (\mathcal{F}_{\rho_{\mathrm{NL}}}, \mathcal{L}_{\rho_{\mathrm{L}}}, \mathcal{H}_\eta) \;\mid\; \mathrm{col}(\rho_{\mathrm{NL}}, \rho_{\mathrm{L}}, \eta) = \theta_{\mathrm{H}} \in \mathbb{R}^{n_{\rho_{\mathrm{NL}}} + n_{\rho_{\mathrm{L}}} + n_\eta} \right\}. \tag{2.16}$$

2.2.2.5 Reformulation of the model

The error prediction minimization optimization problem is not convex in general. However, it can be clearly seen from the parametrization (2.9) that the model (2.15) can be rewritten in order to obtain a linear regression structure. By combining the first two equations in (2.15), the model can be rewritten as :

$$\mathcal{M}_{\theta_{\mathrm{H}}} \begin{cases} A(q^{-1},\rho_{\mathrm{L}})\chi(t_k) = B(q^{-1},\rho_{\mathrm{L}}) \sum_{i=1}^{l} \alpha_i(\rho_{\mathrm{NL}})\gamma_i(u(t_k)) \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{2.17}$$

which can be expanded as (note that for clarity's sake $\gamma_i(u(t_k))$ is denoted $u_i(t_k)$ in the sequel)

$$\mathcal{M}_{\theta_{\mathrm{H}}} \begin{cases} A(q^{-1},\rho_{\mathrm{L}})\chi(t_k) = \sum_{i=1}^{l} \underbrace{\alpha_i(\rho_{\mathrm{NL}})B(q^{-1},\rho_{\mathrm{L}})}_{B_i(q^{-1},\rho_{\mathrm{NL}},\rho_{\mathrm{L}})} \underbrace{\gamma_i(u(t_k))}_{u_i(t_k)} \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k). \end{cases} \tag{2.18}$$

Under these modelling settings, the nonlinearity model and the linear process model can be combined into the process model, denoted by $\mathcal{G}_\rho$ and defined in the form :

$$\mathcal{G}_\rho : \big(A(q^{-1},\rho), B_i(q^{-1},\rho)\big) \tag{2.19}$$

where the polynomials A and B_i are given by

$$\mathcal{G}_\rho \quad \begin{cases} A(q^{-1},\rho) = 1 + \displaystyle\sum_{i=1}^{n_{\mathrm{a}}} a_i q^{-i}, \\ B_i(q^{-1},\rho) = \alpha_i \displaystyle\sum_{j=0}^{n_{\mathrm{b}}} b_j q^{-j}, i = 1 \dots l, \quad \alpha_1 = 1. \end{cases}$$

The associated model parameters are stacked columnwise in the parameter vector ρ,

$$\rho = \begin{bmatrix} \mathbf{a} \\ \alpha_1 \mathbf{b} \\ \vdots \\ \alpha_l \mathbf{b} \end{bmatrix} \in \mathbb{R}^{n_\rho}, \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \end{bmatrix} \in \mathbb{R}^{n_a}, \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix} \in \mathbb{R}^{n_b+1}, \tag{2.21}$$

with $n_\rho = n_a + l(n_b + 1)$.
Introduce also $\mathcal{G} = \{\mathcal{G}_\rho \mid \rho \in \mathbb{R}^{n_\rho}\}$, as the collection of all process models in the form of (2.19).
Finally, with respect to the given process and noise part $(\mathcal{G}_\rho, \mathcal{H}_\eta)$, the parameters can be collected as $\theta = [\ \rho^\top \quad \eta^\top\ ]$ and the signal relations of the Hammerstein BJ model, denoted in the sequel as $\mathcal{M}_\theta$, are defined as :

$$\mathcal{M}_\theta : y(t_k) = \frac{\sum_{i=1}^{l} B_i(q^{-1},\rho)u_i(t_k)}{A(q^{-1},\rho)} + \frac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k), \tag{2.22}$$

with $B_i(q^{-1},\rho) = \alpha_i B(q^{-1},\rho)$ and $u_i(t_k) = \gamma_i(u(t_k))$.

Based on this model structure, the whole model set including the process ($\mathcal{G}_\rho$) and noise ($\mathcal{H}_\eta$) models parameterized independently, is denoted as $\mathcal{M}$ and takes finally the form

$$\mathcal{M} = \left\{ (\mathcal{G}_\rho, \mathcal{H}_\eta) \;\mid\; \mathrm{col}(\rho, \eta) = \theta \in \mathbb{R}^{n_\rho + n_\eta} \right\}. \tag{2.23}$$

The set (2.23) corresponds to the set of candidate models in which we seek the best fitting model using data gathered from $\mathcal{S}_\mathrm{o}$ under a given identification criterion (cost function).

Remarks

It has to be noticed that this model transforms the Hammerstein structure into an augmented LTI *Multi Input Single Ouput* model structure such as presented in Figure 2.2. Consequently, the number of parameters to be estimated is not minimal as $n_\rho = n_a + l(n_b + 1)$ which is in general greater than $n_{\rho_\mathrm{L}} + n_{\rho_\mathrm{NL}} = n_a + l + (n_b + 1)$. Therefore, as the model is not minimal, the optimal estimation of this augmented MISO model does not correspond to the optimal estimates of the true Hammerstein model. Nonetheless, the gain granted using this modelling is the possible linear regression form and therefore, the convexification of the optimization problem. In order to define the identification problem it is firstly necessary to define minimization criterion. Nonetheless, the augmented model structure given in (2.22) is now an LTI structure, and therefore, the PEM framework from [Lju99] as described in Section 1.4 can be directly used here.

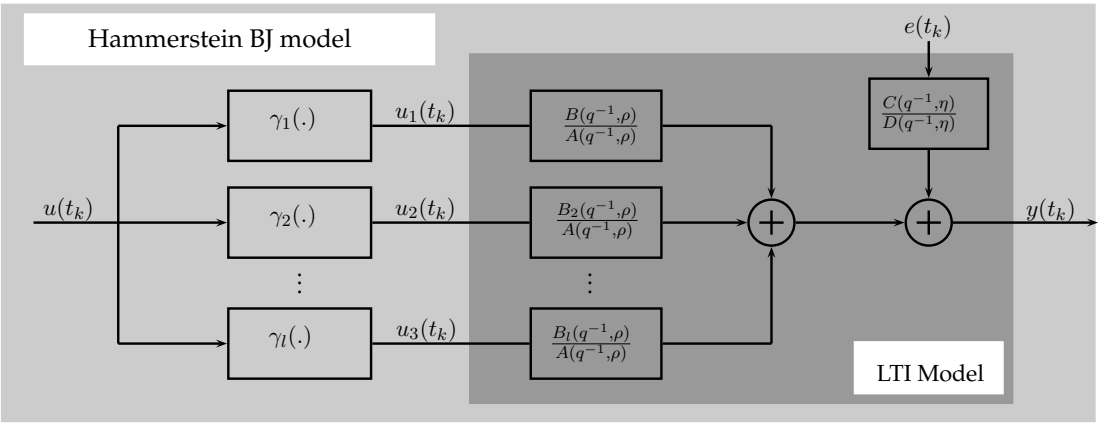


FIG. 2.2 – Hammerstein augmented model

2.2.3 Identification problem statement

Based on the previous considerations, the identification problem addressed can now be stated.

Problem 2 *Given a discrete time Hammerstein data generating system $\mathcal{S}_\mathrm{o}$ defined as in (2.1) and a data set $\mathcal{D}_N$ collected from $\mathcal{S}_\mathrm{o}$. Based on the Hammerstein BJ model structure $\mathcal{M}_\theta$ defined by (2.22), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions :*

HA1 *$\mathcal{S}_\mathrm{o} \in \mathcal{M}$, i.e. there exists a θ_o defining a $\mathcal{G}_{\rho_\mathrm{o}} \in \mathcal{G}$ and a $\mathcal{H}_{\eta_\mathrm{o}} \in \mathcal{H}$ such that $(\mathcal{G}_{\rho_\mathrm{o}}, \mathcal{H}_{\eta_\mathrm{o}})$ is equal to $\mathcal{S}_\mathrm{o}$.*

HA2 *$u(t_k)$ is not correlated to $e_\mathrm{o}(t_k)$.*

HA3 *$\mathcal{D}_N$ is informative with respect to $\mathcal{M}$.*

HA4 *G_o is BIBO stable,* i.e *for any bounded input signal u, the output of $\mathcal{S}_o$ is bounded.*

2.2.4 Optimal IV for Hammerstein models

The *Hammerstein RIV* (HRIV) method derives from the RIV algorithm for DT LTI systems. This was evolved by converting the maximum likelihood estimation equations to a pseudo-linear form involving optimal prefilters [You76], [YJ80]. A similar analysis can be utilised in the present situation since the problem is very similar, in both algebraic and statistical terms. The augmented model (2.22) then takes the linear regression form [YGG08] :

$$y(t_k) = \varphi^\top(t_k)\rho + \tilde{v}(t_k) \tag{2.24}$$

where ρ is as described in (2.21), $\tilde{v}(t_k) = A(q^{-1},\rho)v(t_k)$ and

$$\varphi(t_k) = \begin{bmatrix} -\mathbf{y}(t_k) \\ \mathbf{u_1}(t_k) \\ \vdots \\ \mathbf{u_l}(t_k) \end{bmatrix}, \mathbf{y}(t_k) = \begin{bmatrix} y(t_{k-1}) \\ \vdots \\ y(t_{k-n_a}) \end{bmatrix}, \mathbf{u_i}(t_k) = \begin{bmatrix} u_i(t_k) \\ \vdots \\ u_i(t_{k-n_b}) \end{bmatrix},$$

Using the conventional PEM approach on (2.24) leads to the prediction error $\varepsilon_\theta(t_k)$ given as :

$$\varepsilon_\theta(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)} \left\{ y(t_k) - \sum_{i=1}^{l} \frac{B_i(q^{-1},\rho)}{A(q^{-1},\rho)} u_i(t_k) \right\}, \tag{2.25}$$

which can be written as

$$\varepsilon_\theta(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)A(q^{-1},\rho)} \left\{ A(q^{-1},\rho)y(t_k) - \sum_{i=1}^{l} B_i(q^{-1},\rho)u_i(t_k) \right\}, \tag{2.26}$$

where the prefilter $D(q^{-1},\eta)/C(q^{-1},\eta)$ will be recognised as the inverse of the ARMA(n_c,n_d) noise model. However, since the polynomial operators commute in this linear case, (2.26) can be considered in the alternative form :

$$\varepsilon_\theta(t_k) = A(q^{-1},\rho)y_f(t_k) - \sum_{i=1}^{l} B_i(q^{-1},\rho)u_{if}(t_k) \tag{2.27}$$

where $y_f(t_k)$ and $u_{if}(t_k)$ represent the outputs of the prefiltering operation using the filter :

$$Q(q,\theta) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)A(q^{-1},\rho)}. \tag{2.28}$$

Therefore, from (2.27), the associated linear-in-the-parameters model then takes the form :

$$y_f(t_k) = \varphi_f^\top(t_k)\rho + \tilde{v}_f(t_k) \tag{2.29}$$

where

$$\varphi_f(t_k) = \begin{bmatrix} -\mathbf{y}_f(t_k) \\ \mathbf{u_{1}}_f(t_k) \\ \vdots \\ \mathbf{u_{l}}_f(t_k) \end{bmatrix}, \mathbf{y}_f(t_k) = \begin{bmatrix} y_f(t_{k-1}) \\ \vdots \\ y_f(t_{k-n_a}) \end{bmatrix}, \mathbf{u_{i}}_f(t_k) = \begin{bmatrix} u_{if}(t_k) \\ \vdots \\ u_{if}(t_{k-n_b}) \end{bmatrix}, \tag{2.30}$$

and $\tilde{v}_{\mathrm{f}}(t_k) = Q(q,\theta)\tilde{v}(t_k)$. Consequently when the data is generated from $\mathcal{S}_{\mathrm{o}}$, if $\theta = \theta_{\mathrm{o}}$, then $\tilde{v}_{\mathrm{f}}(t_k) = e_{\mathrm{o}}(t_k)$ which is a white noise.

Therefore, according to the conditions for optimal IV estimates (see C4 in Section 1.5.5), the optimal filtered instrument for the augmented LTI MISO model structure (2.22) depicted in Figure 2.2 is given as :

$$\begin{aligned}\zeta^{\mathrm{opt}}(t_k) = \Big[& -\chi_{\mathrm{o}}(t_{k-1}) \quad \dots \quad -\chi_{\mathrm{o}}(t_{k-n_{\mathrm{a}}}) \quad u_1(t_k) \quad \dots \quad u_1(t_{k-n_b}) \\ & \dots \quad u_l(t_k) \quad \dots \quad u_l(t_{k-n_b}) \Big]^\top,\end{aligned} \tag{2.31}$$

while the optimal filter is given as :

$$L^{\mathrm{opt}}(q) = Q(q,\theta_{\mathrm{o}}) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})A_{\mathrm{o}}(q^{-1})}. \tag{2.32}$$

2.2.5 The Hammerstein RIV (HRIV) algorithm for BJ models

Of course none of $A(q^{-1},\rho_{\mathrm{o}})$, $B_i(q^{-1},\rho_{\mathrm{o}})$, $C(q^{-1},\eta_{\mathrm{o}})$ or $D(q^{-1},\eta_{\mathrm{o}})$ is known and only their estimates are available. Therefore, neither the optimal prefilter nor the optimal instrument can be accessed and they can only be estimated. In the present Hammerstein case as in the LTI case presented in 1.5.5, the 'auxiliary model' used to generate the noise-free output as well as the computation of the associated prefilter (2.28), are updated based on the parameter estimates obtained at the previous iteration to overcome this problem.

Algorithm 2 (HRIV)

Step 1 *Generate an initial estimate of the process model parameter* $\hat{\rho}^{(0)}$ *(*e.g. *using the LS method). Set* $C(q^{-1},\hat{\eta}^{(0)}) = D(q^{-1},\hat{\eta}^{(0)}) = 1$. *Set* $\tau = 0$.

Step 2 *Compute an estimate of* $\chi(t_k)$ *via*

$$\hat{\chi}(t_k) = \frac{\sum_{i=1}^{l} B_i(q^{-1},\hat{\rho}^{(\tau)})u_i(t_k)}{A(q^{-1},\hat{\rho}^{(\tau)})},$$

where $\hat{\rho}^{(\tau)}$ *is the estimate obtained at the previous iteration. According to assumption HA4 each* $\hat{\chi}$ *is bounded (if* $A(q^{-1},\hat{\rho}^{(\tau)})$ *is not stable it has to be stabilized).*

Step 3 *Compute the filter as in* (2.28) *:*

$$L(q,\hat{\theta}^{(\tau)}) = \frac{D(q^{-1},\hat{\eta}^{(\tau)})}{C(q^{-1},\hat{\eta}^{(\tau)})A(q^{-1},\hat{\rho}^{(\tau)})}$$

and the associated filtered signals $\{u_{i\mathrm{f}} = \gamma_i(u)_{\mathrm{f}}\}_{i=1}^{l}$, y_{f} *and* $\{\hat{\chi}_{\mathrm{f}}\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$.

Step 4 *Build the filtered regressor* $\varphi_{\mathrm{f}}(t_k)$ *and in terms of C4 (Section 1.5.5) the filtered instrument* $\hat{\zeta}_{\mathrm{f}}(t_k)$ *which equal in the given context :*

$$\begin{aligned}\varphi_{\mathrm{f}}(t_k) = \Big[& -y_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -y_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad u_{1\mathrm{f}}(t_k) \quad \dots \quad u_{1\mathrm{f}}(t_{k-n_b}) \\ & \dots \quad u_{l\mathrm{f}}(t_k) \quad \dots \quad u_{l\mathrm{f}}(t_{k-n_b}) \Big]^\top \\ \hat{\zeta}_{\mathrm{f}}(t_k) = \Big[& -\hat{\chi}_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -\hat{\chi}_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad u_{1\mathrm{f}}(t_k) \quad \dots \quad u_{1\mathrm{f}}(t_{k-n_b}) \\ & \dots \quad u_{l\mathrm{f}}(t_k) \quad \dots \quad u_{l\mathrm{f}}(t_{k-n_b}) \Big]^\top\end{aligned} \tag{2.33}$$

Step 5 *The IV optimization problem can be stated in the form*

$$\hat{\rho}^{(\tau+1)}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k)\right]\rho - \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)y_{\mathrm{f}}(t_k)\right]\right\|^2 \tag{2.34}$$

where the solution is obtained as

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k)\right]^{-1}\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)y_{\mathrm{f}}(t_k).$$

The resulting $\hat{\rho}^{(\tau+1)}(N)$ is the IV estimate of the process model associated parameter vector at iteration $\tau+1$ based on the prefiltered input/output data.

Step 6 *An estimate of the noise signal v is obtained as*

$$\hat{v}(t_k) = y(t_k) - \hat{\chi}(t_k, \hat{\rho}^{(\tau)}). \tag{2.35}$$

Based on $\hat{v}$, the estimation of the noise model parameter vector $\hat{\eta}^{(\tau+1)}$ follows, using in this case the `ARMA` *estimation algorithm of the* `MATLAB` *identification toolbox (an IV approach can also be used for this purpose, see [You08]).*

Step 7 *If $\theta^{(\tau+1)}$ has converged or the maximum number of iterations is reached, then stop, else increase τ by 1 and go to Step 2.*

At the end of the iterative process, coefficients $\hat{\alpha}_i$ are not directly accessible. They are however deduced from polynomial $\hat{B}_i(q^{-1})$ as $B_i(q^{-1},\rho) = \alpha_i B(q^{-1},\rho)$. The hypothesis $\alpha_1 = 1$ guarantees that $B_1(q^{-1},\rho) = B(q^{-1},\rho)$ and $\hat{\alpha}_i$ can be computed from :

$$\hat{\alpha}_i = \frac{1}{n_b+1}\sum_{j=0}^{n_b}\frac{\hat{b}_{i,j}}{\hat{b}_{1,j}}, \tag{2.36}$$

where $\hat{b}_{i,j}$ is the j^{th} coefficient of polynomial term $B_i(q^{-1},\rho)$ for $i = 2\dots l$. It must be pointed out that the separation of α_i and b_i coefficients can be executed using more powerful methods such as *single value decomposition* for example.
Moreover, after the convergence is complete, it is possible to compute the estimated parametric error covariance matrix $\hat{\mathbf{P}}_\rho$ from the expression (see Section 1.5.4) :

$$\hat{\mathbf{P}}_\rho = \hat{\sigma}_e^2\left(\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)\hat{\zeta}_{\mathrm{f}}^{\top}(t_k)\right)^{-1} \tag{2.37}$$

where $\hat{\zeta}$ is the IV vector obtained at convergence and $\hat{\sigma}_e^2$ is the estimated residual variance.

Comments

By using the described algorithm, if convergence occurs, it can be seen that all conditions C1-C6 from Section 1.5.5 except for C5 are respected (a discussion about how C1 is fulfilled can be found in [SS83]) : the `HRIV` estimates might be statistically optimal for the augmented model proposed, but the minimal number of parameters needed for representing the MISO structure and the Hammerstein structure are not equal. Consequently, the HRIV estimates cannot be statistically optimal for the Hammerstein model structure. Nonetheless, even if not optimal, the performance of the `HRIV` estimates are depicted on relevant examples in the result section 2.2.7.

2.2.6 HSRIV algorithm for OE models

A simplified version of HRIV algorithm named HSRIV follows the exact same theory for estimation of Hammerstein output error models. It is mathematically described by, $C(q^{-1},\eta^j) = C_o(q^{-1}) = 1$ and $D(q^{-1},\eta^j) = D_o(q^{-1}) = 1$. All previous given equations remain true, and it suffices to estimate ρ^j as $\theta^j = \rho^j$. The implementation of HSRIV is much simpler than HRIV as there is no noise model estimation in the algorithm.

2.2.7 Performance evaluation of the proposed HRIV and HSRIV algorithms

This section presents numerical evaluation of both suggested HRIV and HSRIV methods. For the presented example, the nonlinear block has a polynomial form, *i.e.* $\gamma_i(u(t_k)) = u^i(t_k), \forall i$ and the system to identify is given by

$$\mathcal{S}_{\text{o}} \begin{cases} \bar{u}(t_k) = u(t_k) + 0.5u^2(t_k) + 0.25u^3(t_k), \\ G_o(q) = \dfrac{0.5q^{-1} + 0.2q^{-2}}{1 + q^{-1} + 0.5q^{-2}}, \\ H_o(q) = \dfrac{1}{1 - q^{-1} + 0.2q^{-2}}. \end{cases}$$

where $u(t_k)$ follows a uniform distribution with values between -2 and 2.
The models considered for estimation are :

$$\mathcal{M}_{\texttt{HRIV}} \begin{cases} G(q,\rho) = \dfrac{b_1q^{-1} + b_2q^{-2}}{1 + a_1q^{-1} + a_2q^{-2}}, \\ H(q,\eta) = \dfrac{1}{1 + d_1q^{-1} + d_2q^{-1}}, \\ f(u(t_k)) = u(t_k) + \alpha_1u^2(t_k) + \alpha_2u^3(t_k) \end{cases} \tag{2.38}$$

for the HRIV method which fulfills [HA1] and

$$\mathcal{M}_{\texttt{HSRIV}} \begin{cases} G(q,\rho) = \dfrac{b_1q^{-1} + b_2q^{-2}}{1 + a_1q^{-1} + a_2q^{-2}}, \\ H(q,\eta) = 1, \\ f(u(t_k)) = u(t_k) + \alpha_1u^2(t_k) + \alpha_2u^3(t_k) \end{cases} \tag{2.39}$$

for the HSRIV method which only fulfills $G_{\text{o}} \in \mathcal{G}$ ($H_{\text{o}} \notin \mathcal{H}$).
The result of a Monte Carlo simulation (MCs) analysis is shown in Table 2.1 and the algorithms considered are : HRIV, HSRIV and LSQNONLIN. The LSQNONLIN is a nonlinear optimization algorithm from the MATLAB® optimization toolbox. It assumes the same model as the HRIV method ($S_{\text{o}} \in \mathcal{M}$) and hands out the statistically optimal estimates if the method is properly initialized. In order to place the LSQNONLIN method at its advantage, it is initialized with the true parameter values and therefore this method can be considered as the ground truth.
The MCs results are based on $N_{\text{run}} = 100$ random realization, with the Gaussian white noise input to the ARMA noise model being selected randomly for each realization. In order to compare the statistical performance of the different approaches, the computed mean and standard deviation of the estimated parameters are presented. The noise added at the output is adjusted such that it

TAB. 2.1 – Estimation results of the proposed algorithm

		b_0	b_1	a_1	a_2	α_1	α_2	d_1	d_2
method	true value	0.5	0.2	1	0.5	0.5	0.25	-1	0.2
`LSQNONLIN`	$mean(\hat{\theta})$	0.4991	0.1983	0.9984	0.4992	0.5011	0.2512	-1.0004	0.2001
	$std(\hat{\theta})$	0.0159	0.0109	0.0114	0.0059	0.0187	0.0194	0.0224	0.0219
`HSRIV`	$mean(\hat{\theta})$	0.4992	0.1975	0.9944	0.4976	0.4956	0.2657	X	X
	$std(\hat{\theta})$	0.0402	0.0471	0.0186	0.0071	0.1107	0.0999	X	X
`HRIV`	$mean(\hat{\theta})$	0.5004	0.2009	0.9984	0.4992	0.5006	0.2487	-1.0011	0.2007
	$std(\hat{\theta})$	0.0193	0.0208	0.0114	0.0059	0.0384	0.0397	0.0224	0.0220

corresponds to a *Signal-to-Noise-Ratio* (SNR) of 5dB using :

$$\text{SNR} = 10\log\left(\frac{P_\chi}{P_{v_o}}\right), \tag{2.40}$$

where P_g is the average power of signal g. The number of samples is chosen as $N = 2000$.
As expected, Table 2.1 shows that the proposed algorithms produce unbiased estimates of the Hammerstein model parameters. It can be further noticed that the standard deviation of the estimates remains low even under the unrealistic noise level of 5dB. Even though, the ratio between the `HSRIV` and `HRIV` estimate standard deviation equals to 2. This can be logically explained by the fact that the `HSRIV` algorithm assumes a wrong noise model and such result remains acceptable in practical applications. Finally it can be depicted that the `HRIV` provides the statistical optimal estimates for the parameters which are not replicated inside the parameter vector ρ, that is a_1, a_2, d_1 and d_2. Concerning the other coefficients the standard deviation is approximately multiplied by 2 but the absolute value remains acceptable considering the level of noise added. It can be concluded that the presented algorithms, even if not optimal in the Hammerstein case, constitute good candidates for practical applications where the noise is unknown, and can be a strong help for initializing optimal methods such as `LSQNONLIN`.

2.3 Discrete-time Hammerstein model identification in closed loop

Closed-loop systems are widely used in practice and various attempts have been made to handle linear system identification in the presence of feedback. Indeed, closed-loop system identification leads to several difficulties due to the correlation between the disturbances and the control signal induced by the loop. Several methods have therefore been developed to cope with this problem see *e.g.* [DL80], [Gev93], [Gev05], [SS89], [VdH98], [FL99], [Zhe03]. The problem of nonlinear plants operating in closed loop has not received a lot of attention so far. Indeed, this problem is handled only in a few papers as in *e.g* [FAD01], [LAD01], [DAL02], and [WJC07].
The closed-loop model identification problem is not directly addressed in Section 1.2. Nonetheless, the field of identification is very important for plants operating in closed loop and especially for the

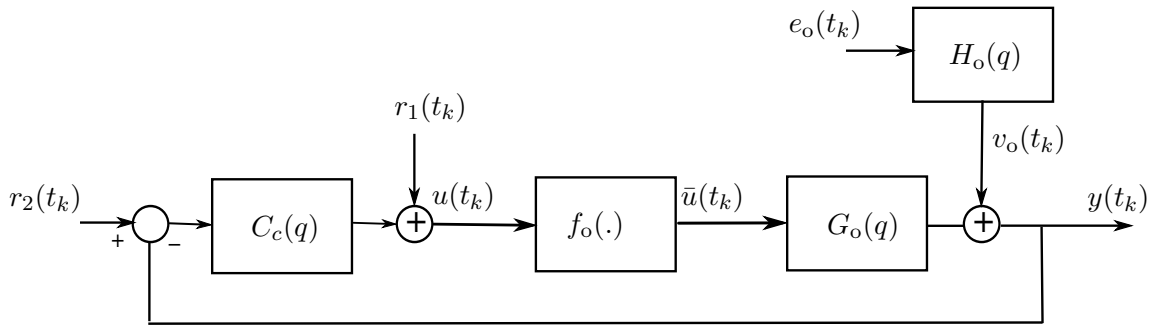


FIG. 2.3 – Closed-loop Hammerstein system

one where the loop cannot be unplugged. Therefore, this section presents how the RIV method can be extended to Hammerstein systems operating in closed loop.

2.3.1 System description

Consider the stable, SISO closed-loop data-generating nonlinear system represented in Figure 2.3 and assume that both control input and output signals, $u(t_k)$ and $y(t_k)$ are uniformly sampled at a constant sampling time $\mathtt{T_s}$ over N samples.
The system $\mathcal{S}_o$ to be identified is described by the following input-output relationships :

$$\mathcal{S}_o \begin{cases} y(t_k) = G_o(q)\bar{u}(t_k) + H_o(q)e_o(t_k), \\ \bar{u}(t_k) = f_o(u(t_k)), \\ u(t_k) = r(t_k) - C_c(q)y(t_k), \\ r(t_k) = r_1(t_k) + C_c(q)r_2(t_k), \end{cases} \tag{2.41}$$

where $G_o(q)$ denotes the process and $C_c(q)$ the controller. $u(t_k)$ describes the process input signal, $y(t_k)$ the process output signal. $r_1(t_k)$ can be a reference value, a setpoint or a noise disturbance on the regulator output and $r_2(t_k)$ can be a setpoint or a measurement noise on the output signal. It can be further noticed that both $r_1(t_k)$ and $r_2(t_k)$ are considered for theoretical developments. Nonetheless in practice, either one or the other is excited while the other remains null. $B_o(q^{-1})$ and $A_o(q^{-1})$ are polynomials in the shift operator q^{-1} with respective degree n_b and n_a.
Again, the method presented is based on the identification of a Box–Jenkins model, where the linear and the noise models are not constrained to have common poles. The coloured noise associated with the sampled output measurement $y(t_k)$ has rational spectral density and can be represented by a discrete-time autoregressive moving average ARMA model :

$$v_o(t_k) = H_o(q)e(t_k) = \frac{C_o(q^{-1})}{D_o(q^{-1})}e_o(t_k) \tag{2.42}$$

where $C_o(q^{-1})$ and $D_o(q^{-1})$ are monic polynomials with constant coefficients and with respective degree n_c and n_d. Furthermore, all roots of $z^{n_d}D_o(z^{-1})$ and $z^{n_c}C_o(z^{-1})$ are inside the unit disc. It can be noticed that in case $C_o(q^{-1}) = D_o(q^{-1}) = 1$, (2.4) defines an OE noise model.
The approach considered here is a direct identification method (using signals $u(t_k)$ and $y(t_k)$) assuming that :

– The controller $C_c(q)$ is *a priori* known.
– The controller $C_c(q)$ insures the global BIBO stability of the closed-loop system $\mathcal{S}_o$.
Therefore, the problem faced here is very similar to the open-loop case under the additive constraint that $u(t_k)$ is correlated to $y(t_k)$ and therefore to $v_o(t_k)$.

2.3.2 Model considered

The discussion about the use of an augmented model as driven in Section 2.2.2 leads to the exact same conclusions in the closed-loop case. Consequently the same model will be used in which the closed-loop equation is added.
The nonlinearity is again considered as a sum of basis functions such as given in (2.9) under the exact same conditions as exposed in Section 2.2.2.2. Consequently, with respect to the given process and noise part $(\mathcal{G}_\rho, \mathcal{H}_\eta)$ such as defined in Section 2.2.2, the parameters can be collected as $\theta = [\ \rho^\top \quad \eta^\top\ ]$ and the signal relations of the Hammerstein BJ model operating in closed loop, denoted in the sequel as $\mathcal{M}_\theta$, are defined as :

$$\mathcal{M}_\theta : \begin{cases} y(t_k) = \dfrac{\sum_{i=1}^{l} B_i(q^{-1},\rho)u_i(t_k)}{A(q^{-1},\rho)} + \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)}e(t_k), \\ u(t_k) = r(t_k) - C_c(q)y(t_k), \\ r(t_k) = r_1(t_k) + C_c(q)r_2(t_k), \end{cases} \tag{2.43}$$

with $B_i(q^{-1},\rho) = \alpha_i B(q^{-1},\rho)$ and $u_i(t_k) = \gamma_i(u(t_k))$.
Based on this model structure, the whole model set including the process $(\mathcal{G}_\rho)$ and noise $(\mathcal{H}_\eta)$ models parameterized independently, is denoted as $\mathcal{M}$ and takes finally the form

$$\mathcal{M} = \left\{ (\mathcal{G}_\rho, \mathcal{H}_\eta) \ \mid \ \mathrm{col}(\rho,\eta) = \theta \in \mathbb{R}^{n_\rho + n_\eta} \right\}. \tag{2.44}$$

The set (2.44) corresponds to the set of candidate models in which we seek the best fitting model using data gathered from $\mathcal{S}_o$ under a given identification criterion (cost function). Again, and for the same reasons as in the open-loop case, the PEM framework from [Lju99] as described in Section 1.4 can be directly used here.

2.3.3 Identification problem statement

Therefore, based on the previous considerations, the identification problem addressed can now be defined.

Problem 3 *Given a discrete-time closed loop Hammerstein data generating system $\mathcal{S}_o$ defined as in (2.41) and a data set $\mathcal{D}_N$ collected from $\mathcal{S}_o$ and based on the Hammerstein BJ model structure $\mathcal{M}_\theta$ defined by (2.22), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions :*

HCLA1 *$\mathcal{S}_o \in \mathcal{M}$, i.e. there exits a θ_o defining a $\mathcal{G}_{\rho_o} \in \mathcal{G}$ and a $\mathcal{H}_{\eta_o} \in \mathcal{H}$ such that $(\mathcal{G}_{\rho_o}, \mathcal{H}_{\eta_o})$ is equal to $\mathcal{S}_o$.*

HCLA2 *$r_1(t_k)$ and $r_2(t_k)$ are not correlated to $e_o(t_k)$.*

HCLA3 *$\mathcal{D}_N$ is informative with respect to $\mathcal{M}$.*

HCLA4 *The open-loop process G_o is BIBO stable,* i.e *for any bounded input signals r_1 and r_2, the output of $\mathcal{S}_o$ is bounded.*

HCLA5 *The controller C_c is supposed to be* a priori *known and insures the BIBO stability of the whole system.*

2.3.4 Optimal IV for CL Hammerstein models

If the process to be identified belongs to the model set previously defined, then $y(t_k)$ can be written under the same filtered regression form as in the open-loop case (2.29) :

$$y_f(t_k) = \varphi_f^\top(t_k)\rho + \tilde{v}_f(t_k) \tag{2.45}$$

where, φ is as given in (2.30) and ρ as in (2.21) and the filter used is $Q(q,\theta)$ from (2.28). Under these conditions again, when the data is generated from $\mathcal{S}_o$ and if $\theta = \theta_o$ then $\tilde{v}_f(t_k) = e_o(t_k)$ which is a white noise.

Nonetheless, even if the description of the problem is in many points similar to the open-loop case, estimating the process between u and y using an IV method is a little complicated by the fact that the feedback loop carries noisy information. Therefore, using the optimal instrument such as given in (2.31) will not lead to fullfil C2 (see Section 1.5.5) as the signal u is corrupted by noise. Therefore, not only the noise-free output signal has to be reconstructed but also the noise-free input signal. Consequently, by extending the RIV method presented in Section 2.2.4 to systems operating in closed-loop, the optimal instrument for the augmented model is :

$$\zeta^{\text{opt}}(t_k) = \begin{bmatrix} -\mathring{\mathbf{X}}(t_k) \\ \gamma_1(\mathring{\mathbf{u}}(t_k)) \\ \vdots \\ \gamma_l(\mathring{\mathbf{u}}(t_k)) \end{bmatrix}, \mathring{\mathbf{X}}(t_k) = \begin{bmatrix} \mathring{\chi}(t_{k-1}) \\ \mathring{\chi}(t_{k-2}) \\ \vdots \\ \mathring{\chi}(t_{k-n_a}) \end{bmatrix}, \mathring{\mathbf{u}}(t_k) = \begin{bmatrix} \mathring{u}(t_{k-1}) \\ \mathring{u}(t_{k-2}) \\ \vdots \\ \mathring{u}(t_{k-n_b}) \end{bmatrix}, \tag{2.46}$$

where $\mathring{u}$ and $\mathring{\chi}$ are the process input and output respectively of the noise-free system $\mathring{\mathcal{S}}_o$ as depicted in Figure 2.4.

For the reader's ease, some clarification about the different notations $\mathring{\chi}(t_k)$ and $\chi_o(t_k)$ has to be given. In this closed-loop framework, $\chi_o(t_k)$ refers to the output of the process linked to the system $\mathcal{S}_o$ depicted in Figure 2.3 and is therefore corrupted by noise through the feedback loop. By opposition $\mathring{\chi}(t_k)$ refers to the output of the noise-free system $\mathring{\mathcal{S}}_o$ depicted in Figure 2.4 and described by :

$$\mathring{\mathcal{S}}_o \begin{cases} \mathring{\chi}(t_k) = G_o(q)\bar{\mathring{u}}(t_k), \\ \bar{\mathring{u}}(t_k) = f_o(\mathring{u}(t_k)), \\ \mathring{u}(t_k) = r(t_k) - C_c(q)\mathring{\chi}(t_k), \\ r(t_k) = r_1(t_k) + C_c(q)r_2(t_k). \end{cases} \tag{2.47}$$

In other word, $\mathring{\mathcal{S}}_o$ is equivalent to $\mathcal{S}_o$ for $e_o(t_k) = 0, \forall k$.

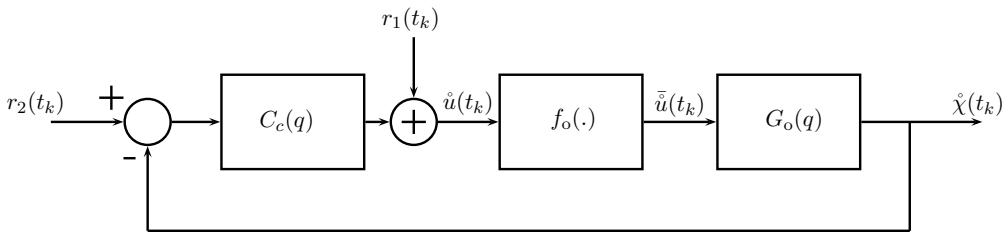


FIG. 2.4 – Noise-free system

2.3.5 Hammerstein closed-loop RIV (HCLRIV) algorithm for BJ models

Using this given instrument, the RIV algorithm can be applied to Hammerstein systems operating in closed loop.

Algorithm 3 (HCLRIV)

Step 1 *Generate an initial estimate of the process model parameter* $\hat{\rho}^{(0)}$ *(*e.g. *using the LS method). Set* $C(q^{-1}, \hat{\eta}^{(0)}) = D(q^{-1}, \hat{\eta}^{(0)}) = 1$. *Set* $\tau = 0$.

Step 2 *Compute an estimate of* $\mathring{\chi}(t_k)$ *and* $\mathring{u}(t_k)$ *via the simulation of the auxiliary model* $\mathring{\mathcal{S}}(\hat{\rho}^{\tau})$ *given in Figure 2.5 :*

$$\begin{cases} \hat{\mathring{\chi}}(t_k) = \dfrac{\sum_{i=1}^{l} B_i(q^{-1}, \hat{\rho}^{(\tau)}) \gamma_i(\hat{\mathring{u}}(t_k))}{A(q^{-1}, \hat{\rho}^{(\tau)})} \\ \hat{\mathring{u}}(t_k) = r(t_k) - C_c(q)\hat{\mathring{\chi}}(t_k), \end{cases} \tag{2.48}$$

where $\hat{\rho}^{(\tau)}$ *is the estimate obtained at the previous iteration. According to assumption HCLA4 each* $\hat{\mathring{\chi}}$ *is bounded.*

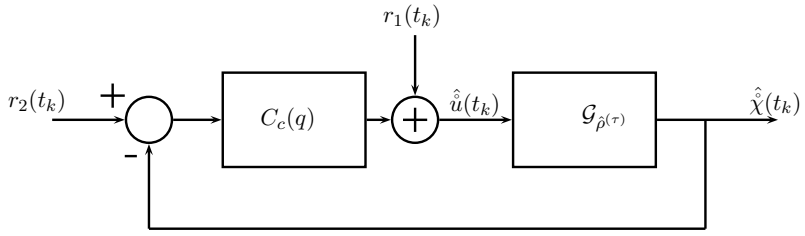


FIG. 2.5 – Auxiliary model used for simulation

Step 3 *Compute the filter as in* (2.28) *:*

$$L(q, \hat{\theta}^{(\tau)}) = \frac{D(q^{-1}, \hat{\eta}^{(\tau)})}{C(q^{-1}, \hat{\eta}^{(\tau)}) A(q^{-1}, \hat{\rho}^{(\tau)})}$$

and the associated filtered signals $\{u_{i\mathrm{f}} = (\gamma_i(u))_{\mathrm{f}}\}_{i=1}^{l}$, y_{f} *and* $\{\hat{\chi}_{\mathrm{f}}\}_{i=1,l=0}^{n_{\mathrm{a}}, n_\alpha}$.

Step 4 *Build the filtered regressor* $\varphi_{\mathrm{f}}(t_k)$ *and in terms of C4 (Section 1.5.5) the filtered instrument*

$\hat{\zeta}_{\mathrm{f}}(t_k)$ *which equal in the given context :*

$$\begin{aligned}\varphi_{\mathrm{f}}(t_k) = \Big[& -y_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -y_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad u_{1\mathrm{f}}(t_k) \quad \dots \quad u_{1\mathrm{f}}(t_{k-n_b}) \\ & \dots \quad u_{l\mathrm{f}}(t_k) \quad \dots \quad u_{l\mathrm{f}}(t_{k-n_b}) \Big]^\top \\ \hat{\zeta}_{\mathrm{f}}(t_k) = \Big[& -\hat{\hat{\chi}}_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -\hat{\hat{\chi}}_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad \hat{\hat{u}}_{1\mathrm{f}}(t_k) \quad \dots \quad \hat{\hat{u}}_{1\mathrm{f}}(t_{k-n_b}) \\ & \dots \quad \hat{\hat{u}}_{l\mathrm{f}}(t_k) \quad \dots \quad \hat{\hat{u}}_{l\mathrm{f}}(t_{k-n_b}) \Big]^\top \end{aligned} \tag{2.49}$$

Step 5 *The IV optimization problem can be stated in the form*

$$\hat{\rho}^{(\tau+1)}(N) = \arg \min_{\rho \in \mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) \varphi_{\mathrm{f}}^\top(t_k) \right] \rho - \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}(t_k) \right] \right\|^2 \tag{2.50}$$

where the solution is obtained as

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) \varphi_{\mathrm{f}}^\top(t_k) \right]^{-1} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}(t_k).$$

The resulting $\hat{\rho}^{(\tau+1)}(N)$ *is the IV estimate of the process model associated parameter vector at iteration* $\tau + 1$ *based on the prefiltered input/output data.*

Step 6 *An estimate of the noise signal* v *is obtained as*

$$\hat{v}(t_k) = y(t_k) - \frac{\sum_{i=1}^{l} B_i(q^{-1}, \hat{\rho}^{(\tau)}) \gamma_i(u(t_k))}{A(q^{-1}, \hat{\rho}^{(\tau)})}. \tag{2.51}$$

Based on $\hat{v}$*, the estimation of the noise model parameter vector* $\hat{\eta}^{(\tau+1)}$ *follows, using in this case the* `ARMA` *estimation algorithm of the* `MATLAB` *identification toolbox (an IV approach can also be used for this purpose, see [You08]).*

Step 7 *If* $\theta^{(\tau+1)}$ *has converged or the maximum number of iterations is reached, then stop, else increase* τ *by* 1 *and go to Step 2.*

As in the open-loop case, at the end of the iterative process, coefficients $\hat{\alpha}_i$ are not directly accessible. They are however deduced from polynomial $\hat{B}_i(q^{-1})$ as $B_i(q^{-1}, \rho) = \alpha_i B(q^{-1}, \rho)$. The hypothesis $\alpha_1 = 1$ guarantees that $B_1(q^{-1}, \rho) = B(q^{-1}, \rho)$ and $\hat{\alpha}_i$ can be computed from :

$$\hat{\alpha}_i = \frac{1}{n_b + 1} \sum_{j=0}^{n_b} \frac{\hat{b}_{i,j}}{\hat{b}_{1,j}}, \tag{2.52}$$

where $\hat{b}_{i,j}$ is the j^{th} coefficient of polynomial term $B_i(q^{-1}, \rho)$ for $i = 2 \dots l$. It must be pointed out that the separation of α_i and b_i coefficients can be executed using more powerful methods such as *single value decomposition* for example.

Moreover, after the convergence is complete, it is possible to compute the estimated parametric error covariance matrix $\hat{\mathbf{P}}_\rho$ from the expression (see Section 1.5.4) :

$$\hat{\mathbf{P}}_\rho = \hat{\sigma}_e^2 \left(\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) \hat{\zeta}_{\mathrm{f}}^\top(t_k) \right)^{-1} \tag{2.53}$$

where $\hat{\zeta}$ is the IV vector obtained at convergence and $\hat{\sigma}_e^2$ is the estimated residual variance.

Comments

By using the described algorithm, if convergence occurs, it can be seen that all conditions C1-C6 from Section 1.5.5 except for C5 are respected (a discussion about how C1 is fulfilled can be found in [SS83]) : the `HCLRIV` estimates might be statistically optimal for the augmented model proposed, but the minimal number of parameters needed for representing the MISO structure and the Hammerstein structure are not equal. Consequently, the `HCLRIV` estimates cannot be statistically optimal for the Hammerstein model structure. Nonetheless, even if not optimal, the HCLRIV estimates are unbiased with a low variance as it will be seen in the result section 2.3.7.

2.3.6 `HCLSRIV` algorithm for OE models

A simplified version of `HCLRIV` algorithm named `HCLSRIV` follows the exact same theory for estimation of Hammerstein output error models. It is mathematically described by, $C(q^{-1},\eta^j) = C_o(q^{-1}) = 1$ and $D(q^{-1},\eta^j) = D_o(q^{-1}) = 1$. All previous given equations remain true, and it suffices to estimate ρ^j as $\theta^j = \rho^j$.

2.3.7 Performance evaluation of the proposed `HCLRIV` and `HCLSRIV` algorithms

2.3.7.1 Performance analysis of the closed-loop estimates

This section presents numerical illustration of this problem using both suggested `HCLRIV` and `HCLSRIV` methods on a closed-loop example. Considering the closed-loop case, the nonlinear function has to be bounded and therefore it is a fair assumption to consider the nonlinear block as a sum of sine functions *i.e.* $\gamma_i(u(t_k)) = \sin(i.u(t_k)), \forall i \in \mathbb{N}$ and the system considered here is :

$$\mathcal{S}_{\text{o}} \begin{cases} \bar{u}(t_k) = \sin(u(t_k)) - 0.5\sin(2u(t_k)) + 0.4\sin(3u(t_k)). \\ G_o(q) = \dfrac{0.0997q^{-1} - 0.0902q^{-2}}{1 - 1.8858q^{-1} + 0.9048q^{-2}}, \\ H_o(q) = \dfrac{1 + 0.5q^{-1}}{1 - 0.85q^{-1}}, \\ C_c(q) = \dfrac{10.75 - 9.25q^{-1}}{1 - q^{-1}}. \end{cases}$$

$r_1(t_k)$ follows a uniform distribution with values between -2 and 2 while $r_2(t_k) = 0$.
The models considered for estimation are :

$$\mathcal{M}_{\texttt{HCLRIV}} \begin{cases} G(q,\rho) = \dfrac{b_1q^{-1} + b_2q^{-2}}{1 + a_1q^{-1} + a_2q^{-2}}, \\ H(q,\eta) = \dfrac{1 + c_1q^{-1}}{1 + d_1q^{-1}}, \\ f(u(t_k)) = \sin(u(t_k)) + \displaystyle\sum_{i=2}^{3} \alpha_i \sin(i.u(t_k)) \end{cases} \tag{2.54}$$

for the `HCLRIV` method which fulfills [HCLA1] and

$$\mathcal{M}_{\texttt{HCLSRIV}} \begin{cases} G(q,\rho) = \dfrac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}}, \\ H(q,\eta) = 1, \\ f(u(t_k)) = \sin(u(t_k)) + \displaystyle\sum_{i=2}^{3} \alpha_i \sin(i.u(t_k)) \end{cases} \tag{2.55}$$

for the `HCLSRIV` method which only fulfills $G_{\mathrm{o}} \in \mathcal{G}$ ($H_{\mathrm{o}} \notin \mathcal{H}$).
For illustration purposes, the proposed methods are compared to the parameters of the linear plant process identified using a prediction error minimisation method (`PEMlin`, `OElin`) assuming that α_i are known. The problem reduces therefore to the identification of a linear model in closed loop where the input is $\bar{u}(t_k)$ and the output is $y(t_k)$.
The model considered for the PEM identification using the `PEM` method from the MATLAB identification toolbox is then :

$$\mathcal{M}_{\texttt{PEMlin}} \begin{cases} G(q,\rho) = \dfrac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}}, \\ H(q,\eta) = \dfrac{1 + c_1 q^{-1}}{1 + d_1 q^{-1}}, \\ f(u(t_k)) = \sin(u(t_k)) - 0.5\sin(2u(t_k)) + 0.4\sin(3u(t_k)) \end{cases} \tag{2.56}$$

for the `PEMlin` method for which $S_{\mathrm{o}} \in \mathcal{M}$.
The model considered for the PEM identification using the `OE` method from the MATLAB identification toolbox is then :

$$\mathcal{M}_{\texttt{OElin}} \begin{cases} G(q,\rho) = \dfrac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}}, \\ H(q,\eta) = 1, \\ f(u(t_k)) = \sin(u(t_k)) - 0.5\sin(2u(t_k)) + 0.4\sin(3u(t_k)), \end{cases} \tag{2.57}$$

for the `OElin` method for which $G_{\mathrm{o}} \in \mathcal{G}$ and $H_{\mathrm{o}} \notin \mathcal{H}$.
These linear methods are chosen as their properties are well-known [Lju99].
The result of a Monte Carlo simulation (MCs) analysis is shown in Table 2.2 for the algorithms considered, where the consequences of the modelling error on the noise is displayed.
The MCs results are based on $N_{\mathrm{run}} = 100$ random realization, with the Gaussian white noise input to the ARMA noise model being selected randomly for each realization. In order to compare the statistical performances of the different approaches, Table 2.2 displays the computed mean, the standard deviation and the normalised root mean square error (RMSE) for each estimated parameters. The RMSE is defined as

$$RMSE(\hat{\theta}_j) = \sqrt{\frac{1}{N_{\mathrm{run}}} \sum_{i=1}^{N_{\mathrm{run}}} \left(\frac{\theta_j^o - \hat{\theta}_j(i)}{\theta_j^o} \right)^2}, \tag{2.58}$$

with $\hat{\theta}_j$ the jth component of the parameter vector θ.

TAB. 2.2 – Closed-loop estimation results

method		b_1	b_2	a_1	a_2	α_1	α_2	c_1	d_1
	true value	0.0997	-0.0902	-1.8858	0.9048	-0.5000	0.4000	0.5000	-0.8500
OElin	$mean(\hat{\theta})$	0.0693	-0.0653	-1.8655	0.8867	X	X	X	X
	$std(\hat{\theta})$	0.0016	0.0019	0.0186	0.0171	X	X	X	X
	RMSE	0.2959	0.2686	0.0141	0.0267	X	X	X	X
PEMlin	$mean(\hat{\theta})$	0.0998	-0.0901	-1.8857	0.9028	X	X	0.4918	-0.8603
	$std(\hat{\theta})$	0.0009	0.0025	0.0247	0.0208	X	X	0.0208	0.0312
	RMSE	0.0089	0.0280	0.0131	0.0229	X	X	0.0421	0.0365
HCLSRIV	$mean(\hat{\theta})$	0.0978	-0.0894	-1.8715	0.8927	-0.4893	0.4006	X	X
	$std(\hat{\theta})$	0.0100	0.0138	0.0943	0.0912	0.0774	0.0495	X	X
	RMSE	0.1010	0.1588	0.0520	0.1042	0.1554	0.1230	X	X
HCLRIV	$mean(\hat{\theta})$	0.1004	-0.0895	-1.8862	0.8923	-0.5024	0.4015	0.4959	-0.8515
	$std(\hat{\theta})$	0.0031	0.0090	0.0829	0.0799	0.0182	0.0153	0.0249	0.0310
	RMSE	0.0313	0.1035	0.0457	0.0915	0.0367	0.0381	0.0496	0.0365

The noise added at the output is adjusted such that it corresponds to a *Signal-to-Noise-Ratio* (SNR) of 10dB using in this closed-loop context :

$$\text{SNR} = 10\log\left(\frac{P_{\mathring{\chi}}}{P_{v_o}}\right), \tag{2.59}$$

where P_g is the average power of signal g. The number of samples is chosen as $N = 2000$.
On the one hand it can be seen in Table 2.2 that even if the noise model assumption is false, which is the case for the `HCLSRIV` and `OElin` methods in this example, the refined IV method gives a parameter estimation centered on the true value with a correct estimated parameter variance (RMSE remaining under 16%), while the prediction error based method `OElin` estimates are strongly biased and have an RMSE near 30%. This fact makes `HCLSRIV` a very interesting method for practical applications where the noise dynamic is not known : by modelling a white noise added to the output, the number of parameters to be estimated is reduced in comparison to a Box–Jenkins model identification.
On the other hand, the presented `HCLRIV` method is not optimal due to the redundancy in the regressor but exposes RMSE in estimated parameters under 11 %. It can be noticed however that the estimated parameter variances are close to those from `PEMlin` estimates : this may therefore be seen as an acceptable achievement since in the `PEMlin` case, the nonlinear part is supposed to be known.

2.3.7.2 Robustness issues with respect to nonlinearity error modelling

In most cases, the expert knowledge can give a hint about the structure of the linear part of the Hammerstein model. However, the nonlinearity model chosen often reflects the absence of knowledge

concerning the true nonlinearity. The linear-in-the-parameter model presented here therefore might not fulfill [HCLA1] in most cases. This problem is referred to as blind estimation. This section presents the performance of `HCLSRIV` on blind estimation of the nonlinear plant operating in close-loop. The studied system used is described by :

$$\mathcal{S}_{\text{o}} \begin{cases} G_o(q) = \frac{0.0997q^{-1}-0.0902q^{-2}}{1-1.8858q^{-1}+0.9048q^{-2}}, \\ H_o(q) = 1, \\ C_c(q) = \frac{10.75-9.25q^{-1}}{1-q^{-1}}. \end{cases} \tag{2.60}$$

In this example, we limit the application to the identification of odd nonlinear functions ($f(-x) = -f(x)$) which are supposed to be continuous in the studied input domain. The nonlinear function used for this example is the following piecewise linear function defined as :

$$f_{\text{o}}(x) = \begin{cases} -1 & \text{if } x < -3 \\ x/3 & \text{if } -3 < x \leq -1.5 \\ -0.5 & \text{if } -1.5 < x \leq -1 \\ x + 0.5 & \text{if } -1.5 < x \leq -0.5 \\ 0 & \text{if } -0.5 < x \leq 0 \\ f(-x) = -f(x) & \forall x \end{cases}$$

This piecewise nonlinear function cannot be exactly modeled using the proposed model as it cannot be exactly expressed as the sum of basis functions.
Let us simply assume that the nonlinear function we look for is continuous for all given u and it is odd. Furthermore the nonlinear function must be bounded as being in a closed-loop. For odd bounded functions, the Fourier serie decomposition is limited to a sum of sine functions. It can be considered that the nonlinear function studied in the range $[\min_k(u(t_k)) \max_k(u(t_k))]$ is a single period of the corresponding periodic signal of period T with $T = 2 \max_k(|u(t_k)|)$. The model considered for estimation is then :

$$\mathcal{M}_{\text{Blind}} \begin{cases} G(q,\rho) = \dfrac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}}, \\ H(q,\eta) = 1, \\ f(u(t_k)) = \sin(\frac{2\pi}{T} u(t_k)) + \sum_{i=2}^{11} \alpha_i \sin(i \frac{2\pi}{T} u(t_k)) \end{cases} \tag{2.61}$$

It can be noticed that the constraint $\alpha_1 = 1$ is not fulfilled in this case. Therefore, it is not possible to determine directly α_i and $B(q)$ from the estimated $B_i(q)$ because there is no unicity of the model. Therefore, for display purposes, the obtained results were further homogenized based on b_1 using a correction factor λ :

$$\begin{cases} \lambda = \dfrac{0.0997}{b_1}, \\ b_i = b_i \lambda \quad i = 1 \dots n_b, \\ f(u(t_k)) = \dfrac{f(u(t_k))}{\lambda}. \end{cases}$$

The result of a Monte Carlo simulation analysis is shown in Table 2.3 and in Figure 2.6 using $N_{\text{run}} = 100$ and under a SNR of 15dB (see (2.59)).

It can be firstly noticed from Figure 2.6 that even though the periodic assumption is wrong, the algorithm can model relatively well the nonlinearity. Naturally, the nonlinear estimation appears to be more valid in the center of the $u(t_k)$ domain, where the periodicity approximation is not crucial while the error is growing when approaching the border of the domain. Nonetheless, some further knowledge such as the fact that f_o is bounded would be needed to extrapolate the nonlinear function outside the domain.

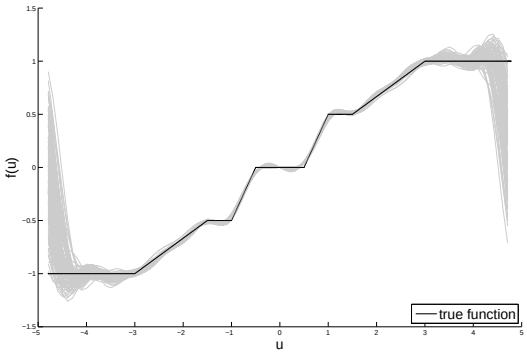


FIG. 2.6 – Nonlinear blind estimation results

Considering the linear part estimation shown in Table 2.3, it has to be noticed that b_1 always equals to 0.0997 due to homogenization factor and that the presented variance on this particular parameter is not representative. It can be further depicted that even if the nonlinear function is not correctly parameterized and can only be approximated, the linear part parameters of the plant process are estimated with a low variance (RMSE under 1%) and are unbiased.

Such a blind estimation is an essential step in practical applications where the nonlinear part is unknown : it firstly delivers a good estimate of the denominator dynamic of the linear process and further provides a shape for the nonlinear function which can be subsequently analyzed to create a better parametrization for this function and therefore find a more appropriate method for the identification of the studied system.

TAB. 2.3 – Blind Estimation Results

	b_1	b_2	a_1	a_2
true value	0.0997	-0.0902	-1.8858	0.9048
$mean(\hat{\theta})$	0.0997	-0.0902	-1.8875	0.9062
$std(\hat{\theta})$	0	0.0005	0.0030	0.0029
RMSE	0	0.0051	0.0018	0.0035

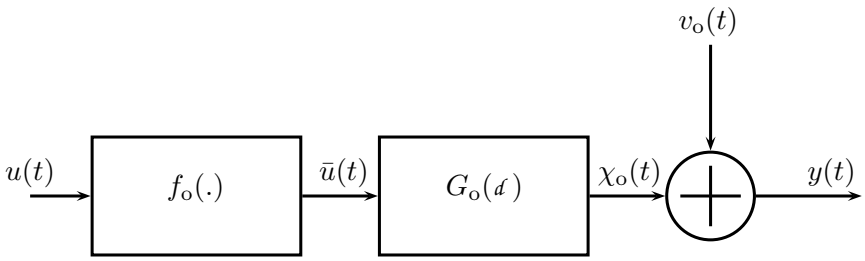


FIG. 2.7 – CT Hammerstein block representation

2.4 Continuous-time Hammerstein model identification

Even if measured data are sampled, the underlying dynamic of a real system is continuous and direct continuous-time model identification methods regained interest in the recent years [GW08]. The advantage of using direct continuous-time model identification has been pointed out in many different contexts in the LTI framework [RG04], [GW08], [GL09], [GL10b], [GL10a], [GGYdH09], [GdH05]. Nonetheless, a survey by Rao and Unbehauen [RH06] shows that CT model identification methods applied to Hammerstein models are poorly represented in literature and only a few methods can be found. In [PU95], the authors focus on the time-derivative approximation problems while solving the optimization problem using least squares. A non-parametric method can be found in [Gre00] while an approach dedicated to periodic input signals can be found in [ZRBW06]. To the best of the author's knowledge, the parametric estimation problem has not been addressed yet for CT Hammerstein models which focus with some colored added noise. Consequently, this section presents an RIV algorithm for direct CT model identification for CT Hammerstein models.

2.4.1 System description

Consider the CT Hammerstein data generating system depicted in Figure 2.7 corresponding to the following input-output relationship :

$$\mathcal{S}_o \begin{cases} \bar{u}(t) = f_o(u(t)), \\ \chi_o(t) = G_o(\not{d})\bar{u}(t), \\ y(t) = \chi(t) + v_o(t), \end{cases} \tag{2.62}$$

where

$$G_o(\not{d}) = \frac{B_o(\not{d})}{A_o(\not{d})} \tag{2.63}$$

and $B_o(\not{d})$ and $A_o(\not{d})$ are polynomials in the differential operator $\not{d}$ ($\not{d}^i x(t) = \frac{\mathrm{d}^i x(t)}{\mathrm{d}t^i}$) of respective degree n_b and n_a ($n_a \geq n_b$).

In terms of identification we can assume that sampled measurements of (y, u) are available at a sampling time $k\mathtt{T_s} > 0$. Hence, we will denote the discrete time samples of these signals as $u(t_k) = u(k\mathtt{T_s})$, where $k \in \mathbb{Z}$. The basic idea to solve the noisy *continuous-time* (CT) modelling problem is to assume that the CT noise process $v_o(t)$ can be written at the sampling instances as a *discrete-time* (DT) white noise process filtered by a DT transfer function (see Section 1.6.1). Again,

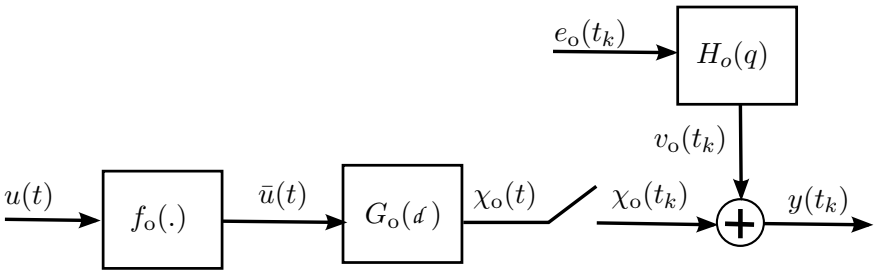


FIG. 2.8 – Hybrid Hammerstein block representation

the practically general case is considered where the colored noise associated with the sampled output measurement $y(t_k)$ is assumed to have a rational spectral density which might has no relation to the actual process dynamics. Therefore, v_o is represented by a discrete-time *autoregressive moving average* (ARMA) model :

$$v_o(t_k) = H_o(q)e_o(t_k) = \frac{C_o(q^{-1})}{D_o(q^{-1})}e_o(t_k) \tag{2.64}$$

where $e_o(t_k)$ is a DT zero mean white noise process, q^{-1} is the backward time shift operator, i.e. $q^{-i}u(t_k) = u(t_{k-i})$, and C_o with D_o are monic polynomials with constant coefficients with respective degree n_c and n_d. This avoids the rather difficult mathematical problem of treating sampled CT random process [GL09] and their equivalent in terms of a filtered piecewise constant CT noise source (see [PSR00]). Furthermore, all roots of $z^{n_d}D_o(z^{-1})$ and $z^{n_c}C_o(z^{-1})$ are inside the unit disc. Therefore, consider the Hammerstein system represented in Figure 2.8 and assume that both input and output signals, $u(t)$ and $y(t)$ are uniformly sampled at a constant sampling time $\mathtt{T}_s$ over N samples.

Consequently, in terms of (2.64), the Hammerstein system $\mathcal{S}_o$ (2.62), is described by the following input-output relationship :

$$\mathcal{S}_o \begin{cases} \bar{u}(t) = f_o(u(t)), \\ \chi_o(t) = G_o(\mathcal{A})\bar{u}(t), \\ v_o(t_k) = H_o(q)e(t_k), \\ y(t_k) = \chi_o(t_k) + v_o(t_k), \end{cases} \tag{2.65}$$

This corresponds to a so-called Hammerstein hybrid Box-Jenkins system concept already used in CT identification of LTI systems (see [PSR00], [Joh94], [YGG08]). Furthermore, in terms of (2.4), exactly the same noise assumption is made as in the classical DT Box-Jenkins models [Lju99].

2.4.2 Model considered

2.4.2.1 Process modelling

Similarly to the discrete-time case, by aiming at the convexification of the optimization problem, the static nonlinearity model is modelled as the linear sum of basis functions :

$$f(u(t),\rho)=\sum_{i=1}^{l}\alpha_i(\rho)\gamma_i(u(t)),\qquad \alpha_1=1 \tag{2.66}$$

while the CT linear part can be parameterized such that :

$$\chi(t)=G(\mathscr{d},\rho)\bar{u}(t)=\frac{B(\mathscr{d},\rho)}{A(\mathscr{d},\rho)}f(u(t),\rho) \tag{2.67}$$

with

$$\begin{cases} A(\mathscr{d},\rho)=\mathscr{d}^{n_a}+\sum_{i=1}^{n_\mathrm{a}}a_i\mathscr{d}^{n_a-i}, \\ B(\mathscr{d},\rho)=\sum_{j=0}^{n_\mathrm{b}}b_j\mathscr{d}^{n_b-j}. \end{cases}$$

Just as in the DT case, both equations (2.66) and (2.67) can be combined such that :

$$\chi(t)=\frac{B(\mathscr{d},\rho)}{A(\mathscr{d},\rho)}\sum_{i=1}^{l}\alpha_i(\rho)\gamma_i(u(t))=\frac{1}{A(\mathscr{d},\rho)}\sum_{i=1}^{l}\underbrace{\alpha_i(\rho)B(\mathscr{d},\rho)}_{B_i(\mathscr{d},\rho)}\underbrace{\gamma_i(u(t))}_{u_i(t)}. \tag{2.69}$$

Under these modelling settings, the nonlinearity model and the linear process model can be combined into a process model, denoted by $\mathcal{G}_\rho$ and defined in the form :

$$\mathcal{G}_\rho : (A(\mathscr{d},\rho),B_i(\mathscr{d},\rho)) \tag{2.70}$$

where the polynomials A and B_i are parameterized as

$$\mathcal{G}_\rho \quad \begin{cases} A(\mathscr{d},\rho)=1+\sum_{i=1}^{n_\mathrm{a}}a_i\mathscr{d}^{n_a-i}, \\ B_i(\mathscr{d},\rho)=\alpha_i\sum_{j=0}^{n_\mathrm{b}}b_j\mathscr{d}^{n_b-j}, i=1\dots l \end{cases}$$

The associated model parameters are stacked columnwise in the parameter vector ρ,

$$\rho=\begin{bmatrix}\mathbf{a}\\ \alpha_1\mathbf{b}\\ \vdots\\ \alpha_l\mathbf{b}\end{bmatrix}\in\mathbb{R}^{n_\rho},\mathbf{a}=\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_{n_a}\end{bmatrix}\in\mathbb{R}^{n_a},\mathbf{b}=\begin{bmatrix}b_0\\ b_1\\ \vdots\\ b_{n_b}\end{bmatrix}\in\mathbb{R}^{n_b+1}, \tag{2.72}$$

with $n_\rho=n_a+l(n_b+1)$.
Introduce also $\mathcal{G}=\{\mathcal{G}_\rho \mid \rho\in\mathbb{R}^{n_\rho}\}$, as the collection of all process models in the form of (2.70).

2.4.2.2 Noise model

The noise model is denoted by $\mathcal{H}$ and defined as an LTI transfer function :

$$\mathcal{H}_\eta : (H(q,\eta)) \tag{2.73}$$

where H is a monic rational function given in the form of

$$H(q,\eta) = \frac{C(q^{-1},\eta)}{D(q^{-1},\eta)} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_\mathrm{c}} q^{-n_\mathrm{c}}}{1 + d_1 q^{-1} + \ldots + d_{n_\mathrm{d}} q^{-n_\mathrm{d}}}. \tag{2.74}$$

The associated model parameters η are stacked columnwise in the parameter vector,

$$\eta = \begin{bmatrix} c_1 & \ldots & c_{n_\mathrm{c}} & d_1 & \ldots & d_{n_\mathrm{d}} \end{bmatrix}^\top \in \mathbb{R}^{n_\eta}, \tag{2.75}$$

where $n_\eta = n_\mathrm{c} + n_\mathrm{d}$. Additionally, denote $\mathcal{H} = \{\mathcal{H}_\eta \mid \eta \in \mathbb{R}^{n_\eta}\}$, the collection of all noise models in the form of (2.73).

2.4.2.3 Whole model

With respect to the given process and noise parts $(\mathcal{G}_\rho, \mathcal{H}_\eta)$, the parameters can be collected as $\theta = \begin{bmatrix} \rho^\top & \eta^\top \end{bmatrix}$ and the signal relations of the CT Hammerstein BJ model, denoted in the sequel as $\mathcal{M}_\theta$, are defined as :

$$\mathcal{M}_\theta \begin{cases} \chi(t) = \dfrac{\sum_{i=1}^{l} B_i(\not{d}\,,\rho) u_i(t)}{A(\not{d}\,,\rho)}, \\ v(t_k) = \frac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k), \\ y(t_k) = \chi(t_k) + v(t_k), \end{cases} \tag{2.76}$$

with $B_i(\not{d}\,,\rho) = \alpha_i(\rho) B(\not{d}\,,\rho)$ and $u_i(t) = \gamma_i(u(t))$.
Based on this model structure, the model set, denoted as $\mathcal{M}$, with the linear process $(\mathcal{G}_\rho)$ and noise $(\mathcal{H}_\eta)$ models parameterized independently, takes the form

$$\mathcal{M} = \left\{ (\mathcal{G}_\rho, \mathcal{H}_\eta) \;\mid\; \mathrm{col}(\rho,\eta) = \theta \in \mathbb{R}^{n_\rho + n_\eta} \right\}. \tag{2.77}$$

This set corresponds to the set of candidate models in which we seek the best fitting model using data gathered from $\mathcal{S}_\mathrm{o}$ under a given identification criterion (cost function).

Remarks

It has to be noticed that, just as in the DT case, this model transforms the Hammerstein structure into an augmented LTI *Multi Input Single Ouput* model structure such as presented in Figure 2.9. Consequently, the number of parameters to be estimated is not minimal as $n_\rho = n_a + l(n_b+1)$ which is in general greater than $n_{\rho_\mathrm{L}} + n_{\rho_\mathrm{NL}} = n_a + l + (n_b+1)$. Therefore, as the model is not minimal, the optimal estimation of this augmented MISO model does not correspond to the optimal estimates of the true Hammerstein model. Nonetheless, the gain granted using this modelling is the possible linear regression form and consequently, the convexification of the optimization problem. In order to define the identification problem it is firstly necessary to define minimization criterion.

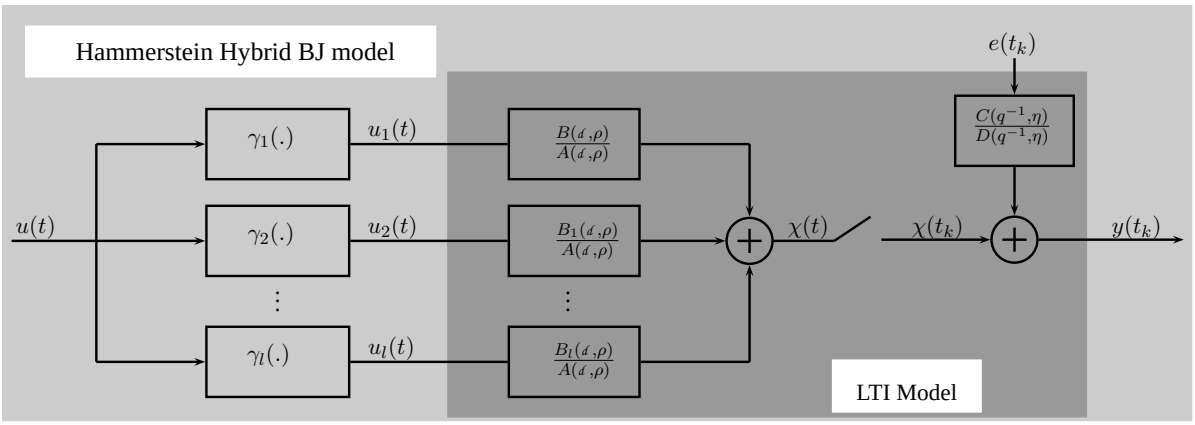


FIG. 2.9 – CT Hammerstein augmented model

2.4.3 Predictors and prediction error

As the model structure given in (2.76) is now equivalent to a *Multi Input Single Ouput* LTI structure, the PEM framework can be defined in the presented CT case by extending the DT framework from [Lju99]. Similarly to the DT case, one is concerned about finding a model in a given Hammerstein model structure $\mathcal{M}$, which minimizes the statistical mean of the squared prediction error based on past samples of (y, u).

2.4.3.1 System reformulation and prediction error

Even if the prediction error minimization is commonly used in the presented hybrid context (mostly in the LTI case), it is important for the reader to get a clear presentation of the prediction error framework for the presented Hammerstein model in order to address the identification problem. The augmented model structure given in (2.76) is now an LTI structure, and therefore, this section logically presents the extension of the PEM framework for DT models from [Lju99] as described in Section 1.4.

If the system belongs to the model set defined, it is possible to express the CT Hammerstein system as a MISO LTI system by rewriting the signal relations of (2.65) as

$$\mathcal{S}_{\mathrm{o}} \begin{cases} \chi_{\mathrm{o}}(t) = \frac{B_{\mathrm{o}}(\mathcal{A})}{A_{\mathrm{o}}(\mathcal{A})} \sum_{i=1}^{l} \alpha_{\mathrm{o},i} \gamma_i(u(t)) = \frac{1}{A_{\mathrm{o}}(\mathcal{A})} \sum_{i=1}^{l} \underbrace{\alpha_{\mathrm{o},i} B_{\mathrm{o}}(\mathcal{A})}_{B_{\mathrm{o},i}(\mathcal{A})} \underbrace{\gamma_i(u(t))}_{u_i(t)} \\ v_{\mathrm{o}}(t_k) = \frac{C_{\mathrm{o}}(q^{-1})}{D_{\mathrm{o}}(q^{-1})} e(t_k) \\ y(t_k) = \chi_{\mathrm{o}}(t_k) + v_{\mathrm{o}}(t_k) \end{cases} \tag{2.78}$$

where $\alpha_{\mathrm{o},1} = 1$. Note that in this way, the process part of the Hammerstein BJ model is rewritten as a *Multiple-Input Single-Output* (MISO) system with l inputs $\{u_i\}_{i=1}^{l}$. By using (2.78), (2.65) can be rewritten in terms of the sampled output signal $y(t_k)$ as

$$y(t_k) = \left(\frac{\sum_{i=1}^{l} B_{\mathrm{o},i}(\mathcal{A}) u_i}{A_{\mathrm{o}}(\mathcal{A})} \right)(t_k) + \frac{C_{\mathrm{o}}(q^{-1})}{D_{\mathrm{o}}(q^{-1})} e_{\mathrm{o}}(t_k). \tag{2.79}$$

which is a sampled LTI representation of the system defined in (2.65).

Given the assumption that $v_{\mathrm{o}}(t_k) = H_{\mathrm{o}}(q)e_{\mathrm{o}}(t_k)$, that $C(q^{-1})$ and $D(q^{-1})$ are monic polynomials, (2.64) can be rewritten in the form

$$v_{\mathrm{o}}(t_k) = e_{\mathrm{o}}(t_k) + \sum_{i=1}^{\infty} h_i e_{\mathrm{o}}(t_{k-i}). \tag{2.80}$$

This guarantees that the knowledge of $\{v_{\mathrm{o}}(\tau)\}_{\tau \leq k-1}$ ensures the knowledge of $\{e_{\mathrm{o}}(\tau)\}_{\tau \leq k-1}$. Therefore, by using the classical approach [Lju99], the prediction of $v_{\mathrm{o}}(t_k)$ is considered as the conditional expectation of $v_{\mathrm{o}}(t_k)$ based on $\{e_{\mathrm{o}}(\tau)\}_{\tau \leq k-1}$ which is given according to (2.80) :

$$\hat{v}_{\mathrm{o}}(t_k) = \hat{v}_{\mathrm{o}}(t_k|t_{k-1}) = \mathbb{E}\{v_{\mathrm{o}}(t_k)|\{e_{\mathrm{o}}(\tau)\}_{\tau \leq k-1}\} = \sum_{i=1}^{\infty} h_i e_{\mathrm{o}}(t_{k-i}) \tag{2.81}$$

where $\mathbb{E}$ is the expectation operator. Assuming that H_{o} is invertible such that $e_{\mathrm{o}}(t_k) = H_{\mathrm{o}}^{-1}(q)v_{\mathrm{o}}(t_k)$, then the classical one-step-ahead predictor can be given as [Lju99]

$$\hat{v}_{\mathrm{o}}(t_k) = v_{\mathrm{o}}(t_k|t_{k-1}) = v_{\mathrm{o}}(t_k) - e_{\mathrm{o}}(t_k) = (1 - H_{\mathrm{o}}^{-1}(q))v_{\mathrm{o}}(t_k). \tag{2.82}$$

Consequently, for the considered Hammerstein system formulated as in (2.79), the one-step-ahead predictor of $y(t_k)$ (defined as the conditional expectation $\hat{y}(t_k|t_{k-1})$ of $y(t_k)$) is given by

$$\begin{aligned}
\hat{y}(t_k) &= \left(\frac{\sum_{i=1}^{l} B_{\mathrm{o},i}(\not{d})u_i}{A_{\mathrm{o}}(\not{d})}\right)(t_k) + \hat{v}_{\mathrm{o}}(t_k) \\
\hat{y}(t_k) &= \left(\frac{\sum_{i=1}^{l} B_{\mathrm{o},i}(\not{d})u_i}{A_{\mathrm{o}}(\not{d})}\right)(t_k) + (1 - H_{\mathrm{o}}^{-1}(q))v_{\mathrm{o}}(t_k) \\
\hat{y}(t_k) &= \left(\frac{\sum_{i=1}^{l} B_{\mathrm{o},i}(\not{d})u_i}{A_{\mathrm{o}}(\not{d})}\right)(t_k) + (1 - H_{\mathrm{o}}^{-1}(q))\left[y(t_k) - \left(\frac{\sum_{i=1}^{l} B_{\mathrm{o},i}(\not{d})u_i}{A_{\mathrm{o}}(\not{d})}\right)(t_k)\right] \\
\hat{y}(t_k) &= H_{\mathrm{o}}^{-1}(q)\left(\frac{\sum_{i=1}^{l} B_{\mathrm{o},i}(\not{d})u_i}{A_{\mathrm{o}}(\not{d})}\right)(t_k) + (1 - H_{\mathrm{o}}^{-1}(q))y(t_k)
\end{aligned} \tag{2.83}$$

2.4.3.2 Prediction error model

Based on the same idea as in Subsection 2.4.3, the Hammerstein model (2.76) can be as well expressed into its MISO LTI sampled form :

$$y_\theta(t_k) = \left(\frac{\sum_{i=1}^{l} B_i(\not{d}, \rho)u_i}{A(\not{d}, \rho)}\right)(t_k) + \frac{C(q^{-1}, \eta)}{D(q^{-1}, \eta)}e(t_k). \tag{2.84}$$

Therefore, similarly to the LTI theory, the model (2.84) can be used to express the *one-step ahead prediction error* , which is defined as [Lju99] :

$$\varepsilon_\theta(t_k) = y(t_k) - \hat{y}_\theta(t_k), \tag{2.85}$$

where $\hat{y}_\theta(t_k)$ is the *one step ahead predictor* based on the model (2.76) written as in (2.84) and is defined as (see (2.83)) :

$$\hat{y}_\theta(t_k) = H^{-1}(q, \eta)\left(\frac{\sum_{i=1}^{l} B_i(\not{d}, \rho)u_i}{A(\not{d}, \rho)}\right)(t_k) + (1 - H^{-1}(q, \eta))y(t_k). \tag{2.86}$$

2.4.3.3 Prediction error minimization

Just like in the DT case, denote $\mathcal{D}_N = \{y(t_k), u(t_k)\}_{k=1}^N$ a data sequence of $\mathcal{S}_\text{o}$. Then, in order to provide an estimate of θ based on the minimization of ε_θ, an identification criterion $W(\mathcal{D}_N, \theta)$ can be introduced, like the *least squares* criterion

$$W(\mathcal{D}_N, \theta) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon_\theta^2(t_k), \tag{2.87}$$

such that the parameter estimate is

$$\hat{\theta}_N = \arg \min_{\theta \in \mathbb{R}^{n_\rho + n_\eta}} W(\mathcal{D}_N, \theta). \tag{2.88}$$

2.4.4 Identification problem statement

Based on the previous considerations, the identification problem addressed in the sequel can now be defined.

Problem 4 *Given a CT Hammerstein data-generating system $\mathcal{S}_\text{o}$ defined as in (2.62) and a data set $\mathcal{D}_N$ collected from $\mathcal{S}_\text{o}$. Assuming that the intersample behavior of the input is* a priori *known. Based on the Hammerstein BJ model structure $\mathcal{M}_\theta$ defined by (2.76), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions :*

HCTA1 *$\mathcal{S}_\text{o} \in \mathcal{M}$, i.e. there exits a θ_o defining a $\mathcal{G}_{\rho_\text{o}} \in \mathcal{G}$ and a $\mathcal{H}_{\eta_\text{o}} \in \mathcal{H}$ such that $(\mathcal{G}_{\rho_\text{o}}, \mathcal{H}_{\eta_\text{o}})$ is equal to $\mathcal{S}_\text{o}$.*

HCTA2 *$u(t_k)$ is not correlated to $e_\text{o}(t_k)$.*

HCTA3 *$\mathcal{D}_N$ is informative with respect to $\mathcal{M}$.*

HCTA4 *$\mathcal{S}_\text{o}$ is BIBO stable : for any bounded input signal u, the output of $\mathcal{S}_\text{o}$ is bounded.*

2.4.5 Optimal IV for CT Hammerstein BJ models

Using the LTI model (2.76), $y(t_k)$ can be written in the regression form :

$$y^{(n_a)}(t_k) = \varphi^\top(t_k)\rho + \tilde{v}(t_k) \tag{2.89}$$

where,

$$\begin{aligned}
\varphi(t_k) &= [\ -y^{(n_\text{a}-1)}(t_k) \quad \dots \quad -y(t_k) \quad u_1^{(n_\text{b})}(t_k) \quad \dots \quad u_1(t_k) \quad \dots \quad u_l^{(n_\text{b})}(t_k) \quad \dots \quad u_l(t_k)\ ]^\top \\
\rho &= [\ a_1 \quad \dots \quad a_{n_\text{a}} \quad b_0 \quad \dots \quad b_{n_\text{b}} \quad \dots \quad \alpha_l b_0 \quad \dots \quad \alpha_l b_{n_\text{b}}\ ]^\top
\end{aligned}$$

and

$$\tilde{v}(t_k) = A(\mathcal{A}, \rho) v(t_k).$$

where $x^{(n)}(t_k)$ denotes the sample of the n^th derivative of the signal $x(t)$ sampled at time t_k. The conventional PEM approach on (2.89) leads to the prediction error $\varepsilon_\theta(t_k)$ given as (see (1.53)) :

$$\varepsilon_\theta(t_k) = \frac{D(q^{-1}, \eta)}{C(q^{-1}, \eta)} \frac{1}{A(\mathcal{A}, \rho)} \left(A(\mathcal{A}, \rho) y(t_k) - \sum_{i=1}^{l} B_i(\mathcal{A}, \rho) u_i(t_k) \right). \tag{2.90}$$

It can be seen that (2.90) mixes CT filtering and DT filtering. A preliminary discussion, in which we considered the practically feasible situation such that only sampled measurements of the CT

signals (y, u) are available, can be found in Section 1.6. In order to calculate (2.90) and to use it in any identification criterion to estimate $\mathrm{col}(\rho, \eta)$, we need to compute $\varepsilon_\theta(t_k)$ in terms of (2.90) based on DT samples. Under this assumption, and considering that a CT filter can only be applied to sampled data through numerical approximation, the commutativity property holds between a DT filter and the numerical approximation of a CT filter. Consequently, the DT filtering of $\frac{D(q^{-1},\eta)}{C(q^{-1},\eta)}$ and the numerical approximation of the CT filter $A(\mathscr{d}\,, \rho)$ in (2.90) commute.
The later allows us to rewrite *the one-step ahead prediction error* (2.90) associated to (2.89) as

$$\varepsilon_\theta(t_k) = \left(A(\mathscr{d}\,, \rho) y_{\mathrm{f}}\right)(t_k) - \sum_{i=1}^{l} B_i(\mathscr{d}\,, \rho) u_{i\mathrm{f}}(t_k) \tag{2.91}$$

where $y_{\mathrm{f}}(t_k)$ and $u_{i,\mathrm{f}}(t_k)$ represent the outputs of the complete hybrid prefiltering operation, involving the continuous-time filtering operation using the filter (see [YGG08]) :

$$Q_{\mathrm{c}}(\mathscr{d}\,, \rho) = \frac{1}{A(\mathscr{d}\,, \rho)}, \tag{2.92}$$

and the discrete-time filtering operation using the filter :

$$Q_{\mathrm{d}}(q^{-1}, \eta) = \frac{D(q^{-1}, \eta)}{C(q^{-1}, \eta)}. \tag{2.93}$$

In other words :

$$y_{\mathrm{f}}(t_k) = \frac{D(q^{-1}, \eta)}{C(q^{-1}, \eta)} \left[\left(\frac{1}{A(\mathscr{d}\,, \rho)} y \right)(t_k) \right], \tag{2.94}$$

Based on (2.91), the associated linear-in-the-parameters model takes the form [YGG08] :

$$y_{\mathrm{f}}^{(n_a)}(t_k) = \varphi_{\mathrm{f}}^{\top}(t_k)\rho + \tilde{v}_{\mathrm{f}}(t_k), \tag{2.95}$$

where

$$\begin{aligned} \varphi_{\mathrm{f}}(t_k) &= \left[-y_{\mathrm{f}}^{(n_{\mathrm{a}}-1)}(t_k) \quad \dots \quad -y_{\mathrm{f}}(t_k) \quad u_{1\mathrm{f}}^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_{1\mathrm{f}}(t_k) \quad \dots \quad u_{l\mathrm{f}}^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_{l\mathrm{f}}(t_k) \right]^{\top} \\ \tilde{v}_{\mathrm{f}}(t_k) &= Q_{\mathrm{d}}(q, \eta) Q_{\mathrm{c}}(\mathscr{d}\,, \rho) \tilde{v}(t_k). \end{aligned} \tag{2.96}$$

Consequently when the data is generated from $\mathcal{S}_{\mathrm{o}}$, if $\theta = \theta_{\mathrm{o}}$, then $\tilde{v}_{\mathrm{f}}(t_k) = e_{\mathrm{o}}(t_k)$ which is a white noise.
Therefore, according to the conditions for optimal IV estimates (see C4 in Section 1.5.5), the optimal filtered instrument for the augmented LTI MISO model structure (2.76) depicted in Figure 2.9 is given as :

$$\zeta^{\mathrm{opt}}(t_k) = \left[-\chi_{\mathrm{o}}^{(n_{\mathrm{a}}-1)}(t_k) \quad \dots \quad -\chi_{\mathrm{o}}(t_k) \quad u_1^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_1(t_k) \quad \dots \quad u_l^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_l(t_k) \right]^{\top} \tag{2.97}$$

while the optimal filter is given as the filter chain involving the continuous-time filtering operation using the filter (see [YGG08]) :

$$L_{\mathrm{c}}^{\mathrm{opt}} = Q_{\mathrm{c}}(\mathscr{d}\,, \rho_{\mathrm{o}}) = \frac{1}{A_{\mathrm{o}}(\mathscr{d}\,)}, \tag{2.98}$$

and the discrete-time filtering operation using the filter :

$$L_{\mathrm{d}}^{\mathrm{opt}} = Q_{\mathrm{d}}(q, \eta_{\mathrm{o}}) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})}. \tag{2.99}$$

In other words :

$$y_{\mathrm{f}}(t_k) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})} \left[\left(\frac{1}{A_{\mathrm{o}}(\mathscr{d}\,)} y \right)(t_k) \right], \tag{2.100}$$

2.4.6 Hammerstein RIVC (HRIVC) algorithm for BJ models

In the present CT Hammerstein case, as in the LTI case presented in 1.5.5, the 'auxiliary model' used to generate the instrumental variables, as well as the associated prefilters (2.92) and (2.93) , are updated, based on the parameter estimates obtained at the previous iteration.

Algorithm 4 (HRIVC)

Step 1 *Generate an initial estimate* $\hat{\rho}^{(0)}$ *using the* SRIVC *algorithm (by simply firstly forcing* $f(u(t)) = u(t)$*). Set* $C(\text{ƈ}, \hat{\eta}^{(0)}) = D(\text{ƈ}, \hat{\eta}^{(0)}) = 1$*. Set* $\tau = 0$*.*

Step 2 *Compute an estimate of* $\chi(t_k)$ *via*

$$\hat{\chi}(t_k) = \frac{\sum_{i=1}^{l} B_i(\text{ƈ}, \hat{\rho}^{(\tau)}) u_i(t_k)}{A(\text{ƈ}, \hat{\rho}^{(\tau)})},$$

where $\hat{\rho}^{(\tau)}$ *is the estimate obtained at the previous iteration. According to assumption HCTA4 each* $\hat{\chi}$ *is bounded.*

Step 3 *Compute the continuous-time filter* $L_{\mathrm{c}}(\text{ƈ}, \hat{\rho}^{(\tau)})$ *:*

$$L_{\mathrm{c}}(\text{ƈ}, \hat{\rho}^{(\tau)}) = \frac{1}{A(\text{ƈ}, \hat{\rho}^{(\tau)})} \tag{2.101}$$

Step 4 *Use the CT filter* $L_{\mathrm{c}}(\text{ƈ}, \hat{\rho}^{(\tau)})$ *as well as* $\hat{\chi}(t_k)$ *in order to generate the derivatives needed :*

$$\begin{aligned}
&\{L_{\mathrm{c}}(\text{ƈ}, \hat{\rho}^{(\tau)}) \hat{\chi}^{(i)}(t_k)\}_{i=0}^{n_{\mathrm{a}}-1} \\
&\{L_{\mathrm{c}}(\text{ƈ}, \hat{\rho}^{(\tau)}) y^{(i)}(t_k)\}_{i=0}^{n_{\mathrm{a}}-1} \\
&\{L_{\mathrm{c}}(\text{ƈ}, \hat{\rho}^{(\tau)}) u_i^{(j)}(t_k)\}_{i=1,j=0}^{l,n_{\mathrm{b}}}
\end{aligned}$$

Step 5 *Compute the discrete-time filter :*

$$L_{\mathrm{d}}(q^{-1}, \hat{\eta}^{(\tau)}) = \frac{D(q^{-1}, \hat{\eta}^{(\tau)})}{C(q^{-1}, \hat{\eta}^{(\tau)})}$$

Step 6 *By applying the DT filter on the estimated derivatives obtained from Step 4, the needed filtered signals* $\{y_{\mathrm{f}}^{(i)}\}_{i=0}^{n_{\mathrm{a}}-1}$, $\{\hat{\chi}_{\mathrm{f}}^{(i)}\}_{i=0}^{n_{\mathrm{a}}-1}$ *and* $\{u_{\mathrm{if}}^{(j)}\}_{i=1,j=0}^{l,n_{\mathrm{b}}}$ *are computed.*

Step 7 *Build the filtered regressor* $\varphi_{\mathrm{f}}(t_k)$ *and in terms of C4 (Section 1.5.5) the filtered estimated instrument* $\hat{\zeta}_{\mathrm{f}}(t_k)$ *which equal in the given context :*

$$\begin{aligned}
\varphi_{\mathrm{f}}(t_k) = \Big[& -y_{\mathrm{f}}^{(n_{\mathrm{a}}-1)}(t_k) \quad \dots \quad -y_{\mathrm{f}}(t_k) \quad u_{1\mathrm{f}}^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_{1\mathrm{f}}(t_k) \\
& \dots \quad u_{l\mathrm{f}}^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_{l\mathrm{f}}(t_k) \Big]^{\top} \\
\hat{\zeta}_{\mathrm{f}}(t_k) = \Big[& -\hat{\chi}_{\mathrm{f}}^{(n_{\mathrm{a}}-1)}(t_{k-1}) \quad \dots \quad -\hat{\chi}_{\mathrm{f}}(t_k) \quad u_{1\mathrm{f}}^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_{1\mathrm{f}}(t_k) \\
& \dots \quad u_{l\mathrm{f}}^{(n_{\mathrm{b}})}(t_k) \quad \dots \quad u_{l\mathrm{f}}(t_k) \Big]^{\top}
\end{aligned} \tag{2.102}$$

Step 8 *The IV optimization problem can be stated in the form*

$$\hat{\rho}^{(\tau+1)}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k)\right]\rho - \left[\frac{1}{N}\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)y_{\mathrm{f}}^{(n_{\mathrm{a}})}(t_k)\right]\right\|^2 \tag{2.103}$$

where the solution is obtained as

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\varphi_{\mathrm{f}}^{\top}(t_k)\right]^{-1} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)y_{\mathrm{f}}^{(n_{\mathrm{a}})}(t_k).$$

The resulting $\hat{\rho}^{(\tau+1)}(N)$ is the IV estimate of the process model associated parameter vector at iteration $\tau+1$ based on the prefiltered input/output data.

Step 9 *An estimate of the noise signal v is obtained as*

$$\hat{v}(t_k) = y(t_k) - \chi(t_k, \hat{\rho}^{(\tau)}). \tag{2.104}$$

Based on $\hat{v}$, the estimation of the noise model parameter vector $\hat{\eta}^{(\tau+1)}$ follows, using in this case the `ARMA` *estimation algorithm of the* `MATLAB` *identification toolbox (an IV approach can also be used for this purpose, see [You08]).*

Step 10 *If $\theta^{(\tau+1)}$ has converged or the maximum number of iterations is reached, then stop, else increase τ by 1 and go to Step 2.*

At the end of the iterative process, coefficients $\hat{\alpha}_i$ are not directly accessible. They are however deduced from polynomial $\hat{B}_i(\mathscr{A})$ as $B_i(\mathscr{A},\rho) = \alpha_i B(\mathscr{A},\rho)$. The hypothesis $\alpha_1 = 1$ guarantees that $B_1(\mathscr{A},\rho) = B(\mathscr{A},\rho)$ and $\hat{\alpha}_i$ can be computed from :

$$\hat{\alpha}_i = \frac{1}{n_b+1}\sum_{k=0}^{n_b} \frac{\hat{b}_{i,k}}{\hat{b}_{1,k}}, \tag{2.105}$$

where $\hat{b}_{i,k}$ is the kth coefficient of polynomial term $B_i(\mathscr{A},\rho)$ for $i = 2\ldots l$. Again, it must be pointed out that the separation of α_i and b_i coefficients can be executed using more powerful methods such as *single value decomposition* for example.
Moreover, after the convergence is complete, it is also possible to compute the estimated parametric error covariance matrix $\hat{\mathbf{P}}_\rho$ from the expression (see Section 1.5.4) :

$$\hat{\mathbf{P}}_\rho = \hat{\sigma}_e^2 \left(\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\hat{\zeta}_{\mathrm{f}}^{\top}(t_k)\right)^{-1} \tag{2.106}$$

where $\hat{\zeta}$ is the IV vector obtained at convergence and $\hat{\sigma}_e^2$ is the estimated residual variance.

Comments

By using the described algorithm, if convergence occurs, it can be seen that all conditions C1-C6 from Section 1.5.5 except for C5 are respected (a discussion about how C1 is fulfilled can be found in [SS83]) : the `HRIVC` estimates might be statistically optimal for the augmented model proposed, but the minimal number of parameters needed for representing the MISO structure and the Hammerstein structure are not equal. Consequently, the `HRIVC` estimates cannot be statistically optimal for the CT Hammerstein model structure. Nonetheless, even if not optimal, the HRIVC estimates are unbiased with a low variance as it will be seen in the result section 2.4.8.

2.4.7 HSRIVC algorithm for OE models

A simplified version of HRIVC algorithm named HSRIVC follows the exact same theory for estimation of Hammerstein output error models. It is mathematically described by, $C(q^{-1},\eta^j) = C_o(q^{-1}) = 1$ and $D(q^{-1},\eta^j) = D_o(q^{-1}) = 1$. All previous given equations remain true, and it suffices to estimate ρ^j as $\theta^j = \rho^j$. The implementation of HSRIVC is much simpler than HRIVC as there is no noise model estimation in the algorithm : the same algorithm is used except for steps 5, 6 and 9 which are simply removed.

2.4.8 Performance evaluation of the proposed HRIVC and HSRIVC algorithms

2.4.8.1 Performances of the methods

This section presents numerical evaluation of both suggested HRIVC and HSRIVC methods. For the presented example, the nonlinear block has a polynomial form, *i.e.* $\gamma_i(u(t)) = u^i(t), \forall i$ and

$$\bar{u}(t) = u(t) + 0.5u^2(t) + 0.25u^3(t), \tag{2.107}$$

where $u(t)$ follows a uniform distribution with values between -2 and 2. The system is simulated using a zero-order-hold on the input.
The linear dynamic block is first a second-order system described by :

$$G_o(\mathcal{A}) = \frac{10\mathcal{A} + 30}{\mathcal{A}^2 + \mathcal{A} + 5}. \tag{2.108}$$

The sampling period equals $\mathtt{T_s} = 0.48$s. Based on this process, two different systems $\mathcal{S}_1$ and $\mathcal{S}_2$ are defined according to the noise process associated. $\mathcal{S}_1$ is a Hammerstein output error model and therefore

$$H_o(q) = 1.$$

while $\mathcal{S}_2$ is a Hammerstein Box–Jenkins model with :

$$H_o(q) = \frac{1}{1 - q^{-1} + 0.2q^{-2}}.$$

The models considered for estimation are :

$$\mathcal{M}_{\mathtt{HRIVC}} \begin{cases} G(\mathcal{A},\rho) = \dfrac{b_0\mathcal{A} + b_1}{\mathcal{A}^2 + a_1\mathcal{A} + a_2}, \\ H(q,\eta) = \dfrac{1}{1 + d_1q^{-1} + d_2q^{-2}}, \\ f(u(t)) = u(t) + \alpha_1u^2(t) + \alpha_2u^3(t) \end{cases} \tag{2.109}$$

for the HRIVC method and

$$\mathcal{M}_{\mathtt{HSRIVC}} \begin{cases} G(\mathcal{A},\rho) = \dfrac{b_0\mathcal{A} + b_1}{\mathcal{A}^2 + a_1\mathcal{A} + a_2}, \\ H(q,\eta) = 1, \\ f(u(t)) = u(t) + \alpha_1u^2(t) + \alpha_2u^3(t) \end{cases} \tag{2.110}$$

for HSRIVC.

The result of a Monte Carlo simulation (MCs) analysis is shown in Tables 2.4 and 2.5 for the algorithms considered. The MCs results are based on $N_{\text{run}} = 500$ random realization, with the Gaussian white noise input to the ARMA noise model being selected randomly for each realization. In order to compare the statistical performance of the different approaches, the computed mean, standard deviation and RMSE (as in (2.58)) of the estimated parameters are presented. The noise added at the output is adjusted such that it corresponds to a SNR of 15dB and 5dB. The number of samples is $N = 2000$.

Table 2.4 shows that according to the theory, the `HRIVC` and `HSRIVC` methods provide similar, unbiased estimates of the model parameters. Both methods seem to be robust even at unrealistic noise level of 5dB as the RMSE remain under 22% for both methods. Results obtained using the `HRIVC` algorithm, have standard deviations which are always smaller than the ones produced by `HSRIVC`. Even though, the `HSRIVC` algorithm based on an Output Error model is a reasonable alternative to the full `HRIVC` algorithm based on a Box–Jenkins model : in practice the noise model cannot be exactly known and therefore the use of the `HRIVC` algorithm would simply raise the number of parameters to be estimated. The consequence for raising the number of parameters still leads to acceptable estimates as it can be seen for system $\mathcal{S}_1$ in Table 2.4. If the noise model is correctly assumed, it is as well correctly estimated as shown in Table 2.5.

2.4.8.2 The advantages of direct continuous-time estimation

This section shows through a chosen example the interest of using the direct CT methods with respect to the indirect DT methods. The linear part of the second system is based on a benchmark proposed by Rao and Garnier in [RG02] (see also [Lju03]) designed initially to highlight the advantages of CT estimation. It is a fourth-order, non-minimum phase system with complex poles. Its transfer function is given by :

$$G_o(\mathit{ɗ}) = \frac{-6400\mathit{ɗ} + 1600}{\mathit{ɗ}^4 + 5\mathit{ɗ}^3 + 408\mathit{ɗ}^2 + 416\mathit{ɗ} + 1600}. \tag{2.111}$$

The sampling frequency is chosen to be about ten times the bandwidth of the system under study which leads to $\mathtt{T_s} = 0.0314$s. White noise is added to the output samples.
The models take the forms :

$$\mathcal{M}_{\texttt{HSRIVC}} \begin{cases} G(\mathit{ɗ}, \rho) = \dfrac{b_o\mathit{ɗ} + b_1}{\mathit{ɗ}^4 + a_1\mathit{ɗ}^3 + a_2\mathit{ɗ}^2 + a_3\mathit{ɗ} + a_4}, \\ H(q, \eta) = 1, \\ f(u(t)) = u(t) + \alpha_1 u^2(t) + \alpha_2 u^3(t) \end{cases} \tag{2.112}$$

for `HSRIVC` and

$$\mathcal{M}_{\texttt{HSRIV}} \begin{cases} G(q, \rho) = \dfrac{\tilde{b_0}q^{-1} + \tilde{b_1}q^{-2} + \tilde{b_2}q^{-3} + \tilde{b_3}q^{-4}}{1 + \tilde{a_1}q^{-1} + \tilde{a_2}q^{-2} + \tilde{a_3}q^{-3} + \tilde{a_4}q^{-4}}, \\ H(q, \eta) = 1, \\ f(u(t)) = u(t) + \alpha_1 u^2(t) + \alpha_2 u^3(t) \end{cases} \tag{2.113}$$

for `HSRIV`.

TAB. 2.4 – Estimation results for different noise models

				b_0	b_1	a_1	a_2	α_1	α_2
system	SNR	method							
$\mathcal{S}_1$			true value	10	30	1	5	0.5	0.25
	15dB	HSRIVC	$mean(\hat{\theta})$	9.9869	30.0251	1.0003	4.9996	0.5005	0.2507
			$std(\hat{\theta})$	0.3053	0.6984	0.0084	0.0236	0.0113	0.0116
			RMSE	0.0305	0.0233	0.0084	0.0047	0.0227	0.0464
	5dB	HSRIVC	$mean(\hat{\theta})$	9.9834	29.8845	0.9987	4.9960	0.5061	0.2553
			$std(\hat{\theta})$	0.9508	2.3145	0.0267	0.0789	0.0359	0.0383
			RMSE	0.0950	0.0772	0.0267	0.0158	0.0728	0.1544
$\mathcal{S}_2$			true value	10	30	1	5	0.5	0.25
	15dB	HSRIVC	$mean(\hat{\theta})$	9.9957	29.8760	1.0001	4.9991	0.5026	0.2523
			$std(\hat{\theta})$	0.3670	1.5660	0.0170	0.0436	0.0201	0.0180
			RMSE	0.0367	0.0523	0.0169	0.0087	0.0405	0.0723
		HRIVC	$mean(\hat{\theta})$	9.9906	30.0172	1.0006	5.0020	0.5008	0.2506
			$std(\hat{\theta})$	0.2497	0.8954	0.0119	0.0265	0.0118	0.0115
			RMSE	0.0250	0.0298	0.0119	0.0053	0.0236	0.0460
	5dB	HSRIVC	$mean(\hat{\theta})$	10.0882	29.6146	1.0010	4.9814	0.5080	0.2604
			$std(\hat{\theta})$	1.0764	4.4585	0.0517	0.1291	0.0610	0.0542
			RMSE	0.1079	0.1490	0.0517	0.0261	0.1230	0.2208
		HRIVC	$mean(\hat{\theta})$	10.049	30.0277	0.9998	4.9980	0.5015	0.2522
			$std(\hat{\theta})$	0.7861	2.8278	0.0379	0.0871	0.0369	0.0366
			RMSE	0.0787	0.0942	0.0378	0.0174	0.0738	0.1466

TAB. 2.5 – Noise estimation results for different noise levels

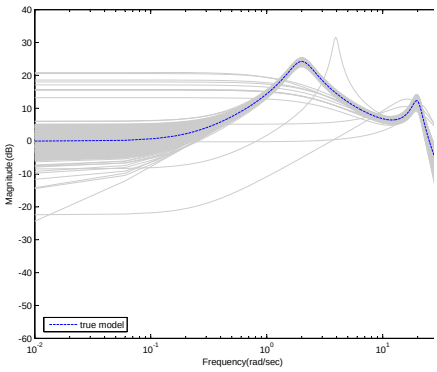
System	SNR	method		d_1	d_2
$\mathcal{S}_2$			true value	-1	0.2
	15dB	HRIVC	$mean(\hat{\theta})$	-1.0002	0.2005
			$std(\hat{\theta})$	0.0219	0.0223
			RMSE	0.0218	0.1112
	5dB	HRIVC	$mean(\hat{\theta})$	-0.9997	0.1994
			$std(\hat{\theta})$	0.0227	0.0219
			RMSE	0.0227	0.1096

The result of an MCs is displayed in Figures 2.10(a) and 2.10(b) for the `HSRIVC` and `HRIVC` algorithms. The MCs results are based on $N_{\text{run}} = 500$ random realization, with the Gaussian white

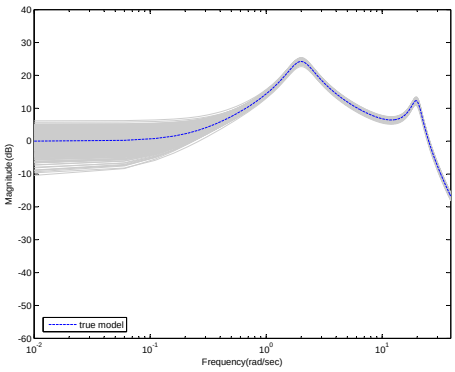
noise being selected randomly for each realization. The noise added at the output is adjusted such that it corresponds to a SNR of 5dB. The number of samples is $N = 2000$.
Figures 2.10(a) and 2.10(b) display the magnitude Bode plots of the DT and CT estimated linear models. It can be firstly noticed that both linear models present similar results for low frequencies whereas for high frequencies, the estimated CT models are much better. Both methods correctly estimate both resonance peaks. On the other side, the DT method appears to be less reliable, as for some realizations, the algorithm does not converge to acceptable values even though the initialization step is the same for both methods. By only looking at the magnitude Bode diagrams and only considering realizations which converged, both methods give satisfactory results. However, when looking at nonlinear function estimations (Figures 2.10(c) and 2.10(d)), the DT method hands out results with a very large variance while the CT approach delivers a set of estimated functions centered nearly exactly on the true nonlinear function. This can be explained by two facts : the DT version of the Hammerstein model (assuming the appropriate zero order hold) increases the number of parameters to be estimated for the numerator polynomial and therefore results in parameters estimates with large variance. Furthermore, in the DT case, the numerator coefficients are so close to null that a small absolute error produces a large relative error. Estimated $\hat{\alpha}_i$ coefficients, which are directly deduced from $\hat{B}$ (see (2.105)), dramatically suffer from this particular situation.

2.5 Conclusion

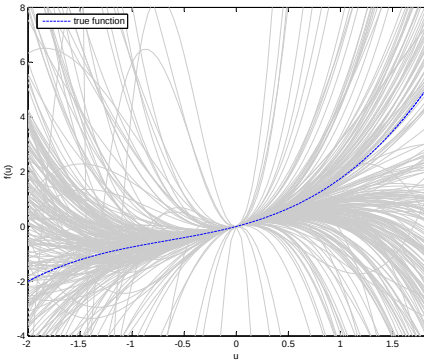
In this chapter, some methods dedicated to the Hammerstein nonlinear models were investigated. Both contributions to the closed-loop and continuous-time case have been published and can be found in [LGGY08] and [LGG09]. Through a relevant set of examples, it was possible to show that the `HRIV` approach is robust to noise conditions and to noise error modelling. Moreover, in most cases, the nonlinearity is used as an approximation of the model complexity and it was shown that in the blind estimation case, the RIV-based method resulted in some interesting results. The presented methods are suboptimal as they estimate a larger number of parameters than the minimum needed for the system description. However, the linear regression approach guarantees convergence, if it occurs, to the global minimum. Nonetheless, the variance in the estimated parameters is acceptable in practical conditions, and if not satisfactory, the estimates can be used as initialisation values for some optimal method which usually are posed into some nonlinear optimization problems and often rely on some robust initialisation. The refined instrumental variable approach for Hammerstein models remains an interesting estimation method for practical applications where the noise condition are unknown. The estimation of CT Hammerstein systems operating in closed loop is not yet developed. Nonetheless, all the keys for solving this problem can be extrapolated from this chapter, and the remaining problems lie in the field of implementation.



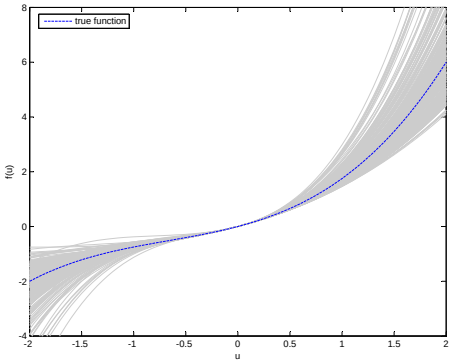
(a) Magnitude Bode plots of the identified DT `HSRIV` models together with the true system.



(b) Magnitude Bode plots of the identified CT `HSRIVC` models together with the true system.



(c) Nonlinear function estimated with DT `HSRIV` models together with the true nonlinear function.



(d) Nonlinear function estimated with CT `HSRIVC` models together with the true nonlinear function.

FIG. 2.10 – DT vs CT Hammerstein model identification

Chapitre 3

Refined Instrumental Variable Methods for LPV Box-Jenkins Models

3.1 Introduction

Dealing with nonlinear models without any structure is often found infeasible in practice. This rises the need for system descriptions that form an intermediate step between *Linear Time-Invariant* (LTI) systems and nonlinear/time-varying plants. To cope with these expectations, the model class of *Linear Parameter-Varying* (LPV) systems provides an attractive candidate. In LPV systems the signal relations are considered to be linear just as in the LTI case, but the parameters are assumed to be functions of a measurable time-varying signal, the so-called *scheduling variable* $p : \mathbb{Z} \mapsto \mathbb{P}$. The compact set $\mathbb{P} \subseteq \mathbb{R}^{n_\mathbb{P}}$ denotes the *scheduling space*. The scheduling variable p represents the nonlinear or time-varying nature of the modeled dynamics, e.g. p describes the changes in the operating conditions of the plant, time-variation of the system dynamics or external effects like temperature changes. This LPV modelling concept allows for a wide representation capability of physical processes but the real practical significance of the LPV framework until now lays in its well worked out and industrially reputed control synthesis approaches, e.g. [AG95, Sch96, WD06b], that have led to many successful applications of LPV control in practice [MB04, ØSB08, WNP+06, DS01].
The existing LPV identification approaches are almost exclusively formulated in discrete-time, commonly assuming static dependence on the scheduling parameter (dependence only on the instantaneous value of the scheduling variable), and they are mainly characterized by the type of LPV model structure used : *Input-Output* (IO) [BG00, BG02a, WD06a] *State-Space* (SS) [LM07, vWV09, dSRdC07] or *Orthogonal Basis Functions* (OBFs) based models [THV09] (see [Tót10] for an overview of existing methods).

Often, an important advantage of IO models (especially for SISO systems) is that they can be directly derived from physical/chemical laws in their continuous form such as in the considered field of application.

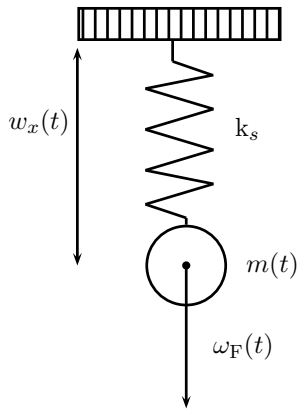


FIG. 3.1 – Varying mass connected to a spring

Example

For example, let us investigate the modeling of the motion of a varying mass connected to a spring (see [Tót10]). This problem is one of the typical phenomena occurring in systems with time-varying masses like in motion control (robotics, rotating crankshafts, rockets, conveyor systems, excavators, cranes), biomechanics, and in fluid-structure interaction problems. Denote by ω_x the position of the varying mass m. Let $\mathrm{k}_s > 0$ be the spring constant, introduce ω_{F} as the force acting on the mass, and assume that there is no damping. By Newton's second law of motion, the following equation holds :

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(m\frac{\mathrm{d}}{\mathrm{d}t}\omega_x\right) = \omega_{\mathrm{F}} - \mathrm{k}_s\omega_x, \tag{3.1}$$

or equivalently

$$m\frac{\mathrm{d}^2}{\mathrm{d}^2t}\omega_x + \left(\frac{\mathrm{d}}{\mathrm{d}t}m\right)\frac{\mathrm{d}}{\mathrm{d}t}\omega_x + \mathrm{k}_s\omega_x = \omega_{\mathrm{F}}. \tag{3.2}$$

It is immediate that by taking m as a scheduling variable, the behavior of this plant can be described as an IO LPV system, preserving the physical insight of Newton's second law.

■

Therefore, it might be more natural to express a given physical system through an IO operator form or transfer function modelling. A comparison between IO and SS model based approaches can be found in [SJ00] for linear systems.

In the LPV case, most of the methods developed for IO models based identification are derived under a linear regression form [WD06a, BG00, BG99]. By using the concepts of the LTI PE framework, recursive LS and IV methods have been also introduced [GBFB06, BKL08]. However, it has been only recently understood how a prediction error framework can be established for the estimation of general LPV models [Tót10]. Due to the linear regressor based estimation, the usual model structure in existing methods is assumed to be *auto regressive with exogenous input* (ARX). Even if a non-statistically optimal IV method has been recently introduced in [BKL08] for LPV *Output Error* (OE) models, no method has been proposed so far to deal with colored noise. Moreover, it can be shown that it is generally impossible to reach statistically optimal estimates by using linear regression as presented so far in the literature. These imply, that there is lack of an LPV identification approach, which is capable of efficiently estimating LPV-IO models under colored noise conditions, *e.g.* as in

a *Box-Jenkins* (BJ) setting, which is the case in many practical applications and more specifically in the considered field of application.
By aiming at fulfilling this gap, an estimation method is developed in Section 3.2 for LPV-IO BJ discrete-time models in the SISO case which is extended in Section 3.3 to the closed-loop case.
Furthermore, LPV controllers are commonly synthesized in CT as stability and performance requirements of the closed loop behavior can be more conveniently expressed in CT, like in a mixed-sensitivity setting [ZD98]. Therefore, the current design tools focus on continuous-time LPV controller synthesis requiring accurate and low order CT models of system. However, LPV identification methods are almost exclusively developed for DT as in this setting it is much easier to handle the estimation of parameter-varying dynamics. The only available CT method (to the author's knowledge) uses a local, or a so-called multiple-model type of approach, where CT-LTI models of the LPV system are estimated for constant trajectories of p, *i.e.* at certain operating operating points, by using a frequency domain method and then the resulting models are interpolated over $\mathbb{P}$. However such an approach is unable to capture the global behavior of the system (limited number of identified CT-LTI models, no information about transient behavior from one operating condition to the other) and is affected by interpolation problems described in [MSJS99, TFHV07]. As developing CT-LPV models based on first principle laws is a very costly and time consuming process, often resulting in a high order model unsuitable for control design in industrial world. Concerning natural systems with an LPV nature, no method is yet able to exactly capture their behavior using their natural CT expression. Consequently, there is a growing need of the LPV framework for efficient identification methods that directly deliver reliable CT models.
In practice, CT systems can only be identified based on sampled measured data records. Thus in general, for delivering a CT model estimate, the available approaches in system identification can be categorized as follows [RH06] :

- **Indirect approaches :** These methods involve the identification of a DT model in a completely DT setting which is followed by the transformation of the DT model estimate into a CT form.
- **Direct approaches :** The methods formulate the identification of the CT model directly based on samples of the measured CT signals.

Unfortunately, transformation of DT-LPV models to CT-LPV models is more complicated than in the LTI case and despite recent advances in LPV discretization theory (see [TLHV09, THV10]), the theory of CT realization of DT models is still in an immature state. Thus, it is often difficult in practice to obtain an adequate CT realization from an identified DT model. In order to illustrate the underlaying problems, consider the following simple CT-LPV model

$$\frac{\mathrm{d}}{\mathrm{d}t}y(t) + p(t)y(t) = bu(t), \tag{3.3}$$

where p, y and u are the scheduling, output and input signals of the system respectively and $b \in \mathbb{R}$ is a constant parameter. When approximating the derivative in DT by, for example, using the backward Euler approximation : $\frac{\mathrm{d}}{\mathrm{d}t}y(t_k) = \frac{y(t_k)-y(t_{k-1})}{\mathtt{T_s}}$ with $\mathtt{T_s} > 0$ the sampling period, (3.3) transforms into :

$$y(t_k) = \frac{1}{1+\mathtt{T_s}p(t_k)}y(t_{k-1}) + \frac{b\mathtt{T_s}}{1+\mathtt{T_s}p(t_k)}u(t_k). \tag{3.4}$$

This discretized model has now two p-dependent coefficients to be estimated instead of the one single constant parameter in (3.3). Moreover, the dependence of the coefficients on p is not linear

anymore but rational with a singularity whenever $p(t_k) = -\frac{1}{\mathtt{T_s}}$. An alternative way to approximate derivatives in DT is to apply a forward Euler approximation : $\frac{\mathrm{d}}{\mathrm{d}t}y(t_k) = \frac{y(t_{k+1})-y(t_k)}{\mathtt{T_s}}$, which turns the CT model into :

$$y(t_k) = (1 - \mathtt{T_s}p(t_{k-1}))y(t_{k-1}) + b\mathtt{T_s}u(t_{k-1}). \tag{3.5}$$

This discretized model has only one p-dependent coefficient and the linearity of the dependence is preserved, however now the model equation is dependent on $p(t_{k-1})$ instead of $p(t_k)$. This so-called dynamic dependence (dependence of the model coefficients on time-shifted versions of p) is a common result of model transformations in the LPV case and rises problems in LPV system identification and control alike (see [THV08]). Furthermore, it is well known in numerical analysis that the forward Euler approximation is more sensitive for the choice of $\mathtt{T_s}$ in terms of numerical stability than the backward Euler approximation [Atk89]. This means that (3.5) requires much faster sampling rate than (3.4) to give a stable approximation of the system and it is more sensitive to parameter uncertainties which rises problems if (3.5) is used for estimation. Consequently it can be seen that even for a very simple CT-LPV model, estimation of a DT model with the purpose of obtaining afterwards a CT realization is a tedious task with many underlaying problems for which there are no general theoretical solutions available.
In this chapter, we aim through Section 3.4 to provide the very first step towards filling the lack of LPV direct identification methods via the introduction of a direct CT identification approach that benefits from the properties of IV methods.

3.2 Discrete-time open-loop LPV model identification

3.2.1 System description

Consider the data generating LPV system described by the following equations :

$$\mathcal{S}_{\mathrm{o}} \begin{cases} A_{\mathrm{o}}(p_k, q^{-1})\chi_{\mathrm{o}}(t_k) = B_{\mathrm{o}}(p_k, q^{-1})u(t_k) \\ y(t_k) = \chi_{\mathrm{o}}(t_k) + v_{\mathrm{o}}(t_k) \end{cases} \tag{3.6}$$

where p_k is the value of the scheduling parameter p at sample time t_k, χ_{o} is the noise-free output, u is the input, v_{o} is the additive noise with bounded spectral density, y is the noisy output of the system and q is the time-shift operator, *i.e.* $q^{-i}u(t_k) = u(t_{k-i})$. $A_{\mathrm{o}}(p_k, q^{-1})$ and $B_{\mathrm{o}}(p_k, q^{-1})$ are polynomials in q^{-1} of degree n_{a} and n_{b} respectively :

$$A_{\mathrm{o}}(p_k, q^{-1}) = 1 + \sum_{i=1}^{n_{\mathrm{a}}} a_i^{\mathrm{o}}(p_k)q^{-i}, \tag{3.7a}$$

$$B_{\mathrm{o}}(p_k, q^{-1}) = \sum_{j=0}^{n_{\mathrm{b}}} b_j^{\mathrm{o}}(p_k)q^{-j}, \tag{3.7b}$$

where the coefficients a_i and b_j are real meromorphic functions ($f : \mathbb{R}^n \mapsto \mathbb{R}$ is a real meromorphic function if $f = g/h$ with g, h analytic and $h \neq 0$) with static dependence on p. It is assumed that these coefficients are non-singular on $\mathbb{P}$, thus the solutions of $\mathcal{S}_{\mathrm{o}}$ are well-defined and the process part is completely characterized by the coefficient functions $\{a_i^{\mathrm{o}}\}_{i=1}^{n_{\mathrm{a}}}$ and $\{b_j^{\mathrm{o}}\}_{j=0}^{n_{\mathrm{b}}}$.

Most of existing methods in the literature assume an ARX type of data generating system, which means that the noise process v_o can be written as

$$e_\mathrm{o}(t_k) = A_\mathrm{o}(p_k, q^{-1})v_\mathrm{o}(t_k), \tag{3.8}$$

where e_o is a zero-mean, discrete-time white noise process with a normal distribution $\mathcal{N}(0, \sigma_\mathrm{o}^2)$, σ_o^2 being the variance. Even if in some specific applications, the dependence of the noise on p can be considered as a fair assumption, the structure of (3.8) is often found unrealistic as it assumes that both the noise and the process part of $\mathcal{S}_\mathrm{o}$ contain the same dynamics. Again, a more general case is considered where the colored noise associated with the sampled output measurement $y(t_k)$ is assumed to have a rational spectral density which might has no relation to the actual process dynamics of $\mathcal{S}_\mathrm{o}$. As a preliminary step towards the case of a p-dependent noise, it is also assumed that this rational spectral density is not dependent on p : this corresponds to a more realistic assumption than (3.8), especially in case of measurement noise. Therefore, v_o is represented by a discrete-time *autoregressive moving average* (ARMA) model :

$$v_\mathrm{o}(t_k) = H_\mathrm{o}(q)e_\mathrm{o}(t_k) = \frac{C_\mathrm{o}(q^{-1})}{D_\mathrm{o}(q^{-1})}e_\mathrm{o}(t_k), \tag{3.9}$$

where $C_\mathrm{o}(q^{-1})$ and $D_\mathrm{o}(q^{-1})$ are monic polynomials with constant coefficients and with respective degree n_c and n_d. Furthermore, all roots of $z^{n_\mathrm{d}}D_\mathrm{o}(z^{-1})$ and $z^{n_\mathrm{c}}C_\mathrm{o}(z^{-1})$ are inside the unit disc. It can be noticed that in case $C_\mathrm{o}(q^{-1}) = D_\mathrm{o}(q^{-1}) = 1$, (3.9) defines an OE model.

3.2.2 Model considered

Next we introduce a discrete-time LPV Box-Jenkins (BJ) type of model structure that we propose for the identification of the data-generating system (3.6) with noise model (3.9). In the proposed model structure, the noise model and the process model are parameterized separately.

3.2.2.1 Process model

The process model is denoted by $\mathcal{G}_\rho$ and defined in an LPV-IO representation form :

$$\mathcal{G}_\rho : \big(A(p_k, q^{-1}, \rho), B(p_k, q^{-1}, \rho)\big) = (\mathcal{A}_\rho, \mathcal{B}_\rho) \tag{3.10}$$

where the p-dependent polynomials A and B are parameterized as

$$\mathcal{A}_\rho \quad \begin{cases} A(p_k, q^{-1}, \rho) = 1 + \displaystyle\sum_{i=1}^{n_\mathrm{a}} a_i(p_k)q^{-i}, \\ a_i(p_k) = a_{i,0} + \displaystyle\sum_{l=1}^{n_\alpha} a_{i,l}f_l(p_k) \quad i = 1, \ldots, n_\mathrm{a} \end{cases}$$

$$\mathcal{B}_\rho \quad \begin{cases} B(p_k, q^{-1}, \rho) = \displaystyle\sum_{j=0}^{n_\mathrm{b}} b_j(p_k)q^{-j}, \\ b_j(p_k) = b_{j,0} + \displaystyle\sum_{l=1}^{n_\beta} b_{j,l}g_l(p_k) \quad j = 0, \ldots, n_\mathrm{b} \end{cases}$$

In this parameterization, $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are meromorphic functions of p, with static dependence, allowing the identifiability of the model (pairwise orthogonal functions on $\mathbb{P}$ for example). The associated model parameters ρ are stacked columnwise :

$$\rho = \begin{bmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_{n_\mathrm{a}} & \mathtt{b}_0 & \dots & \mathtt{b}_{n_\mathrm{b}} \end{bmatrix}^\top \in \mathbb{R}^{n_\rho}, \tag{3.12}$$

where

$$\begin{aligned} \mathtt{a}_i &= \begin{bmatrix} a_{i,0} & a_{i,1} & \dots & a_{i,n_\alpha} \end{bmatrix} \in \mathbb{R}^{n_\alpha+1} \\ \mathtt{b}_j &= \begin{bmatrix} b_{j,0} & b_{j,1} & \dots & b_{j,n_\beta} \end{bmatrix} \in \mathbb{R}^{n_\beta+1} \end{aligned}$$

and $n_\rho = n_\mathrm{a}(n_\alpha+1) + (n_\mathrm{b}+1)(n_\beta+1)$. Introduce also $\mathcal{G} = \{\mathcal{G}_\rho \mid \rho \in \mathbb{R}^{n_\rho}\}$, as the collection of all process models in the form of (3.10).

3.2.2.2 Noise model

The noise model is denoted by $\mathcal{H}$ and defined as an LTI transfer function :

$$\mathcal{H}_\eta : (H(q,\eta)) \tag{3.13}$$

where H is a monic rational function given in the form of

$$H(q,\eta) = \frac{C(q^{-1},\eta)}{D(q^{-1},\eta)} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_\mathrm{c}} q^{-n_\mathrm{c}}}{1 + d_1 q^{-1} + \ldots + d_{n_\mathrm{d}} q^{-n_\mathrm{d}}}. \tag{3.14}$$

The associated model parameters η are stacked columnwise in the parameter vector,

$$\eta = \begin{bmatrix} c_1 & \dots & c_{n_\mathrm{c}} & d_1 & \dots & d_{n_\mathrm{d}} \end{bmatrix}^\top \in \mathbb{R}^{n_\eta}, \tag{3.15}$$

where $n_\eta = n_\mathrm{c} + n_\mathrm{d}$. Additionally, denote $\mathcal{H} = \{\mathcal{H}_\eta \mid \eta \in \mathbb{R}^{n_\eta}\}$, the collection of all noise models in the form of (3.13).

3.2.2.3 Whole model

With respect to a given process and noise part $(\mathcal{G}_\rho, \mathcal{H}_\eta)$, the parameters can be collected as $\theta = \begin{bmatrix} \rho^\top & \eta^\top \end{bmatrix}$ and the signal relations of the LPV-BJ model, denoted in the sequel as $\mathcal{M}_\theta$, are defined as :

$$\mathcal{M}_\theta \begin{cases} A(p_k, q^{-1}, \rho)\chi(t_k) = B(p_k, q^{-1}, \rho)u(t_k) \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{3.16}$$

Based on this model structure, the model set, denoted as $\mathcal{M}$, with process $(\mathcal{G}_\rho)$ and noise $(\mathcal{H}_\eta)$ models parameterized independently, takes the form

$$\mathcal{M} = \left\{ (\mathcal{G}_\rho, \mathcal{H}_\eta) \;\mid\; \mathrm{col}(\rho,\eta) = \theta \in \mathbb{R}^{n_\rho + n_\eta} \right\}. \tag{3.17}$$

This set corresponds to the set of candidate models in which we seek the model that explains data gathered from $\mathcal{S}_\mathrm{o}$ the best, under a given identification criterion (cost function).

3.2.3 Predictors and prediction error

Similar to the LTI case, in the LPV prediction error framework, one is concerned about finding a model in a given LPV model structure $\mathcal{M}$, which minimizes the statistical mean of the squared prediction error based on past samples of (y, u, p). However in the LPV case, no transfer function representation of systems is available. Furthermore, multiplication with q is not commutative over the p-dependent coefficients, meaning that :

$$\begin{aligned} q^{-1}B(p_k, q^{-1})u(t_k) &\neq B(p_k, q^{-1})q^{-1}u(t_k) \\ &\text{as} \\ q^{-1}B(p_k, q^{-1})u(t_k) &= B(p_{k-1}, q^{-1})u(t_{k-1}) \\ B(p_k, q^{-1})q^{-1}u(t_k) &= B(p_k, q^{-1})u(t_{k-1}). \end{aligned} \tag{3.18}$$

Therefore to define predictors with respect to models $\mathcal{M}_\theta \in \mathcal{M}$, a convolution type representation of the system dynamics, *i.e.* an LPV *Impulse Response Representation* (IRR), is used where the coefficients have dynamic dependence on p (dependence on future and past samples of p) [Tót8]. This means that $\mathcal{S}_0$ with an asymptotically stable process and noise part is written as

$$y(t_k) = \underbrace{(G_\mathrm{o}(q) \diamond p)(t_k)u(t_k)}_{\chi_\mathrm{o}(t_k)} + \underbrace{(H_\mathrm{o}(q) \diamond p)(t_k)e_\mathrm{o}(t_k)}_{v_\mathrm{o}(t_k)} \tag{3.19}$$

where

$$(G_\mathrm{o}(q) \diamond p)(t_k) = \sum_{i=0}^{\infty} (\alpha_i^\mathrm{o} \diamond p)(t_k) q^{-i}, \tag{3.20a}$$

$$(H_\mathrm{o}(q) \diamond p)(t_k) = 1 + \sum_{i=1}^{\infty} (\beta_i^\mathrm{o} \diamond p)(t_k) q^{-i}, \tag{3.20b}$$

with $\alpha_i^\mathrm{o} \diamond p$ expressing dynamic dependence of α_i on p, *i.e.* $\alpha_i^\mathrm{o} \diamond p = \alpha_i(p, qp, q^{-1}p, q^2p, \ldots)$. Now if p is deterministic and there exits a convergent adjoint $H_\mathrm{o}^\dagger$ of H_o such that

$$e_\mathrm{o}(t_k) = (H_\mathrm{o}^\dagger(q) \diamond p)(t_k) v_\mathrm{o}(t_k), \tag{3.21}$$

then it is possible to show (see [Tót10]) that the *one-step ahead predictor* of y is

$$y(t_k \mid t_{k-1}) = \Big((H_\mathrm{o}^\dagger(q) G_\mathrm{o}(q)) \diamond p\Big)(t_k)\ u(t_k) + \Big((1 - H_\mathrm{o}^\dagger(q)) \diamond p\Big)(t_k)\ y(t_k). \tag{3.22}$$

Example

As an example, consider an LPV-IO representation given in the following discrete time filter form :

$$y = -0.1pq^{-1}y - 0.2q^{-2}y + \sin(p)q^{-1}u,$$

with $\mathbb{P} = [0, 1]$. *By recursive substitution of this equation*

$$\begin{cases} q^{-1}y = -0.1q^{-1}pq^{-2}y - 0.2q^{-3}y + \sin(q^{-1}p)q^{-2}u, \\ q^{-2}y = -0.1q^{-1}pq^{-3}y - 0.2q^{-4}y + \sin(q^{-2}p)q^{-3}u, \\ \vdots \end{cases} \tag{3.23}$$

the following series-expansion in terms of the pulse basis functions $\{q^{-1}, q^{-2}, \ldots\}$ *results in :*

$$y = \underbrace{\sin(p)}_{\alpha_1 \diamond p} q^{-1}u + \underbrace{\big(-0.1p\sin(q^{-1}p)\big)}_{\alpha_2 \diamond p} q^{-2}u + \underbrace{\big(0.01p(q^{-1}p) - 0.2\big)\sin(q^{-2}p)}_{\alpha_3 \diamond p} q^{-3}u \ldots \tag{3.24}$$

■

In case the noise model is not dependent on p, like in (3.9),

$$(H_{\mathrm{o}}(q) \diamond p)(t_k) = H_{\mathrm{o}}(q)(t_k) = \frac{C_{\mathrm{o}}(q^{-1})}{D_{\mathrm{o}}(q^{-1})}$$

and

$$(H_{\mathrm{o}}^{\dagger}(q) \diamond p)(t_k) = H_{\mathrm{o}}^{-1}(q) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})}.$$

With respect to a parameterized model structure, and using the IRR representation, by simply using the PEM framework as defined in [Lju99], we can define the *one-step ahead prediction error* as

$$\varepsilon_\theta(t_k) = y(t_k) - \hat{y}(t_k \mid t_{k-1}), \tag{3.25}$$

where

$$\hat{y}(t_k \mid t_{k-1}) = \Big((H^{\dagger}(q,\theta)G(q,\theta)) \diamond p\Big)(t_k)\ u(t_k) + \Big((1 - H^{\dagger}(q,\theta)) \diamond p\Big)(t_k)\ y(t_k) \tag{3.26}$$

with $G(q,\theta)$ and $H(q,\theta)$ the IRR's of the process and noise part respectively. Denote $\mathcal{D}_N = \{y(t_k), u(t_k), p(t_k)\}_{k=1}^{N}$ a data sequence of $\mathcal{S}_{\mathrm{o}}$. Then, to provide an estimate of θ based on the minimization of ε_θ, an identification criterion $W(\mathcal{D}_N, \theta)$ can be introduced, like the *least squares* criterion

$$W(\mathcal{D}_N, \theta) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon_\theta^2(t_k), \tag{3.27}$$

such that the parameter estimate is

$$\hat{\theta}_N = \arg \min_{\theta \in \mathbb{R}^{n_\rho + n_\eta}} W(\mathcal{D}_N, \theta). \tag{3.28}$$

3.2.4 Persistency of excitation for LPV models

In order to estimate an adequate model in a given model set, most PEM algorithms like least squares or instrumental variable methods require that a *persistency of excitation* condition with respect to $\mathcal{D}_N$ collected from the system is satisfied. Such a condition is required to guarantee consistency and convergence of the algorithm providing estimates. However, in the LPV case it turns out that persistency of excitation in terms of (u, p) for a given order, as it is understood in the LTI case (see [Lju99]), does not guarantee consistency of the estimated parameters. The reason is that even if identifiability of the given parametrization is satisfied under the considered identification criterion, statistically global minimum of the criterion function is not guaranteed with respect to such data. This means that the terminology of persistency of excitation with order n is ill-defined in the

LPV case. Instead, the informativity of the data sets (see [GBBM09]) with respect to the assumed coefficient parametrization and model order is needed to be satisfied in order to ensure consistency of the estimates and the convergence of the estimation method. However, conditions of informative data sets have not been investigated directly in the LPV literature (for some preliminary work with conservative conditions see [BG02a, WD06a]). The question whether a data set is informative in the LPV case remains open. In terms of the upcoming analysis, it is assumed that the considered data sets satisfy this property. However in practice, the absence of a solid criterion restricts the user to the paradigm to excite as much as possible the system.

3.2.5 Identification problem statement

Based on the previous considerations, the identification problem addressed in the sequel can now be defined.

Problem 5 *Given a discrete time LPV data generating system $\mathcal{S}_\mathrm{o}$ defined as in (3.6) and a data set $\mathcal{D}_N$ collected from $\mathcal{S}_\mathrm{o}$. Based on the LPV-BJ model structure $\mathcal{M}_\theta$ defined by (3.16), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions :*

LPVA1 *$\mathcal{S}_\mathrm{o} \in \mathcal{M}$, i.e. there exits a θ_o defining a $\mathcal{G}_{\rho_\mathrm{o}} \in \mathcal{G}$ and a $\mathcal{H}_{\eta_\mathrm{o}} \in \mathcal{H}$ such that $(\mathcal{G}_{\rho_\mathrm{o}}, \mathcal{H}_{\eta_\mathrm{o}})$ is equal to $\mathcal{S}_\mathrm{o}$.*

LPVA2 *In the parametrization $\mathcal{A}_\rho$ and $\mathcal{B}_\rho$, $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are chosen such that $(\mathcal{G}_\mathrm{o}, \mathcal{H}_\mathrm{o})$ is identifiable for any trajectory of p.*

LPVA3 *$u(t_k)$ is not correlated to $e_\mathrm{o}(t_k)$.*

LPVA4 *Two different models inside $\mathcal{M}$ produce different prediction errors given $\mathcal{D}_N$.*

LPVA5 *$\mathcal{S}_\mathrm{o}$ is globally BIBO stable, i.e. for any trajectory of $p : \mathbb{R} \mapsto \mathbb{P}$ and any bounded input signal u, the output of $\mathcal{S}_\mathrm{o}$ is bounded [Tót10].*

3.2.6 On the use of linear regression framework and statistical optimality

LPV-IO parametric identification methods proposed in the literature so far are based on LS methods such as least squares or instrumental variable methods [BG02a, BKL08]. The currently accepted view in the literature is that if the system belongs to the model set defined in (3.17), then $y(t_k)$ can be written in the linear regression form :

$$y(t_k) = \varphi^\top(t_k)\rho + \tilde{v}(t_k) \tag{3.29}$$

with ρ as defined in (3.12) and

$$\begin{aligned}
\varphi(t_k) &= \Big[-y(t_{k-1}) \quad \dots \quad -y(t_{k-n_\mathrm{a}}) \quad -y(t_{k-1})f_1(p_k) \quad \dots \quad -y(t_{k-n_\mathrm{a}})f_{n_\alpha}(p_k) \\
&\qquad u(t_k) \quad \dots \quad u(t_{k-n_\mathrm{b}}) \quad u(t_k)g_1(p_k) \quad \dots \quad u(t_{k-n_\mathrm{b}})g_{n_\beta}(p_k) \Big]^\top && \text{(3.30a)} \\
\tilde{v}(t_k) &= A(p_k, q^{-1}, \rho)v(t_k). && \text{(3.30b)}
\end{aligned}$$

In this section, it is shown why such a linear regression cannot lead to statistically optimal (unbiased and minimal variance) estimates when the model structure is an LPV Box–Jenkins. Let us first

introduce the adjoint $A^\dagger$ of A, such that $\chi = A^\dagger(p_k, q^{-1}, \rho)u \Leftrightarrow A(p_k, q^{-1}, \rho)\chi = u$. Note that the adjoint always exits in a IRR sense with respect to an asymptotically stable A. In the LTI case, $A^\dagger = \frac{1}{A}$, however, in the LPV case, $A^\dagger \neq \frac{1}{A}$ due to the non-commutativity of the multiplication by q.

3.2.6.1 The conclusion brought in the literature

By considering (3.29) and the associated extended regressor in (3.30a), it is well known that the LS method leads to optimal estimate only if the noise model is ARX ($\tilde{v}(t_k)$ is a white noise). This condition implies that $v(t_k) = A^\dagger(p_k, q^{-1}, \rho)e(t_k)$ but is not fulfilled in many practical situations as v_o is often not directly related to the process itself and might not depend on p_k. Therefore it is proposed in [BKL08] to use an IV method where the instrument is built using the simulated data generated from an estimated auxiliary ARX model :

Algorithm 5 (One-step IV method (`OSIV`))

Step 1 *Estimate an ARX model by the LS method (minimizing* (3.27)*) using the extended regressor* (3.30a).

Step 2 *Generate an estimate* $\hat{\chi}(t_k)$ *of* $\chi(t_k)$ *based on the resulting ARX model of the previous step. Build an instrument based on* $\hat{\chi}(t_k)$ *and then estimate* ρ *using the IV method.*

As previously seen, instrumental variable methods produce unbiased estimates if the instrument is not correlated to the measurement noise. Based on the numerical simulation given in [BKL08], the following conclusions have been proposed :

– In case $\mathcal{S}_\mathrm{o}$ corresponds to an LPV-OE model ($v_\mathrm{o} = e_\mathrm{o}$), Algorithm 5 leads to an unbiased estimate.
– The variance of the estimated parameters is much larger than in a LS estimation as it is well-known.
– The estimation result can be improved if one uses a multi-step algorithm such as in [Lju99].

3.2.6.2 Existing methods and optimal estimates

The conclusions stated in [BKL08] can only be partially agreed to. It is true that the results can be improved and that the IV estimates are unbiased but this chapter will demonstrate that :

– Even by using multi-step algorithm of [Lju99] (designed for LTI systems), the optimal estimate cannot be reached with the linear regression form (3.29).
– For LPV-BJ models, estimates that are close to the statistically optimal solution can be reached by using IV methods and the variance of the estimated parameters is close to the variance of the LS estimator in given situations.

In the following part, it is shown why these statements hold true. In order to show why statistically optimal estimation of the model (3.16) cannot be reached under the viewpoint (3.29), it is first necessary to revisit the result of optimal prediction error in the LTI case.

3.2.6.3 The LPV case

Following the above introduced PEM approach in the LPV case (which is again maximum likelihood estimation because $e_{\mathrm{o}}(t_k) \in \mathcal{N}(0, \sigma_{\mathrm{o}}^2)$), the prediction error $\varepsilon_\theta(t_k)$ of (3.29) with respect to (3.16) is

$$\varepsilon_\theta(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)} A^\dagger(p_k, q^{-1}, \rho)\Big(A(p_k, q^{-1}, \rho)y(t_k) - B(p_k, q^{-1}, \rho)u(t_k)\Big) \tag{3.31}$$

where $D(q^{-1},\eta)/C(q^{-1},\eta)$ can be again recognized as the inverse of the ARMA(n_{c},n_{d}) noise model of (3.16). In contrast to the LTI case, the polynomial operators do not commute as it has been shown in Section 3.2.3. Hence, no filter can be chosen such that both conditions

$$\begin{aligned} A(p_k, q^{-1}, \rho)y_{\mathrm{f}}(t_k) &= \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)} A^\dagger(p_k, q^{-1}, \rho)A(p_k, q^{-1}, \rho)y(t_k) \\ B(p_k, q^{-1}, \rho)u_{\mathrm{f}}(t_k) &= \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)} A^\dagger(p_k, q^{-1}, \rho)B(p_k, q^{-1}, \rho)u(t_k) \end{aligned}$$

are fulfilled simultaneously. Consequently, no filtering of the data can lead to a regression equation

$$y_{\mathrm{f}}(t_k) = \varphi_{\mathrm{f}}^\top(t_k)\rho + \tilde{v}_{\mathrm{f}}(t_k) \tag{3.32}$$

which is equivalent to (3.29) and where $\tilde{v}_{\mathrm{f}}$ is white. In other words, by choosing φ such as in (3.30a) and therefore by assuming (3.29) (as in [BG02a] and [BKL08]) it is not possible to transform the estimation problem of (3.16) into a maximum likelihood estimation problem. The latter implies that no method proposed so far in the literature for solving the estimation of LPV-IO models or LTI-IO models can lead to optimal estimate in the LPV-BJ case by assuming the regression form (3.29). As a consequence, the existing theory needs to be modified in order to solve the identification problem stated in Section 3.2.5.

3.2.7 Reformulation of the model equations

In order to introduce a method which provides a solution to the identification problem of LPV-BJ models, rewrite the signal relations of (3.16) as

$$\mathcal{M}_\theta \begin{cases} \underbrace{\chi(t_k) + \sum_{i=1}^{n_{\mathrm{a}}} a_{i,0}\chi(t_{k-i})}_{F(q^{-1},\rho)\chi(t_k)} + \sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_\alpha} a_{i,l}\underbrace{f_l(p_k)\chi(t_{k-i})}_{\chi_{i,l}(t_k)} = \sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_\beta} b_{j,l}\underbrace{g_l(p_k)u(t_{k-j})}_{u_{j,l}(t_k)} \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)}e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{3.33}$$

where $F(q^{-1},\rho) = 1 + \sum_{i=1}^{n_{\mathrm{a}}} a_{i,0}q^{-i}$ and $g_0(t_k) = 1$. Note that in this way, the LPV-BJ model is rewritten as a *Multiple-Input Single-Output* (MISO) system with $(n_{\mathrm{b}}+1)(n_\beta+1) + n_{\mathrm{a}}n_\alpha$ inputs $\{\chi_{i,l}\}_{i=1,l=1}^{n_{\mathrm{a}},n_\alpha}$ and $\{u_{j,l}\}_{j=0,l=0}^{n_{\mathrm{b}},n_\beta}$ as represented in Fig. 3.2. Given the fact that the polynomial operator commutes in this representation ($F(q^{-1},\rho)$ does not depend on p_k), (3.33) can be rewritten as

$$y(t_k) = -\sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_\alpha} \frac{a_{i,l}}{F(q^{-1},\rho)}\chi_{i,l}(t_k) + \sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_\beta} \frac{b_{j,l}}{F(q^{-1},\rho)}u_{j,l}(t_k) + H(q)e(t_k), \tag{3.34}$$

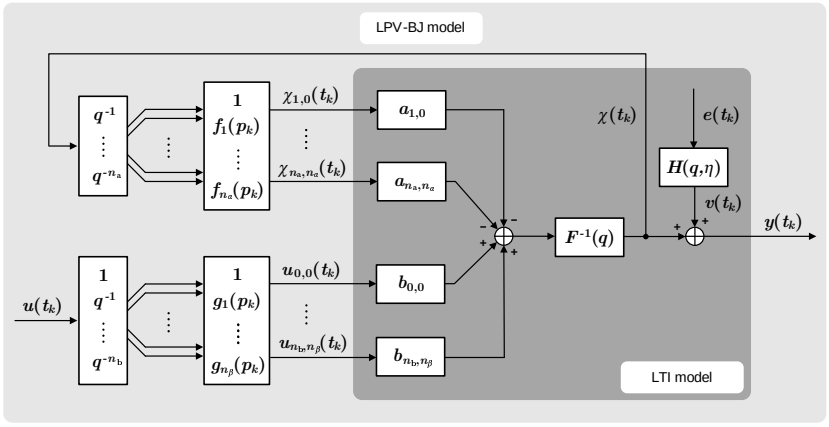


FIG. 3.2 – MISO LTI interpretation of the LPV-BJ model

which is an LTI representation. As (3.34) is an equivalent form of the model (3.16), and under the Assumption LPVA1, it holds that the data generating system $\mathcal{S}_o$ has also a MISO-LTI interpretation. It can be further noticed that unlike in the Hammerstein case, this model is not augmented and the number of parameters needed to describe the model remains minimal.

3.2.8 Optimal instrumental variable for LPV systems

Based on the MISO-LTI formulation (3.34), it becomes possible in theory to achieve optimal PEM using linear regression. This allows to extend the *Refined Instrumental Variable* (RIV) approach of the LTI identification framework to provide an efficient way of identifying LPV-BJ models.
Using (3.34), $y(t_k)$ can be written in the regression form :

$$y(t_k) = \varphi^\top(t_k)\rho + \tilde{v}(t_k) \tag{3.35}$$

where,

$$\begin{aligned}
\varphi(t_k) &= \left[\begin{array}{cccccccc} -y(t_{k-1}) & \dots & -y(t_{k-n_a}) & -\chi_{1,1}(t_k) & \dots & -\chi_{n_a,n_\alpha}(t_k) & u_{0,0}(t_k) & \dots \; u_{n_b,n_\beta}(t_k) \end{array} \right]^\top \\
\rho &= \left[\begin{array}{cccccccc} a_{1,0} & \dots & a_{n_a,0} & a_{1,1} & \dots & a_{n_a,n_\alpha} & b_{0,0} & \dots \; b_{n_b,n_\beta} \end{array} \right]^\top \\
\tilde{v}(t_k) &= F(q^{-1},\rho)v(t_k).
\end{aligned} \tag{3.36}$$

It is important to notice that (3.35) and (3.29) are not equivalent. The extended regressor in (3.35) contains the noise-free output terms $\{\chi_{i,l}\}$. Therefore, by momentary assuming that $\{\chi_{i,l}(t_k)\}_{i=1,l=0}^{n_a,n_\alpha}$ are known *a priori*, the conventional PEM approach on (3.35) leads to the prediction error $\varepsilon_\theta(t_k)$ given as :

$$\varepsilon_\theta(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)F(q^{-1},\rho)} \left(F(q^{-1},\rho)y(t_k) - \Big(-\sum_{i=1}^{n_a}\sum_{l=1}^{n_\alpha} a_{i,l}\chi_{i,l}(t_k) + \sum_{j=0}^{n_b}\sum_{l=0}^{n_b} b_{j,l}u_{j,l}(t_k) \Big) \right) \tag{3.37}$$

where $D(q^{-1},\eta)/C(q^{-1},\eta)$ can be recognized again as the inverse of the ARMA(n_c,n_d) noise model in (3.16). However, since the system written as in (3.34) is equivalent to a LTI system, the polynomial

operators commute and (3.37) can be considered in the alternative form

$$\varepsilon_\theta(t_k) = F(q^{-1},\rho)y_{\rm f}(t_k) - \sum_{i=1}^{n_{\rm a}}\sum_{l=1}^{n_\alpha} a_{i,l}\chi^{\rm f}_{i,l}(t_k) + \sum_{j=0}^{n_{\rm b}}\sum_{l=0}^{n_\beta} b_{j,l}u^{\rm f}_{j,l}(t_k) \tag{3.38}$$

where $y_{\rm f}(t_k)$, $u^{\rm f}_{k,j}(t_k)$ and $\chi^{\rm f}_{i,l}(t_k)$ represent the outputs of the prefiltering operation, using the filter (see [YGG08]) :

$$Q(q,\theta) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)F(q^{-1},\rho)}. \tag{3.39}$$

Based on (3.38), the associated linear-in-the-parameters model takes the form [YGG08] :

$$y_{\rm f}(t_k) = \varphi_{\rm f}^\top(t_k)\rho + \tilde{v}_{\rm f}(t_k), \tag{3.40}$$

where

$$\begin{aligned}\varphi_{\rm f}(t_k) &= \left[\begin{array}{cccccccc} -y_{\rm f}(t_{k-1}) & \dots & -y_{\rm f}(t_{k-n_{\rm a}}) & -\chi^{\rm f}_{1,1}(t_k) & \dots & -\chi^{\rm f}_{n_{\rm a},n_\alpha}(t_k) & u^{\rm f}_{0,0}(t_k) & \dots \quad u^{\rm f}_{n_{\rm b},n_\beta}(t_k)\end{array}\right]^\top \\ \tilde{v}_{\rm f}(t_k) &= F(q^{-1},\rho)v_{\rm f}(t_k).\end{aligned}$$

Consequently when the data is generated from $\mathcal{S}_{\rm o}$, if $\theta = \theta_{\rm o}$, then $\tilde{v}_{\rm f}(t_k) = e_{\rm o}(t_k)$ which is a white noise.

Therefore, according to the conditions for optimal estimates as given in Section 1.5.5, C4 defines the optimal filtered instrument as the noise-free version of the regressor :

$$\zeta^{\rm opt}(t_k) = \left[\begin{array}{cccccccc} -\chi_{\rm o}(t_{k-1}) & \dots & -\chi_{\rm o}(t_{k-n_{\rm a}}) & -\chi^{\rm o}_{1,1}(t_k) & \dots & -\chi^{\rm o}_{n_{\rm a},n_\alpha}(t_k) & u_{0,0}(t_k) & \dots \; u_{n_{\rm b},n_\beta}(t_k)\end{array}\right]^\top \tag{3.41}$$

while the optimal filter is given as :

$$L^{\rm opt}(q) = Q(q,\theta_{\rm o}) = \frac{D_{\rm o}(q^{-1})}{C_{\rm o}(q^{-1})F_{\rm o}(q^{-1})}. \tag{3.42}$$

Remarks on the use of the RIV approach

- Full column rank of $\bar{\mathbb{E}}\{L(q)\varphi(t_k)L(q)\varphi^\top(t_k)\}$ follows under Assumption LPVA4 [BG02a].
- In a practical situation none of $F(q^{-1},\rho)$, $C(q^{-1},\eta)$, $D(q^{-1},\eta)$ or $\{a_{i,l}(\rho)\}_{i=1,l=0}^{n_{\rm a},n_\alpha}$, $\{b_{j,l}(\rho)\}_{j=0,l=0}^{n_{\rm b},n_\beta}$ is known *a priori*. Therefore, the RIV estimation normally involves an iterative (or relaxation) algorithm in which, at each iteration, an 'auxiliary model' is used to generate the instrumental variables (which guarantees C2), as well as the associated prefilters. This auxiliary model is based on the parameter estimates obtained at the previous iteration. Consequently, if convergence occurs, C4 and C6 are fulfilled.
- Convergence of the iterative RIV algorithm has not been proved so far and is only empirically assumed [You08].
- The considered LPV model can be reformulated in a LTI-MISO form only under the condition that the noise-free output terms are *a priori* known (see Section 3.2.7). Therefore, even if the presented method considerably lowers the variance in the estimated parameters, the optimality cannot be guaranteed.

3.2.9 Iterative LPV-RIV algorithm for BJ models

Based on the previous considerations, the iterative scheme of the RIV algorithm dedicated to the LPV case as follows.

Algorithm 6 (LPV-RIV)

Step 1 *Generate an initial estimate of the process model parameter* $\hat{\rho}^{(0)}$ *(*e.g. *using the LS method). Set* $C(q^{-1}, \hat{\eta}^{(0)}) = D(q^{-1}, \hat{\eta}^{(0)}) = 1$. *Set* $\tau = 0$.

Step 2 *Compute an estimate of* $\chi(t_k)$ *via*

$$A(p_k, q^{-1}, \hat{\rho}^{(\tau)})\hat{\chi}(t_k) = B(p_k, q^{-1}, \hat{\rho}^{(\tau)})u(t_k),$$

where $\hat{\rho}^{(\tau)}$ *is the estimated obtained at the previous iteration. Based on* $\mathcal{M}_{\hat{\theta}^{(\tau)}}$, *deduce* $\{\hat{\chi}_{i,l}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$ *as given in* (3.33). *According to Assumption LPVA5, each* $\hat{\chi}_{i,l}$ *is bounded.*

Step 3 *Compute the filter as in* (3.39) *:*

$$L(q^{-1}, \hat{\theta}^{(\tau)}) = \frac{D(q^{-1}, \hat{\eta}^{(\tau)})}{C(q^{-1}, \hat{\eta}^{(\tau)})F(q^{-1}, \hat{\rho}^{(\tau)})}$$

and the associated filtered signals $\{u_{j,l}^{\mathrm{f}}(t_k)\}_{j=0,l=0}^{n_{\mathrm{b}},n_\beta}$, $y_{\mathrm{f}}(t_k)$ *and* $\{\chi_{i,l}^{\mathrm{f}}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$.

Step 4 *Build the filtered estimated regressor* $\hat{\varphi}_{\mathrm{f}}(t_k)$ *and in terms of C4 the filtered instrument* $\hat{\zeta}_{\mathrm{f}}(t_k)$ *as :*

$$\begin{aligned}
\hat{\varphi}_{\mathrm{f}}(t_k) = \Big[& -y_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -y_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad -\hat{\chi}_{1,1}^{\mathrm{f}}(t_k) \\
& \dots \quad -\hat{\chi}_{n_{\mathrm{a}},n_\alpha}^{\mathrm{f}}(t_k) \quad u_{0,0}^{\mathrm{f}}(t_k) \quad \dots \quad u_{n_{\mathrm{b}},n_\beta}^{\mathrm{f}}(t_k) \Big]^\top \\
\hat{\zeta}_{\mathrm{f}}(t_k) = \Big[& -\hat{\chi}_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -\hat{\chi}_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad -\hat{\chi}_{1,1}^{\mathrm{f}}(t_k) \\
& \dots \quad -\hat{\chi}_{n_{\mathrm{a}},n_\alpha}^{\mathrm{f}}(t_k) \quad u_{0,0}^{\mathrm{f}}(t_k) \quad \dots \quad u_{n_{\mathrm{b}},n_\beta}^{\mathrm{f}}(t_k) \Big]^\top
\end{aligned}$$

Step 5 *The IV optimization problem can now be stated in the form*

$$\hat{\rho}^{(\tau+1)}(N) = \arg\min_{\rho \in \mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\hat{\varphi}_{\mathrm{f}}^\top(t_k) \right] \rho - \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}(t_k) \right] \right\|^2 \tag{3.43}$$

where the solution is obtained as

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\hat{\varphi}_{\mathrm{f}}^\top(t_k) \right]^{-1} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}(t_k).$$

The resulting $\hat{\rho}^{(\tau+1)}(N)$ *is the IV estimate of the process model associated parameter vector at iteration* $\tau + 1$ *based on the prefiltered input/output data.*

Step 6 *An estimate of the noise signal* v *is obtained as*

$$\hat{v}(t_k) = y(t_k) - \hat{\chi}(t_k, \hat{\rho}^{(\tau)}). \tag{3.44}$$

Based on $\hat{v}$, *the estimation of the noise model parameter vector* $\hat{\eta}^{(\tau+1)}$ *follows, using in this case the* ARMA *estimation algorithm of the* MATLAB *identification toolbox (an IV approach can also be used for this purpose, see [You08]).*

Step 7 *If $\theta^{(\tau+1)}$ has converged or the maximum number of iterations is reached, then stop, else increase τ by 1 and go to Step 2.*

Moreover, after the convergence is complete, it is possible to compute the estimated parametric error covariance matrix $\hat{\mathbf{P}}_\rho$ from the expression (see Section 1.5.4) :

$$\hat{\mathbf{P}}_\rho = \hat{\sigma}_e^2 \left(\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) \hat{\zeta}_{\mathrm{f}}^\top(t_k) \right)^{-1} \tag{3.45}$$

where $\hat{\zeta}$ is the IV vector obtained at convergence and $\hat{\sigma}_e^2$ is the estimated residual variance.

3.2.10 `LPV-SRIV` algorithm for OE models

Based on a similar concept, the so-called *simplified* `LPV-RIV` (`LPV-SRIV`) method, can also be developed for the estimation of LPV-OE models. This method is based on a model structure (3.16) with $C(q^{-1},\eta) = D(q^{-1},\eta) = 1$ and consequently, Step 6 of Algorithm 6 can be skipped. Naturally, the `LPV-SRIV` does not minimize statistically optimal PEM for LPV-BJ models, however its performance and robustness is studied in Section 3.2.11.

3.2.11 Performance evaluation of the proposed `LPV-RIV` and `LPV-SRIV` algorithms

3.2.11.1 The presented method compared to existing methods

The system taken into consideration is inspired by the example in [BKL08] and is mathematically described as

$$\mathcal{S}_{\mathrm{o}} \begin{cases} A_{\mathrm{o}}(q,p_k) = 1 + a_1^{\mathrm{o}}(p_k)q^{-1} + a_2^{\mathrm{o}}(p_k)q^{-2} \\ B_{\mathrm{o}}(q,p_k) = b_0^{\mathrm{o}}(p_k)q^{-1} + b_1^{\mathrm{o}}(p_k)q^{-2} \\ H_{\mathrm{o}}(q) = \dfrac{1}{1 - q^{-1} + 0.2q^{-2}} \end{cases} \tag{3.46}$$

where $v(t_k) = H_{\mathrm{o}}(q)e(t_k)$ and

$$a_1^{\mathrm{o}}(p_k) = 1 - 0.5p_k - 0.1p_k^2, \tag{3.47a}$$
$$a_2^{\mathrm{o}}(p_k) = 0.5 - 0.7p_k - 0.1p_k^2, \tag{3.47b}$$
$$b_0^{\mathrm{o}}(p_k) = 0.5 - 0.4p_k + 0.01p_k^2, \tag{3.47c}$$
$$b_1^{\mathrm{o}}(p_k) = 0.2 - 0.3p_k - 0.02p_k^2. \tag{3.47d}$$

In the upcoming examples, the scheduling signal p is considered as a periodic function of time : $p_k = 0.5\sin(0.35\pi k) + 0.5$. The input $u(t_k)$ is taken as a white noise with a uniform distribution $\mathcal{U}(-1,1)$ and with length $N = 4000$ to generate data sets $\mathcal{D}_N$ of $\mathcal{S}_{\mathrm{o}}$.
In the sequel, the *One Step Instrumental Variable* (`OSIV`) method presented in [BKL08] and the conventional *Least Square* (`LS`) method dedicated to LPV models such as the one used in [BG02a] are compared to the proposed RIV-based approaches. Both methods assume the following model structure :

$$\mathcal{M}_\theta^{\mathtt{LS,OSIV}} \begin{cases} A(p_k,q^{-1},\rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k,q^{-1},\rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k,q,\rho) = A^\dagger(p_k,q^{-1},\rho) \end{cases}$$

where

$$a_1(p_k) = a_{1,0} + a_{1,1}p_k + a_{1,2}p_k^2 \tag{3.48a}$$
$$a_2(p_k) = a_{2,0} + a_{2,1}p_k + a_{2,2}p_k^2 \tag{3.48b}$$
$$b_0(p_k) = b_{0,0} + b_{0,1}p_k + b_{0,2}p_k^2 \tag{3.48c}$$
$$b_1(p_k) = b_{1,0} + b_{1,1}p_k + b_{1,2}p_k^2 \tag{3.48d}$$

In contrast with these model structures, the proposed LPV *Refined Instrumental Variable* method (`LPV-RIV`) represents the situation $\mathcal{S}_o \in \mathcal{M}$ and assumes the following LPV-BJ model :

$$\mathcal{M}_\theta^{\texttt{LPV-RIV}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = \dfrac{1}{1 + d_1 q^{-1} + d_2 q^{-2}} \end{cases}$$

with $a_1(p_k)$, $a_2(p_k)$, $b_0(p_k)$, $b_1(p_k)$ as given in (3.48a-d), while the LPV *Simplified Refined Instrumental Variable* method (`LPV-SRIV`) represents the case when $\mathcal{G}_o \in \mathcal{G}$, $\mathcal{H}_o \notin \mathcal{H}$ and assumes the following LPV-OE model :

$$\mathcal{M}_\theta^{\texttt{LPV-SRIV}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = 1 \end{cases}$$

The robustness of the proposed and existing algorithms are investigated with respect to different *signal-to-noise ratios* $\text{SNR} = 10\log\frac{P_{\chi_o}}{P_{e_o}}$, where P_{χ_o} and P_{e_o} are the average power of signals χ_o and e_o respectively. To provide representative results, a Monte Carlo simulation is realized based on $N_{\text{run}} = 100$ random realization, with the Gaussian white noise input to the ARMA noise model being selected randomly for each realization at different noise levels : 15dB, 10dB, 5dB and 0dB. An example of noisy signal at 0dB is shown in Figure 3.3. For the Monte Carlo simulation at SNR = 15dB, Tables 3.1 and 3.2 show the detailed results about mean and standard deviation of the estimated parameters. In some practical application, only one realization is accessible and therefore it is not possible to compute the uncertainty through MCS. In this latter case it is important to be able to determine the *standard error* (SE) on the estimated parameters with a *single realization* (SR). Therefore the results of SR are also given in these Tables. Note that it is possible to compute the SR standard error $\text{SE} = \text{diag}(\hat{\mathbf{P}}_\rho)^{1/2}$ from the covariance matrix $\hat{\mathbf{P}}_\rho$ as defined in (3.45).
With respect to the considered methods, Table 3.3 shows the norm of the bias (BN) and variance (VN) of the estimated parameter vector with

$$\begin{cases} \text{BN} = ||\rho_o - \bar{\mathbb{E}}(\hat{\rho})||_2, \\ \text{VN} = ||\bar{\mathbb{E}}(\hat{\rho} - \bar{\mathbb{E}}(\hat{\rho}))||_2, \end{cases} \tag{3.49}$$

where $\bar{\mathbb{E}}$ is the mean operator over the Monte Carlo simulation and $||.||_2$ is the $\mathcal{L}_2$ norm. The table also displays the mean number of iterations (Nit) the algorithms needed to converge to the estimated parameter vector.

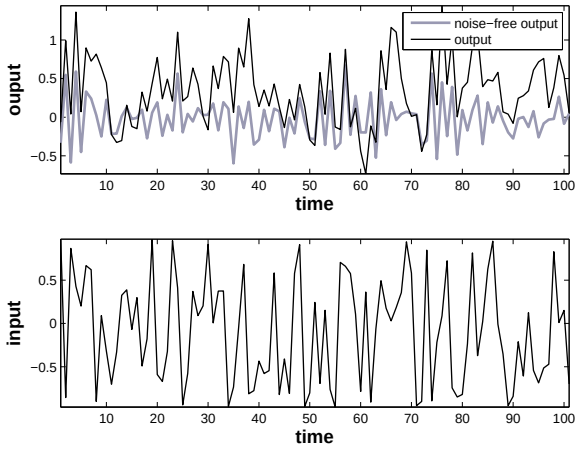


FIG. 3.3 – I/O signals at SNR = 0dB

It can be seen from Table 3.3 that the IV methods are unbiased according to the theoretical results. It might not appear clearly for the `OSIV` method when using SNR under 10dB but considering the variances induced, the bias is only due to the relatively low number of simulation runs. Under 10dB, the results of the `OSIV` cannot be considered as relevant as they induce such large variances. In the present BJ system, the `OSIV` method does not lead to satisfying results and cannot be used in practical applications. It can be seen that for SNR down to 5dB, the `LPV-RIV` method produces variance in the estimated parameters which are very close to the one obtained with the `LS` method, not mentioning that the bias has been completely suppressed. Although the statistical optimality of the algorithm cannot be proved, this latter result shows on this example, that the `LPV-RIV` algorithm dramatically improves the accuracy of the estimates. The suboptimal `LPV-SRIV` method offers satisfying results, considering that the noise model is not correctly assumed. The variance in the estimated parameters is twice as much as for the `LPV-RIV` estimates but it remains close to the variance of the `LS` estimates. Finally, it can be pointed out that the number of iterations is high in comparison to the linear case for `RIV` methods (typically, 4 iterations are needed in a second order linear case). Tables 3.1 and 3.2 show that detailed results lead to the same conclusion as when looking at Table 3.3. It can be finally seen from 3.2 that the `LPV-RIV` method estimates accurately the noise model and that the standard error obtained from a single realization is well correlated to the standard deviation obtained through Monte Carlo simulation.

3.2.11.2 Robustness to modelling error

All the tests realized in the previous section has been achieved at least in the case when $\mathcal{G}_o \in \mathcal{G}$. This assumption is fair for linear systems as most processes can be derived from first principle laws. In the nonlinear models however, the nonlinearity and therefore the scheduling parameter dependence is often only an approximation of the real function. Therefore, it is relevant to investigate the case when $\mathcal{G}_o \notin \mathcal{G}$. However, it is obviously not possible in this case to compare the model parameters

TAB. 3.1 – Mean and standard deviation of the estimated A polynomial parameters at SNR = 15dB

		$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$
method	true value	1	-0.5	-0.1	0.5	-0.7	-0.1
LS	mean	-0.3794	2.2373	-2.0584	-0.1085	-0.0755	-0.4786
	std	0.0219	0.0663	0.0591	0.0125	0.0600	0.0558
OSIV	mean	1.0259	-0.6161	0.0205	0.5092	-0.7510	-0.0377
	std	0.3023	1.0330	0.8605	0.1227	0.4348	0.3986
LPV-SRIV MCS	mean	1.0003	-0.5013	-0.0971	0.5007	-0.7047	-0.0943
	std	0.0313	0.1022	0.0893	0.0106	0.0650	0.074
LPV-SRIV SR	$\hat{\rho}$	0.9801	-0.3743	-0.2120	0.4978	-0.7154	-0.0736
	SE	0.0377	0.1567	0.1486	0.0171	0.1010	0.1099
LPV-RIV MCS	mean	0.9999	-0.5020	-0.0989	0.5005	-0.7050	-0.0962
	std	0.0170	0.0589	0.0610	0.0084	0.0488	0.0523
LPV-RIV SR	$\hat{\rho}$	0.9947	-0.5053	-0.0506	0.4981	-0.7303	-0.0350
	SE	0.0120	0.0479	0.0435	0.0050	0.0330	0.0368

TAB. 3.2 – Mean and standard deviation of the estimated B and D polynomial parameters at SNR = 15dB

		$b_{0,0}$	$b_{0,1}$	$b_{0,2}$	$b_{1,0}$	$b_{1,2}$	$b_{2,2}$	d_1	d_2
method	true value	0.5	-0.4	0.01	0.2	-0.3	-0.02	-1	0.2
LS	mean	0.5043	-0.4045	0.0085	-0.3201	0.7890	-0.7335	X	X
	std	0.0039	0.0233	0.0219	0.0097	0.0284	0.0238	X	X
OSIV	mean	0.4986	-0.3991	0.0110	0.2096	-0.3409	0.0181	X	X
	std	0.0115	0.0564	0.0503	0.1151	0.3731	0.2922	X	X
LPV-SRIV MCS	mean	0.4996	-0.3998	0.0101	0.1997	-0.3004	-0.0190	X	X
	std	0.0038	0.0183	0.0171	0.0104	0.0367	0.0300	X	X
LPV-SRIV SR	$\hat{\rho}$	0.4998	-0.3783	-0.0108	0.1996	-0.2885	-0.0273	X	X
	SE	0.0044	0.0221	0.0216	0.0138	0.0561	0.0493	X	X
LPV-RIV MCS	mean	0.4998	-0.3993	0.0092	0.1998	-0.3008	-0.0194	-1.003	0.2042
	std	0.0020	0.0106	0.0106	0.0055	0.0228	0.0214	0.0171	0.0172
LPV-RIV SR	$\hat{\rho}$	0.5020	-0.4142	0.0245	0.1971	-0.2889	-0.0219	X	X
	SE	0.0016	0.0075	0.0072	0.0042	0.0165	0.0141	X	X

TAB. 3.3 – Estimator bias and variance norm at different SNR

Method		15dB	10dB	5dB	0dB
LS	BN	2.9107	3.2897	3.0007	2.8050
	VN	0.0074	0.0151	0.0215	0.0326
OSIV	BN	0.1961	1.8265	6.9337	10.85
	VN	1.3353	179.42	590.78	11782
LPV -SRIV	BN	0.0072	0.0426	0.1775	0.2988
	VN	0.0149	0.0537	0.4425	0.4781
	Nit	22	22	25	30
LPV -RIV	BN	0.0068	0.0184	0.0408	0.1649
	VN	0.0063	0.0219	0.0696	0.2214
	Nit	31	30	30	32

to the true system parameters. Consequently, the fitness score [Lju09] is exposed

$$\text{fit} = 100\% \cdot \left(1 - \frac{||\hat{\chi}(t_k) - \chi_o(t_k)||_2}{||\chi_o(t_k) - \bar{\mathbb{E}}(\chi_o(t_k))||_2}\right) \tag{3.50}$$

between the noise-free output $\chi_o(t)$ and the output $\hat{\chi}(t)$ simulated using the identified model. The fitness score is 100% if the simulated output coincides exactly to the noise-free output. If the score equals 0% the estimated output fits the noise-free output as good as its mean value. A score under -100% is set to -100. The minimum, maximum and mean value of the fitness score for different approaches are displayed in Table 3.4 (SNR = 10dB) and Table 3.5 (SNR = 0dB) as the result of a Monte Carlo simulation in which $N_{\text{run}} = 20$. The true model for the presented system is $n_a = 2$, $n_b = 2$, $n_c = 0$, $n_d = 2$, $n_\alpha = 2$ and $n_\beta = 2$ denoted as $[n_a\ n_b\ n_c\ n_d\ n_\alpha\ n_\beta] = [220222]$. For example the model [210101] corresponds to

$$\mathcal{M}_{\texttt{210101}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2}, \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1}, \\ H(p_k, q, \eta) = \dfrac{1}{1 + d_1 q^{-1}}, \end{cases}$$

with

$$a_1(p_k) = a_{1,0}, \quad a_2(p_k) = a_{2,0}, \quad b_0(p_k) = b_{0,0} + b_{0,1}p_k.$$

Table 3.4 and Table 3.5 show that the LS method fails to correctly fit the output, especially at SNR = 0dB, where the LS method gives very poor fitting score. Table 3.4 shows that at 10dB, even if the OSIV estimator is not efficient, it can compete with the LPV-RIV estimator in some runs but the values of the fitting score are more sparse. Table 3.5 shows that the OSIV estimator can no longer be used reliably for low SNR. It can be moreover noticed that for all simulations, the best fitting score of the OSIV estimated model is less than the average fitting score of the LPV-RIV estimated model. The LPV-RIV estimator seems to be the less affected by strong noise and is reliable : for strong modelling error (model [110011]) and at SNR = 0dB, Figure 3.4 shows that in average, the estimated model output fits reasonably well the noise-free estimation data.

TAB. 3.4 – Modelling Error Monte Carlo simulation with SNR = 10dB

Model parameters						LS			OSIV			LPV-RIV		
n_a	n_b	n_c	n_d	n_α	n_β	min	max	mean	min	max	mean	min	max	mean
2	2	0	2	2	2	60.2	63.6	61.8	19.1	95.5	80.9	98.3	99.2	98.7
2	2	0	0	1	1	61.2	63.9	62.7	-1.8	97.2	86.4	72.5	99.2	97.2
1	1	0	2	2	2	47.8	52.0	50.4	74.9	78.1	76.5	79.3	79.7	79.5
1	1	0	0	1	1	47.5	52.6	50.1	71.5	76.9	74.8	78.3	78.6	78.5

TAB. 3.5 – Modelling Error Monte Carlo simulation with SNR = 0dB

Model parameters						LS			OSIV			LPV-RIV		
n_a	n_b	n_c	n_d	n_α	n_β	min	max	mean	min	max	mean	min	max	mean
2	2	0	2	2	2	28.3	35.8	33.0	-28.3	84.6	62.4	93.5	97.6	95.9
2	2	0	0	1	1	33.0	41.2	36.2	-100	83.5	34.4	48.6	97.8	87.4
1	1	0	2	2	2	1.6	6.8	4.5	-100	73.5	14.9	22.5	79.4	75.9
1	1	0	0	1	1	1.2	6.0	4.2	-100	75.2	28.2	42.9	79.2	71.9

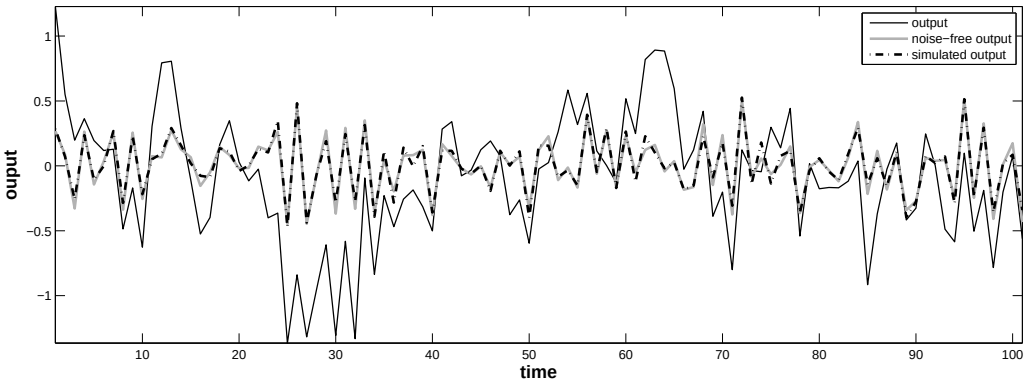


FIG. 3.4 – Noise-free output, noisy output and simulated output for model [110011] identified with LPV-RIV (SNR = 0dB and fit= 79%)

3.2.11.3 LPV noise model

The case of a noise model not depending on the scheduling variable is a particular case studied as a preliminary step with respect to a full-blown LPV-BJ model with p-dependent noise part. This extension is intended as a further step in the development of the `LPV-RIV` approach. The RIV-based methods should be unbiased as long as the instrument is not correlated to noise. Hence, due to the structure of the instrument (see step 4 of Algorithm 6), the method should be also unbiased with respect to p-dependent noise models. It will surely not be optimal, but should be consistent.
The example chosen is derived from the system (3.46) in which the noise ARMA filter is turned into an LPV model and the studied system is expressed by :

$$\mathcal{S}_{\mathrm{o}} \begin{cases} A_{\mathrm{o}}(q, p_k) = 1 + a_1^{\mathrm{o}}(p_k)q^{-1} + a_2^{\mathrm{o}}(p_k)q^{-2}, \\ B_{\mathrm{o}}(q, p_k) = b_0^{\mathrm{o}}(p_k)q^{-1} + b_1^{\mathrm{o}}(p_k)q^{-2}, \\ D_{\mathrm{o}}(q^{-1}, p_k) = 1 + d_1^{\mathrm{o}}(p_k)q^{-1} + d_2^{\mathrm{o}}(p_k)q^{-2}, \end{cases} \tag{3.51}$$

where $D_{\mathrm{o}}(q^{-1}, p_k)v(t_k) = e(t_k)$ and

$$a_1^{\mathrm{o}}(p_k) = 1 - 0.5p_k - 0.1p_k^2, \tag{3.52a}$$
$$a_2^{\mathrm{o}}(p_k) = 0.5 - 0.7p_k - 0.1p_k^2, \tag{3.52b}$$
$$b_0^{\mathrm{o}}(p_k) = 0.5 - 0.4p_k + 0.01p_k^2, \tag{3.52c}$$
$$b_1^{\mathrm{o}}(p_k) = 0.2 - 0.3p_k - 0.02p_k^2, \tag{3.52d}$$
$$d_1^{\mathrm{o}}(p_k) = -1 + 0.2p_k - 0.1p_k^2, \tag{3.52e}$$
$$d_2^{\mathrm{o}}(p_k) = 0.2 - 0.1p_k + 0.05p_k^2. \tag{3.52f}$$

To illustrate the above comments, the results of a Monte Carlo simulation are exposed in Table 3.6. The methods compared are the same methods (and their respective associated models) as the ones exposed in Section 3.2.11.1. The MCs results are based on $N_{\mathrm{run}} = 100$ using a $SNR = 10$dB and a number of samples $N = 4000$. For comparison purpose, the bias and variance norms of the resulting estimates are given in Table 3.6.

Tab. 3.6 – Results for a LPV noise model

Method	LS	OSIV	LPV-SRIV	LPV-RIV
Bias Norm	3.4943	4.2293	0.0335	0.0257
Variance Norm	0.0148	1377	0.0466	0.0318

From this example, it can be seen that the methods `LS`, `OSIV`, `LPV-SRIV` give the similar performance as in the example from Section 3.2.11.1. This can be easily explained as in both examples, the noise assumption is false. Consequently the bias of the `LS` method is identical, the `OSIV` method displays a very large parameter estimate variance, while the `LPV-SRIV` technique keeps a low variance and removes the bias.

The variance for the `LPV-RIV` parameter estimates is larger than in the p-independent noise case as the assumption $H_o \in \mathcal{H}$ is now violated. The variance is still lower than the one obtained with `LPV-SRIV`, as `LPV-RIV` assumes the proper model for the nominal part of the noise (which is the same as in (3.46)). In practice, when the nominal part of the noise model is unknown, it might be better to use `LPV-SRIV` as it uses a more parsimonious model.
Nonetheless, it has to be mentioned that the case of a p-independent noise case such as considered for the method development is still relevant to practical applications. For example, in the case of high-tech applications where the major source of noise is due to measurement error (*e.g.* wafer-scanners), the rational spectral density of the noise which is not dependent on p often holds true.

3.3 Discrete-time closed-loop LPV model identification

Closed-loop LPV-IO parametric identification methods proposed in the literature so far are direct extensions of LTI methods such as CLOE algorithm [BW08] [LK97] or basic instrumental variable [AW09][BKL08]. The solution presented in [AW09] deals with basic IV without variance analysis. In Section 3.2.6, the impossibility to reach efficient estimates using only delayed signals of u and y in the regressor was highlighted. This fact holds in the closed-loop case. The extension of CLOE algorithm presented in [BW08] estimates recursively the noise-free regressor to overcome this problem. It is however designed for Ouput Error (OE) models. In other word, no existing method in the LPV literature can theoretically deal in a statistical optimal way with the identification of closed-loop LPV-BJ data-generating systems.

3.3.1 System description

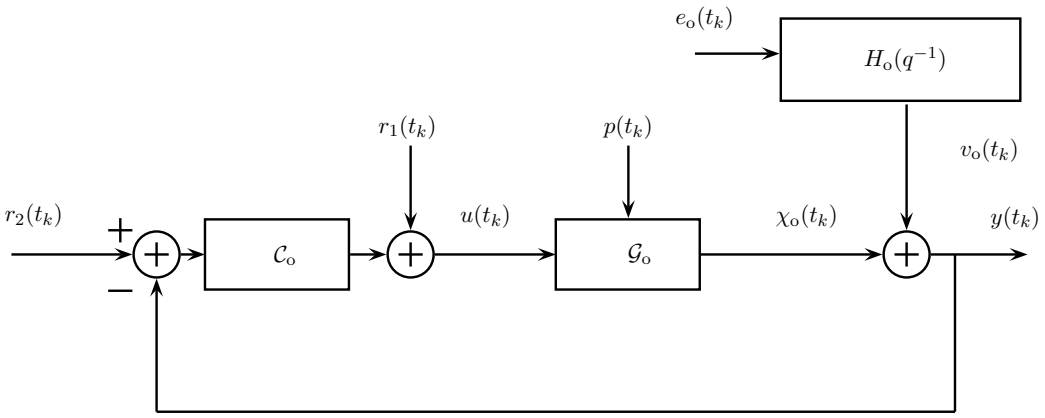


FIG. 3.5 – Closed-loop LPV system

Consider the data generating LPV system $\mathcal{S}_o$ from figure 3.5 defined by the following equation :

$$\mathcal{S}_o \begin{cases} A_o(p_k, q^{-1})\chi_o(t_k) = B_o(p_k, q^{-1})u(t_k) \\ y(t_k) = \chi_o(t_k) + v_o(t_k) \\ u(t_k) = r_1(t_k) + \mathcal{C}_o(r_2(t_k) - y(t_k)) \end{cases} \tag{3.53}$$

where p_k is the value of the scheduling parameter p at sample time t_k, χ_o is the process output, v_o is the additive noise with bounded spectral density, y is the system output and q is the time-shift operator. $r_1(t_k)$ can be a reference value, a setpoint or a noise disturbance on the regulator output and $r_2(t_k)$ can be a setpoint or a measurement noise on the output signal. It can be further noticed that both $r_1(t_k)$ and $r_2(t_k)$ are considered for theoretical developments. Nonetheless in practice, either one or the other is excited while the other remains null. $A_o(p_k, q^{-1})$ and $B_o(p_k, q^{-1})$ are polynomials in q^{-1} of degree n_a and n_b respectively :

$$A_o(p_k, q^{-1}) = 1 + \sum_{i=1}^{n_a} a_i^o(p_k) q^{-i}, \tag{3.54a}$$

$$B_o(p_k, q^{-1}) = \sum_{j=0}^{n_b} b_j^o(p_k) q^{-j}, \tag{3.54b}$$

where the coefficients a_i^o and b_j^o are real meromorphic functions ($f : \mathbb{R}^n \mapsto \mathbb{R}$ is a real meromorphic function if $f = g/h$ with g, h analytic and $h \neq 0$) with static dependence on p. It is assumed that these coefficients are non-singular on $\mathbb{P}$, thus the solutions of $\mathcal{S}_o$ are well-defined and the process part $\mathcal{G}_o$ is completely characterized by the coefficient functions $\{a_i^o\}_{i=1}^{n_a}$ and $\{b_j^o\}_{j=0}^{n_b}$. The noise v_o is represented by a discrete-time *autoregressive moving average* (ARMA) model :

$$v_o(t_k) = H_o(q) e_o(t_k) = \frac{C_o(q^{-1})}{D_o(q^{-1})} e_o(t_k), \tag{3.55}$$

where $C_o(q^{-1})$ and $D_o(q^{-1})$ are monic polynomials with constant coefficients and with respective degree n_c and n_d. Furthermore, all roots of $z^{n_d} D_o(z^{-1})$ and $z^{n_c} C_o(z^{-1})$ are inside the unit disc. It can be noticed that in case $C_o(q^{-1}) = D_o(q^{-1}) = 1$, (3.55) defines an OE noise model.

The method further presented is a direct identification method which can deal with any type of controller assuming that :

– The controller $\mathcal{C}_o$ is *a priori* known.
– The controller $\mathcal{C}_o$ insures the global BIBO stability of the closed-loop system $\mathcal{S}_o$.

Consequently, in the rest of this section, the model description focuses on the process of $\mathcal{S}_o$ to be identified.

3.3.2 Model considered

By using a direct estimation approach such as presented here (direct estimation of the process using $u(t_k)$ and $y(t_k)$), the modelling problem in the closed-loop framework is pretty similar to the open-loop case. Moreover, in the present LPV case, the exact same discussion about optimality as in Section 3.2.6.3 leads to the exact same need for the model reformulation such as derived in Section 3.2.7.

Consequently, with respect to the given process and noise part $(\mathcal{G}_\rho, \mathcal{H}_\eta)$ such as defined in Section 3.2.2, the parameters can be collected as $\theta = [\ \rho^\top \quad \eta^\top\ ]$ and the signal relations of the LPV-BJ model operating in closed loop, denoted in the sequel as $\mathcal{M}_\theta$, are defined (in the same way as in

the open-loop case in which the closed-loop equation is added) as :

$$\mathcal{M}_\theta \begin{cases} \underbrace{\chi(t_k) + \sum_{i=1}^{n_\mathrm{a}} a_{i,0}\chi(t_{k-i})}_{F(q^{-1},\rho)\chi(t_k)} + \sum_{i=1}^{n_\mathrm{a}}\sum_{l=1}^{n_\alpha} a_{i,l}\underbrace{f_l(p_k)\chi(t_{k-i})}_{\chi_{i,l}(t_k)} = \sum_{j=0}^{n_\mathrm{b}}\sum_{l=0}^{n_\beta} b_{j,l}\underbrace{g_l(p_k)u(t_{k-j})}_{u_{j,l}(t_k)} \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)}e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \\ u(t_k) = r_1(t_k) + \mathcal{C}_\mathrm{o}(r_2(t_k) - y(t_k)) \end{cases} \tag{3.56}$$

where $F(q^{-1},\rho) = 1 + \sum_{i=1}^{n_\mathrm{a}} a_{i,0}q^{-i}$ and $g_0(t_k) = 1$. Again, in this way, the LPV-BJ model is rewritten as a *Multiple-Input Single-Output* (MISO) system with $(n_\mathrm{b}+1)(n_\beta+1) + n_\mathrm{a}n_\alpha$ inputs $\{\chi_{i,l}\}_{i=1,l=1}^{n_\mathrm{a},n_\alpha}$ and $\{u_{j,l}\}_{j=0,l=0}^{n_\mathrm{b},n_\beta}$ as represented in Fig. 3.2.
Based on this model structure, the whole model set including the process ($\mathcal{G}_\rho$) and noise ($\mathcal{H}_\eta$) models parameterized independently, is denoted as $\mathcal{M}$ and takes finally the form

$$\mathcal{M} = \left\{(\mathcal{G}_\rho, \mathcal{H}_\eta) \;\middle|\; \mathrm{col}(\rho,\eta) = \theta \in \mathbb{R}^{n_\rho+n_\eta}\right\}. \tag{3.57}$$

The set (3.57) corresponds to the set of candidate models in which we seek the best fitting model using data gathered from $\mathcal{S}_\mathrm{o}$ under a given identification criterion (cost function). In order to define the minimization criterion, the same predictor and prediction error scheme can be defined as in section 3.2.3 and directly applied in order to defined the identification problem.

3.3.3 Identification problem statement

Based on the previous considerations, the identification problem addressed in the sequel can now be defined.

Problem 6 *Given a discrete time closed-loop LPV data generating system $\mathcal{S}_\mathrm{o}$ depicted in Fig. 3.5 and a data set $\mathcal{D}_N = \{y(t_k), u(t_k), r_1(t_k), r_2(t_k), p(t_k)\}_{k=1}^N$ collected from $\mathcal{S}_\mathrm{o}$. Based on the LPV-BJ model structure $\mathcal{M}_\theta$ defined by (3.56), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions :*

LPVCLA1 *$\mathcal{S}_\mathrm{o} \in \mathcal{M}$, i.e. there exits a θ_o defining a $\mathcal{G}_{\rho_\mathrm{o}} \in \mathcal{G}$ and a $\mathcal{H}_{\eta_\mathrm{o}} \in \mathcal{H}$ such that $(\mathcal{G}_{\rho_\mathrm{o}}, \mathcal{H}_{\eta_\mathrm{o}})$ is equal to $\mathcal{S}_\mathrm{o}$.*

LPVCLA2 *In the parametrization $\mathcal{A}_\rho$ and $\mathcal{B}_\rho$, $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are chosen such that $(\mathcal{G}_\mathrm{o}, \mathcal{H}_\mathrm{o})$ is identifiable for any trajectory of p.*

LPVCLA3 *Two different models inside $\mathcal{M}$ produce two different prediction errors given $\mathcal{D}_N$.*

LPVCLA4 *$\mathcal{S}_\mathrm{o}$ is globally BIBO stable, i.e. for any trajectory of $p : \mathbb{R} \mapsto \mathbb{P}$ and any bounded input signal pair $\{r_1, r_2\}$, the output of $\mathcal{S}_\mathrm{o}$ is bounded [Tót10].*

LPVCLA5 *The process of $\mathcal{S}_\mathrm{o}$ is globally BIBO stable, i.e. for any trajectory of $p : \mathbb{R} \mapsto \mathbb{P}$ and any bounded input signal u, the output χ is bounded. [Tót10].*

3.3.4 Optimal IV for CL LPV models

If the process to be identified belongs to the model set previously defined, then $y(t_k)$ can be written under the same filtered regression form as in the open-loop case (3.40) :

$$y_{\mathrm{f}}(t_k) = \varphi_{\mathrm{f}}^{\top}(t_k)\rho + \tilde{v}_{\mathrm{f}}(t_k) \tag{3.58}$$

where, φ and ρ are given in (3.36). Under these conditions, again, when the data is generated from $\mathcal{S}_{\mathrm{o}}$, if $\theta = \theta_{\mathrm{o}}$, then $\tilde{v}_{\mathrm{f}}(t_k) = e_{\mathrm{o}}(t_k)$ which is a white noise.
Nonetheless, even if the description of the problem is in many points similar to the open-loop case, estimating the process between u and y is complicated by the fact that the feedback loop carries noisy information. Therefore, even by using the optimal instrument as given in (3.41) will not lead to fullfil C2 (see Section 1.5.5) as the signal u is corrupted by noise. Therefore, by extending the RIV method presented in Section 3.2 to systems operating in closed-loop, we propose to use the following instrument :

$$\zeta(t_k) \;=\; \left[\; -\mathring{\chi}(t_{k-1}) \dots -\mathring{\chi}(t_{k-n_{\mathrm{a}}}) \quad -\mathring{\chi}_{1,1}(t_k) \dots \; -\mathring{\chi}_{n_{\mathrm{a}},n_\alpha}(t_k) \quad \mathring{u}_{0,0}(t_k) \dots \; \mathring{u}_{n_{\mathrm{b}},n_\beta}(t_k) \;\right]^{\top} \tag{3.59}$$

where $\mathring{u}$ and $\mathring{\chi}$ are the signals from the noise-free system $\mathring{\mathcal{S}}_{\mathrm{o}}$ as presented in Figure 3.6 and where

$$\begin{cases} \mathring{u}_{j,l}(t_k) = g_l(p_k)\mathring{u}(t_{k-j}), \quad j = 0 \dots n_{\mathrm{b}}, l = 0 \dots n_\beta \\ \mathring{\chi}_{i,l}(t_k) = f_l(p_k)\mathring{\chi}(t_{k-i}) \quad i = 1 \dots n_{\mathrm{a}}, l = 0 \dots n_\alpha. \end{cases} \tag{3.60}$$

It can be moreover noticed that there is a difference of notation between :
– χ_{o} which is the notation for the process output of the noisy system $\mathcal{S}_{\mathrm{o}}$ (therefore χ_{o} is corrupted with noise through the feedback loop).
– $\mathring{\chi}$ is the output of the noise-free system $\mathring{\mathcal{S}}_{\mathrm{o}}$.
The optimal filter is the same as the one given in the open-loop case and is given by :

$$L^{\mathrm{opt}}(q) = Q(q, \theta_{\mathrm{o}}) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})F_{\mathrm{o}}(q^{-1})}. \tag{3.61}$$

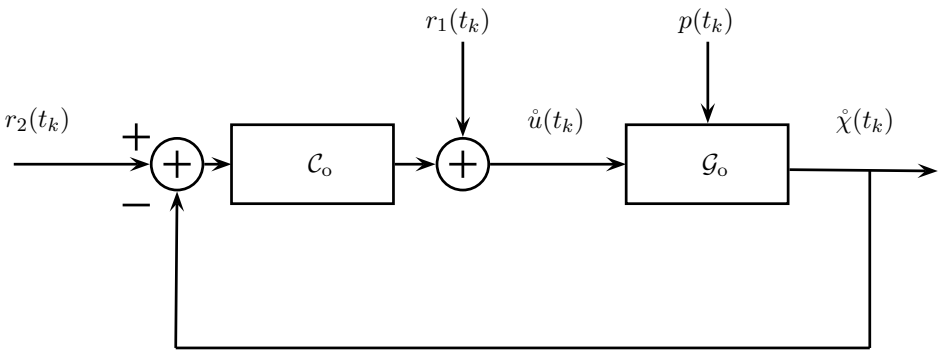


FIG. 3.6 – Noise-free system

3.3.5 LPV-closed-loop RIV (LPV-CLRIV) algorithm for BJ models

Based on the previous considerations, the iterative scheme of the RIV algorithm can be extended to the closed-loop LPV case as follows.

Algorithm 7 (LPV-CLRIV)

Step 1 *Generate an initial estimate of the process model parameter* $\hat{\rho}^{(0)}$ *(*e.g. *using the LS method). Set* $C(q^{-1},\hat{\eta}^{(0)}) = D(q^{-1},\hat{\eta}^{(0)}) = 1$. *Set* $\tau = 0$.

Step 2 *Compute an estimate of* $\chi(t_k)$ *via*

$$A(p_k, q^{-1}, \hat{\rho}^{(\tau)})\hat{\chi}(t_k) = B(p_k, q^{-1}, \hat{\rho}^{(\tau)})u(t_k),$$

where $\hat{\rho}^{(\tau)}$ *the estimate obtained at the previous iteration. Based on* $\mathcal{M}_{\hat{\theta}^{(\tau)}}$, *deduce* $\{\hat{\chi}_{i,l}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$ *as given in* (3.56). *According to Assumption LPVCLA5 each* $\hat{\chi}_{i,l}$ *is bounded.It is important to notice here that* $\hat{\chi}(t_k)$ *is correlated with the noise as it is simulated using the measured* u *signal.*

Step 3 *Compute the filter as in 3.39 :*

$$L(q,\hat{\theta}^{(\tau)}) = \frac{D(q^{-1},\hat{\eta}^{(\tau)})}{C(q^{-1},\hat{\eta}^{(\tau)})F(q^{-1},\hat{\rho}^{(\tau)})}$$

and the associated filtered signals $\{u^{\mathrm{f}}_{j,l}(t_k)\}_{j=0,l=0}^{n_{\mathrm{b}},n_\beta}$, $y_{\mathrm{f}}(t_k)$ *and* $\{\hat{\chi}^{\mathrm{f}}_{i,l}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$.

Step 4 *Build the noisy filtered regressor* $\hat{\varphi}_{\mathrm{f}}(t_k)$ *using the filter* $L(q,\hat{\theta}^{(\tau)})$ *:*

$$\hat{\varphi}_{\mathrm{f}}(t_k) = \Big[\ -y_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -y_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad -\hat{\chi}^{\mathrm{f}}_{1,1}(t_k) \quad \dots \quad -\hat{\chi}^{\mathrm{f}}_{n_{\mathrm{a}},n_\alpha}(t_k) \quad u^{\mathrm{f}}_{0,0}(t_k) \quad \dots \quad u^{\mathrm{f}}_{n_{\mathrm{b}},n_\beta}(t_k)\ \Big]^\top$$

Compute the filtered instrument $\hat{\zeta}_{\mathrm{f}}(t_k)$ *by simulating the estimated auxiliary model as given in figure 3.7 and* y *using the filter* $L(q,\hat{\theta}^{(\tau)})$ *:*

$$\hat{\zeta}_{\mathrm{f}}(t_k) = \Big[\ -\hat{\mathring{\chi}}_{\mathrm{f}}(t_{k-1}) \quad \dots \quad -\hat{\mathring{\chi}}_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \quad -\hat{\mathring{\chi}}^{\mathrm{f}}_{1,1}(t_k) \quad \dots \quad -\hat{\mathring{\chi}}^{\mathrm{f}}_{n_{\mathrm{a}},n_\alpha}(t_k) \quad \hat{\mathring{u}}^{\mathrm{f}}_{0,0}(t_k) \quad \dots \quad \hat{\mathring{u}}^{\mathrm{f}}_{n_{\mathrm{b}},n_\beta}(t_k)\ \Big]^\top$$

According to Assumption LPVCLA4, each $\hat{\mathring{\chi}}_{i,l}$ *and* $\hat{\mathring{u}}_{i,l}$ *is bounded.*

Step 5 *The IV optimization problem can now be stated in the form*

$$\hat{\rho}^{(\tau+1)}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)\hat{\varphi}_{\mathrm{f}}^\top(t_k)\right]\rho - \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)y_{\mathrm{f}}(t_k)\right]\right\|^2 \tag{3.62}$$

where the solution is obtained as

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)\hat{\varphi}_{\mathrm{f}}^\top(t_k)\right]^{-1}\sum_{k=1}^{N}\hat{\zeta}_{\mathrm{f}}(t_k)y_{\mathrm{f}}(t_k).$$

The resulting $\hat{\rho}^{(\tau+1)}(N)$ *is the IV estimate of the process model associated parameter vector at iteration* $\tau+1$ *based on the prefiltered input/output data.*

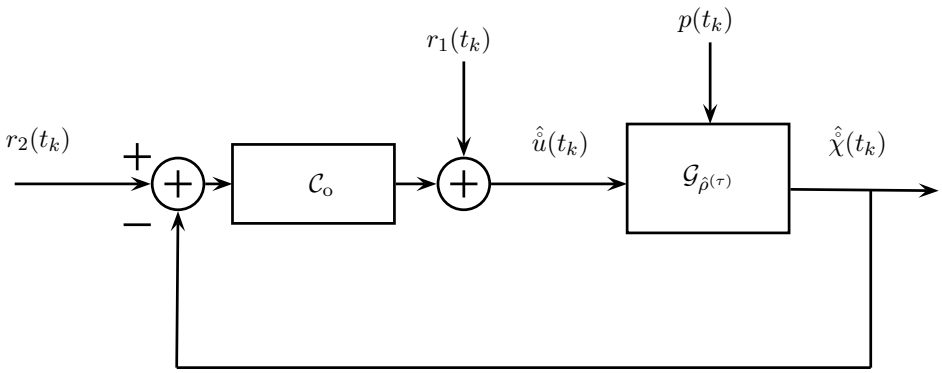


FIG. 3.7 – Estimated auxiliary model

Step 6 *An estimate of the noise signal v is obtained as*

$$\hat{v}(t_k) = y(t_k) - \hat{\chi}(t_k, \hat{\rho}^{(\tau)}). \tag{3.63}$$

Based on $\hat{v}$, the estimation of the noise model parameter vector $\hat{\eta}^{(\tau+1)}$ follows, using in this case the `ARMA` *estimation algorithm of the* `MATLAB` *identification toolbox (an IV approach can also be used for this purpose, see [You08]).*

Step 7 *If $\theta^{(\tau+1)}$ has converged or the maximum number of iterations is reached, then stop, else increase τ by 1 and go to Step 2.*

Moreover, after the convergence is complete, it is possible to compute the estimated parametric error covariance matrix $\hat{\mathbf{P}}_\rho$ from the expression (see Section 1.5.4) :

$$\hat{\mathbf{P}}_\rho = \hat{\sigma}_e^2 \left(\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) \hat{\zeta}_{\mathrm{f}}^{\top}(t_k) \right)^{-1} \tag{3.64}$$

where $\hat{\zeta}$ is the IV vector obtained at convergence and $\hat{\sigma}_e^2$ is the estimated residual variance.

3.3.6 `LPV-CLSRIV` algorithm for OE models

Based on a similar concept, the so-called LPV-CL *simplified* RIV (`LPV-CLSRIV`) method, can also be developed for the estimation of LPV-OE models operating in closed loop. This method is based on a model structure (3.16) with $C(q^{-1}, \eta) = D(q^{-1}, \eta) = 1$ and consequently, Step 6 of Algorithm 7 can be skipped.

3.3.7 Performance evaluation of the presented `LPV-CLRIV` and `LPV-CLSRIV` methods

The system is mathematically described as

$$\mathcal{S}_o \begin{cases} A_o(q, p_k) = 1 + a_1^o(p_k) q^{-1} + a_2^o(p_k) q^{-2} \\ B_o(q, p_k) = b_0^o(p_k) q^{-1} + b_1^o(p_k) q^{-2} \\ H_o(q) = \dfrac{1}{1 - q^{-1} + 0.2 q^{-2}} \end{cases} \tag{3.65}$$

where $v(t_k) = H_o(q)e(t_k)$ and

$$a_1^o(p_k) = 1 - 0.5p_k - 0.1p_k^2, \tag{3.66a}$$
$$a_2^o(p_k) = 0.5 - 0.7p_k - 0.1p_k^2, \tag{3.66b}$$
$$b_0^o(p_k) = 0.5 - 0.4p_k + 0.01p_k^2, \tag{3.66c}$$
$$b_1^o(p_k) = 0.2 - 0.3p_k - 0.02p_k^2. \tag{3.66d}$$

The controller equation is given as :

$$u(t_k) = \frac{1 + 0.5q^{-1}}{1 - 0.85q^{-1}}(r_2(t_k) - y(t_k)) + r_1(t_k) \tag{3.67}$$

In the upcoming examples, the scheduling signal p is considered as a periodic function of time : $p_k = 0.5\sin(0.35\pi k) + 0.5$. The input $r_1(t_k)$ is taken as a white noise with a uniform distribution $\mathcal{U}(-1, 1)$ and with length $N = 4000$ to generate data sets $\mathcal{D}_N$ of $\mathcal{S}_o$.
The proposed LPV *Closed Loop Refined Instrumental Variable* method (`LPV-CLRIV`) represents the situation $\mathcal{S}_o \in \mathcal{M}$ and assumes the following LPV-BJ model :

$$\mathcal{M}_\theta^{\texttt{LPV-CLRIV}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = \dfrac{1}{1 + d_1q^{-1} + d_2q^{-2}} \end{cases}$$

while the LPV *Simplified Closed-Loop Refined Instrumental Variable* method (`LPV-SCLRIV`) represents the case when $\mathcal{G}_o \in \mathcal{G}$, $\mathcal{H}_o \notin \mathcal{H}$ and assumes the following LPV-OE model :

$$\mathcal{M}_\theta^{\texttt{LPV-CLSRIV}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = 1 \end{cases}$$

In both cases,

$$a_1(p_k) = a_{1,0} + a_{1,1}p_k + a_{1,2}p_k^2 \tag{3.68a}$$
$$a_2(p_k) = a_{2,0} + a_{2,1}p_k + a_{2,2}p_k^2 \tag{3.68b}$$
$$b_0(p_k) = b_{0,0} + b_{0,1}p_k + b_{0,2}p_k^2 \tag{3.68c}$$
$$b_1(p_k) = b_{1,0} + b_{1,1}p_k + b_{1,2}p_k^2 \tag{3.68d}$$

In the sequel, the conventional LPV *Least Square* (`LS`) method [BG02a] is compared to the proposed IV approaches. The `LS` method assumes the following LPV-ARX model :

$$\mathcal{M}_\theta^{\texttt{LS}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = A^\dagger(p_k, q^{-1}, \rho). \end{cases}$$

The least square method is not suited for OE models and may not be a good candidate for comparison. Nonetheless, this method results in estimated parameters with a very low variance and is therefore a good indicator for the efficiency of the RIV estimator.

The result of a Monte Carlo simulation are shown in Table 3.7. The MCs results are based on $N_{\text{run}} = 100$ random realization, with the Gaussian white noise input to the ARMA noise model being selected randomly for each realization. The noise added at the output is adjusted such that it corresponds to a SNR (2.59) of 20dB, 15dB, 10dB and 5dB. In order to compare the statistical performance of the different approaches, the computed bias norm (BN) and variance norm (VN) the estimated parameters (3.49) are presented. The number of iterations (Nit) needed for convergence is also given.
It can be seen from Table 3.7 that the IV methods are unbiased according to the theoretical results. For SNR down to 5dB, the `LPV-CLRIV` produces a parameter variance very close to the one obtained with the `LS` method. Although the statistical optimality of the algorithm cannot be proved, this latter result shows on this example, that the LPV-CLRIV algorithm results in accurate estimates. The suboptimal `LPV-CLSRIV` method offers satisfying results, considering that the noise model assumed is wrong. The variance in the estimated parameters is twice as much as in the `LPV-CLRIV` case and it is still close to the variance of the `LS` method.

TAB. 3.7 – Closed-loop estimator bias and variance norm at different SNR

Method		20dB	15dB	10dB	5dB
`LS`	BN	1.5736	2.3922	2.0812	2.0908
	VN	0.0092	0.0140	0.0253	0.0326
`LPV -SRIV`	BN	0.0067	0.0145	0.0125	0.0052
	VN	0.0011	0.0035	0.0110	0.0319
	Nit	16	16	17	19
`LPV -RIV`	BN	0.0060	0.0139	0.0185	0.0146
	VN	0.000623	0.0018	0.0068	0.0206
	Nit	19	19	20	22

3.4 Continuous-time LPV model identification

3.4.1 System description

Consider the data generating CT LPV system with static dependancy on p described by the following equations

$$\mathcal{S}_{\text{o}} \begin{cases} A_{\text{o}}(p_t, \mathcal{d})\chi_{\text{o}}(t) = B_{\text{o}}(p_t, \mathcal{d})u(t) \\ y(t) = \chi_{\text{o}}(t) + v_{\text{o}}(t), \end{cases} \tag{3.69}$$

where $\mathcal{d}$ denotes the differentiation operator with respect to time, i.e. $\mathcal{d} = \frac{\text{d}}{\text{d}t}$, $p : \mathbb{R} \to \mathbb{P}$ is the scheduling variable with $p_t = p(t)$, χ_{o} is the noise-free output, v_{o} is a quasi stationary noise process with bounded spectral density and uncorrelated to p. A_{o} and B_{o} are polynomials in $\mathcal{d}$ with coefficients a_i^{o} and b_i^{o} that are meromorphic functions of p

$$A_{\text{o}}(p_t, \mathcal{d}) = \mathcal{d}^{n_a} + \sum_{i=1}^{n_{\text{a}}} a_i^{\text{o}}(p_t)\mathcal{d}^{n_a - i}, \qquad \text{and} \qquad B_{\text{o}}(p_t, \mathcal{d}) = \sum_{j=0}^{n_{\text{b}}} b_j^{\text{o}}(p_t)\mathcal{d}^{n_b - j}. \tag{3.70}$$

In terms of identification we can assume that sampled measurements of (y, p, u) are available at a sampling time $\mathtt{T_s} > 0$. Hence, we will denote the discrete-time samples of these signals as $u(t_k) = u(k\mathtt{T_s})$, where $k \in \mathbb{Z}$.
Again, like in the Hammerstein case, the basic idea to solve the noisy CT modelling problem of (3.69) is to assume that the CT noise process $v_\mathrm{o}(t)$ can be written at the sampling instances as a DT noise process filtered by a DT transfer function. The practically general case is considered where the colored noise associated with the sampled output measurement $y(t_k)$ is assumed to have a rational spectral density which might has no relation to the actual process dynamics of $\mathcal{S}_\mathrm{o}$. As a preliminary step towards the case of a p-dependent noise, it is also assumed that this rational spectral density is not dependent on p, like *e.g.* in case of measurement noise. Therefore, v_o is represented by a discrete-time *autoregressive moving average* (ARMA) model

$$v_\mathrm{o}(t_k) = H_\mathrm{o}(q)e_\mathrm{o}(t_k) = \frac{C_\mathrm{o}(q^{-1})}{D_\mathrm{o}(q^{-1})}e_\mathrm{o}(t_k), \tag{3.71}$$

where $e_\mathrm{o}(t_k)$ is a DT zero mean white noise process, q^{-1} is the backward time shift operator, i.e. $q^{-i}u(t_k) = u(t_{k-i})$. C_o and D_o are monic polynomials with constant coefficients. This avoids the rather difficult mathematical problem of treating sampled CT random process and their equivalent in terms of a filtered piece-wise constant CT noise source (see [PSR00],[Joh94]). In terms of (3.71), $y(t_k)$ can be written as

$$y(t_k) = \chi_\mathrm{o}(t_k) + v_\mathrm{o}(t_k), \tag{3.72}$$

which corresponds to a so called hybrid Box-Jenkins system concept already used in CT identification of LTI systems (see [PSR00], [Joh94], [LGGY08]). Furthermore, in terms of (3.71), exactly the same noise assumption is made as in the classical DT Box-Jenkins models [Lju99].

3.4.2 Model considered

3.4.2.1 Process model

The process model is denoted by $\mathcal{G}_\rho$ and defined in a form of an LPV-IO representation :

$$\mathcal{G}_\rho : (A(p_t, \mathscr{d}, \rho), B(p_t, \mathscr{d}, \rho)) \tag{3.73}$$

where the p-dependent polynomials A and B given as

$$\begin{aligned} A(p_t, \mathscr{d}, \rho) &= \mathscr{d}^{n_a} + \sum_{i=1}^{n_\mathrm{a}} a_i(p_t)\mathscr{d}^{n_a - i}, \\ B(p_t, \mathscr{d}, \rho) &= \sum_{j=0}^{n_\mathrm{b}} b_j(p_t)\mathscr{d}^{n_b - j}, \end{aligned}$$

are parameterized as

$$\begin{aligned} a_i(p_t) &= a_{i,0} + \sum_{l=1}^{n_\alpha} a_{i,l} f_l(p_t) \quad i = 1, \ldots, n_\mathrm{a} \\ b_j(p_t) &= b_{j,0} + \sum_{l=1}^{n_\beta} b_{j,l} g_l(p_t) \quad j = 0, \ldots, n_\mathrm{b} \end{aligned}$$

with parameters $\rho = [\ \mathtt{a}_1 \quad \dots \quad \mathtt{a}_{n_\mathrm{a}} \quad \mathtt{b}_0 \quad \dots \quad \mathtt{b}_{n_\mathrm{b}}\ ]^\top \in \mathbb{R}^{n_\rho}$, where

$$\begin{aligned} \mathtt{a}_i &= [\ a_{i,0} \quad a_{i,1} \quad \dots \quad a_{i,n_\alpha}\ ] \in \mathbb{R}^{n_\alpha+1} \\ \mathtt{b}_j &= [\ b_{j,0} \quad b_{j,1} \quad \dots \quad b_{j,n_\beta}\ ] \in \mathbb{R}^{n_\beta+1} \end{aligned}$$

and $n_\rho = n_\mathrm{a}(n_\alpha+1) + (n_\mathrm{b}+1)(n_\beta+1)$. Introduce also $\mathcal{G} = \{\mathcal{G}_\rho \mid \rho \in \mathbb{R}^{n_\rho}\}$, as the collection of all process models in the form of (3.73).

3.4.2.2 Noise model

The noise model is denoted by $\mathcal{H}$ and defined as a DT LTI transfer function :

$$\mathcal{H}_\eta : (H(q,\eta)) \tag{3.74}$$

where H is a monic rational function given in the form of

$$H(q,\eta) = \frac{C(q^{-1},\eta)}{D(q^{-1},\eta)} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_\mathrm{c}} q^{-n_\mathrm{c}}}{1 + d_1 q^{-1} + \ldots + d_{n_\mathrm{d}} q^{-n_\mathrm{d}}}. \tag{3.75}$$

The associated model parameters η are stacked columnwise in the parameter vector,

$$\eta = [\ c_1 \quad \dots \quad c_{n_\mathrm{c}} \quad d_1 \quad \dots \quad d_{n_\mathrm{d}}\ ]^\top \in \mathbb{R}^{n_\eta}, \tag{3.76}$$

where $n_\eta = n_\mathrm{c} + n_\mathrm{d}$. Additionally, denote $\mathcal{H} = \{\mathcal{H}_\eta \mid \eta \in \mathbb{R}^{n_\eta}\}$, the collection of all noise models in the form of (3.74).

3.4.2.3 Whole model

With respect to a given process and noise part $(\mathcal{G}_\rho, \mathcal{H}_\eta)$, the parameters can be collected as $\theta = [\ \rho^\top \quad \eta^\top\ ]$ and the signal relations of the LPV-BJ model, denoted in the sequel as $\mathcal{M}_\theta$, are defined as :

$$\mathcal{M}_\theta \begin{cases} A(p_k, \mathrm{d}, \rho)\chi(t) = B(p_k, \mathrm{d}, \rho)u(t) \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{3.77}$$

Based on this model structure, the model set, denoted as $\mathcal{M}$, with process $(\mathcal{G}_\rho)$ and noise $(\mathcal{H}_\eta)$ models parameterized independently, takes the form

$$\mathcal{M} = \left\{ (\mathcal{G}_\rho, \mathcal{H}_\eta) \ \mid \ \mathrm{col}(\rho,\eta) = \theta \in \mathbb{R}^{n_\rho + n_\eta} \right\}. \tag{3.78}$$

This set corresponds to the set of candidate models in which we seek the model that explains data gathered from $\mathcal{S}_\mathrm{o}$ the best, under a given identification criterion (cost function).
SImilarly to the DT case, in order to introduce a method which provides a solution to the identification problem of Hybrid LPV-BJ models, the signal relationships (3.77) can be rewritten as

$$\mathcal{M}_\theta \begin{cases} \underbrace{\chi(t) + \sum_{i=1}^{n_\mathrm{a}} a_{i,0} \chi^{(n_\mathrm{a}-i)}(t)}_{F(\mathrm{d},\rho)\chi(t)} + \sum_{i=1}^{n_\mathrm{a}} \sum_{l=1}^{n_\alpha} a_{i,l} \underbrace{f_l(p(t))\chi^{(n_\mathrm{a}-i)}(t)}_{\chi_{i,l}(t)} = \sum_{j=0}^{n_\mathrm{b}} \sum_{l=0}^{n_\beta} b_{j,l} \underbrace{g_l(p(t))u^{(n_\mathrm{b}-j)}(t)}_{u_{j,l}(t)} \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{3.79}$$

where $F(\not{d}, \rho) = 1 + \sum_{i=1}^{n_\mathrm{a}} a_{i,0} \not{d}^{n_\mathrm{a}-i}$ and $g_0(t) = 1$. In this way, the Hybrid LPV-BJ model is rewritten as a *Multiple-Input Single-Output* (MISO) system with $(n_\mathrm{b}+1)(n_\beta+1) + n_\mathrm{a} n_\alpha$ inputs $\{\chi_{i,l}\}_{i=1,l=1}^{n_\mathrm{a},n_\alpha}$ and $\{u_{j,l}\}_{j=0,l=0}^{n_\mathrm{b},n_\beta}$:

$$\chi_{i,l}(t) = f_l(p(t))\chi^{(n_a-i)}(t) \qquad \{i,l\} \in \{1 \dots n_\mathrm{a}, 1 \dots n_\alpha\}, \tag{3.80}$$

$$u_{j,l}(t) = g_l(p(t))u^{(n_b-j)}(t) \qquad \{j,l\} \in \{1 \dots n_\mathrm{b}, 1 \dots n_\alpha\}. \tag{3.81}$$

Given the fact that the polynomial operator commutes in this representation ($F(\not{d}, \rho)$ does not depend on $p(t)$), (3.79) can be rewritten as

$$y(t_k) = -\left(\sum_{i=1}^{n_\mathrm{a}} \sum_{l=1}^{n_\alpha} \frac{a_{i,l}}{F(\not{d}, \rho)} \chi_{i,l}\right)(t_k) + \left(\sum_{j=0}^{n_\mathrm{b}} \sum_{l=0}^{n_\beta} \frac{b_{j,l}}{F(\not{d}, \rho)} u_{k,j}\right)(t_k) + H(q,\eta)e(t_k). \tag{3.82}$$

which is an LTI representation.

3.4.3 Predictors and prediction error

Similar to the LTI case, in the LPV prediction error framework, one is concerned about finding a model in a given LPV model structure $\mathcal{M}$, which minimizes the statistical mean of the squared prediction error based on past samples of (y, u, p). However in the LPV case, no transfer function representation of systems is available.

Furthermore, multiplication with $\not{d}$ is not commutative over the p-dependent coefficients, meaning that $\not{d}\,(B(p, \not{d})u(t)) = B(\not{d}\,p, \not{d})\not{d}\,u(t)$ which is not equal to $B(p, \not{d})\not{d}\,u(t)$.

In the DT case, in order to define predictors with respect to models $\mathcal{M}_\theta \in \mathcal{M}$, a convolution type representation of the system dynamics, *i.e.* an LPV *Impulse Response Representation*, is used where the coefficients have dynamic dependence on p [Tót10], [LGTG10b]. Considering the CT case, no IRR was developed yet and therefore the same concept cannot be used to define the predictors.

3.4.3.1 System reformulation and prediction error

As (3.82) is an equivalent form of the model (3.77), and under the assumption that the system belongs to the model set defined, it holds that the data generating system $\mathcal{S}_\mathrm{o}$ has also a MISO-LTI interpretation. Therefore, based on the same idea, and with a deterministic p signal, it is possible to express the CT LPV system as a CT MISO LTI system by rewriting the signal relations of (3.69) as

$$\underbrace{\chi_\mathrm{o}^{(n_a)}(t) + \sum_{i=1}^{n_\mathrm{a}} a_{i,0}^\mathrm{o} \chi_\mathrm{o}^{(n_a-i)}(t)}_{F_\mathrm{o}(\not{d})\chi_\mathrm{o}(t)} + \sum_{i=1}^{n_\mathrm{a}} \sum_{l=1}^{n_\alpha} a_{i,l}^\mathrm{o} \underbrace{f_l(p(t))\chi_\mathrm{o}^{(n_a-i)}(t)}_{\chi_{i,l}^\mathrm{o}(t)} = \sum_{j=0}^{n_\mathrm{b}} \sum_{l=0}^{n_\beta} b_{j,l}^\mathrm{o} \underbrace{g_l(p(t))u^{(n_b-j)}(t)}_{u_{j,l}(t)} \tag{3.83}$$

where $g_0(t) = 1$ and the superscript (n) for a signal, like $u^{(n)}$, denotes the n^{th} time-derivative of the signals, e.g. $u^{(n)}(t) = \not{d}^{\,n}u(t)$. Furthermore, $F(\not{d}, \rho) = \not{d}^{\,n_a} + \sum_{i=1}^{n_\mathrm{a}} a_{i,0} \not{d}^{\,n_a-i}$ and $u^{(n)}(t_k)$ represents the value of the signal $u^{(n)}(t)$ sampled at time t_k.

Note that in this way, the time variation of the coefficients is transposed onto the signals $\chi^{\mathrm{o}}_{i,l}(t)$ and $u_{j,l}(t)$:

$$\chi^{\mathrm{o}}_{i,l}(t) = f_l(p(t))\chi^{(n_a-i)}_{\mathrm{o}}(t) \qquad \{i,l\} \in \{1\dots n_{\mathrm{a}}, 1\dots n_\alpha\}, \tag{3.84}$$

$$u_{j,l}(t) = g_l(p(t))u^{(n_b-j)}(t) \qquad \{j,l\} \in \{1\dots n_{\mathrm{b}}, 1\dots n_\alpha\}. \tag{3.85}$$

Therefore, the process part of the LPV-BJ model is rewritten as a *Multiple-Input Single-Output* (MISO) system with $(n_{\mathrm{b}}+1)(n_\beta+1)+n_{\mathrm{a}}n_\alpha$ inputs $\{\chi^{\mathrm{o}}_{i,l}\}^{n_{\mathrm{a}},n_\alpha}_{i=1,l=1}$ and $\{u_{j,l}\}^{n_{\mathrm{b}},n_\beta}_{j=0,l=0}$. By using (3.83), (3.77) can be rewritten in terms of the sampled output signal $y(t_k)$ as

$$y(t_k) = \underbrace{-\left(\sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_\alpha}\frac{a^{\mathrm{o}}_{i,l}}{F_{\mathrm{o}}(\mathcal{d})}\chi^{\mathrm{o}}_{i,l}\right)(t_k) + \left(\sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_\beta}\frac{b^{\mathrm{o}}_{j,l}}{F_{\mathrm{o}}(\mathcal{d})}u_{k,j}\right)(t_k)}_{G_{\mathrm{o}}(\chi_{\mathrm{o}},u,t_k)} + v_{\mathrm{o}}(t_k), \tag{3.86}$$

which is a sampled LTI representation of the system defined in (3.69).
Given the Assumption that $v_{\mathrm{o}}(t_k) = H_{\mathrm{o}}(q)e_{\mathrm{o}}(t_k)$, that $C(q^{-1})$ is a monic polynomial, (3.71) can be rewritten in the form

$$v_{\mathrm{o}}(t_k) = e_{\mathrm{o}}(t_k) + \sum_{i=1}^{\infty} h_i e_{\mathrm{o}}(t_{k-i}). \tag{3.87}$$

This guarantees that the knowledge of $\{v_{\mathrm{o}}(\tau)\}_{\tau\le t_{k-1}}$ ensures the knowledge of $\{e_{\mathrm{o}}(\tau)\}_{\tau\le t_{k-1}}$. Therefore, by using the traditional approach [Lju99], the prediction of $v_{\mathrm{o}}(t_k)$ is considered as the conditional expectation of $v_{\mathrm{o}}(t_k)$ based on $\{e_{\mathrm{o}}(\tau)\}_{\tau\le t_{k-1}}$ which is according to (3.87) :

$$\hat{v}(t_k) = \hat{v}(t_k|t_{k-1}) = \mathbb{E}\{v_{\mathrm{o}}(t_k)|\{e_{\mathrm{o}}(\tau)\}_{\tau\le k-1}\} = \sum_{i=1}^{\infty} h_i e_{\mathrm{o}}(t_{k-i}) \tag{3.88}$$

where $\mathbb{E}$ is the expectation operator. Assuming that H_{o} is invertible such that $e_{\mathrm{o}}(t_k) = H^{-1}_{\mathrm{o}}(q)v_{\mathrm{o}}(t_k)$, then the classical one-step-ahead predictor can be given as [Lju99]

$$\hat{v}(t_k) = v(t_k|t_{k-1}) = v_{\mathrm{o}}(t_k) - e_{\mathrm{o}}(t_k) = (1 - H^{-1}_{\mathrm{o}}(q))v_{\mathrm{o}}(t_k). \tag{3.89}$$

Consequently, for the considered LPV system formulated as in (3.86), the one-step-ahead predictor of $y(t_k)$ (defined as the conditional expectation $\hat{y}(t_k|t_{k-1})$ of $y(t_k)$) is given by

$$\begin{aligned}
\hat{y}(t_k) =& G_{\mathrm{o}}(\chi_{\mathrm{o}},u,t_k) + \hat{v}_{\mathrm{o}}(t_k)\\
\hat{y}(t_k) =& G_{\mathrm{o}}(\chi_{\mathrm{o}},u,t_k) + (1 - H^{-1}_{\mathrm{o}}(q))v_{\mathrm{o}}(t_k)\\
\hat{y}(t_k) =& G_{\mathrm{o}}(\chi_{\mathrm{o}},u,t_k) + (1 - H^{-1}_{\mathrm{o}}(q))(y(t_k) - G_{\mathrm{o}}(\chi_{\mathrm{o}},u,t_k))\\
\hat{y}(t_k) =& H^{-1}_{\mathrm{o}}(q)G_{\mathrm{o}}(\chi_{\mathrm{o}},u,t_k) + (1 - H^{-1}_{\mathrm{o}}(q))(y(t_k))\\
\hat{y}(t_k) =& H^{-1}_{\mathrm{o}}(q)\left(-\left(\sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_\alpha}\frac{a^{\mathrm{o}}_{i,l}}{F_{\mathrm{o}}(\mathcal{d})}\chi^{\mathrm{o}}_{i,l}\right)(t_k) + \left(\sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_\beta}\frac{b^{\mathrm{o}}_{j,l}}{F_{\mathrm{o}}(\mathcal{d})}u_{k,j}\right)(t_k)\right)\\
&\qquad + (1 - H^{-1}_{\mathrm{o}}(q))(y(t_k))
\end{aligned} \tag{3.90}$$

3.4.3.2 Prediction error model

Using the model MISO LTI form (3.82) :

$$y_\theta(t_k) = -\left(\sum_{i=1}^{n_\mathrm{a}}\sum_{l=1}^{n_\alpha}\frac{a_{i,l}}{F(\mathcal{d},\rho)}\chi_{i,l}\right)(t_k) + \left(\sum_{j=0}^{n_\mathrm{b}}\sum_{l=0}^{n_\beta}\frac{b_{j,l}}{F(\mathcal{d},\rho)}u_{k,j}\right)(t_k) + H(q,\eta)e(t_k), \tag{3.91}$$

and similarly to the LTI case theory, the *one-step ahead prediction error* can be expressed and defined as [Lju99] :

$$\varepsilon_\theta(t_k) = y(t_k) - \hat{y}_\theta(t_k) \tag{3.92}$$

where $\hat{y}_\theta(t_k)$ is the *one step ahead predictor* based on the model (3.77) written as in (3.91) and is defined as (see (3.90)) :

$$\hat{y}_\theta(t_k) = H^{-1}(q,\eta)\left(-\left(\sum_{i=1}^{n_\mathrm{a}}\sum_{l=1}^{n_\alpha}\frac{a_{i,l}}{F(\mathcal{d},\rho)}\chi_{i,l}\right)(t_k) + \left(\sum_{j=0}^{n_\mathrm{b}}\sum_{l=0}^{n_\beta}\frac{b_{j,l}}{F(\mathcal{d},\rho)}u_{k,j}\right)(t_k)\right) + \Big(1 - H^{-1}(q,\eta)\Big)y(t_k). \tag{3.93}$$

3.4.3.3 Prediction error minimization

Denote $\mathcal{D}_N = \{y(t_k), u(t_k), p(t_k)\}_{k=1}^N$ a data sequence of $\mathcal{S}_\mathrm{o}$. Then, to provide an estimate of θ based on the minimization of ε_θ, an identification criterion $W(\mathcal{D}_N,\theta)$ can be introduced, like the *least square* criterion

$$W(\mathcal{D}_N,\theta) = \frac{1}{N}\sum_{k=1}^{N}\varepsilon_\theta^2(t_k), \tag{3.94}$$

such that the parameter estimate is

$$\hat{\theta}_N = \arg\min_{\theta\in\mathbb{R}^{n_\rho+n_\eta}} W(\mathcal{D}_N,\theta). \tag{3.95}$$

3.4.4 Identification problem statement

Based on the previous considerations, the identification problem addressed in the sequel can now be defined.

Problem 7 *Given a CT LPV data generating system $\mathcal{S}_\mathrm{o}$ defined as in (3.69) and a data set $\mathcal{D}_N$ collected from $\mathcal{S}_\mathrm{o}$. Given the intersample behavior of $u(t)$ and $p(t)$. Based on the hybrid LPV-BJ model structure $\mathcal{M}_\theta$ defined by (3.77), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions :*

CTLPVA1 *$\mathcal{S}_\mathrm{o} \in \mathcal{M}$, i.e. there exits a θ_o defining a $\mathcal{G}_{\rho_\mathrm{o}} \in \mathcal{G}$ and a $\mathcal{H}_{\eta_\mathrm{o}} \in \mathcal{H}$ such that $(\mathcal{G}_{\rho_\mathrm{o}}, \mathcal{H}_{\eta_\mathrm{o}})$ is equal to $\mathcal{S}_\mathrm{o}$.*

CTLPVA2 *In the parametrization $\mathcal{A}_\rho$ and $\mathcal{B}_\rho$, $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are chosen such that $(\mathcal{G}_\mathrm{o}, \mathcal{H}_\mathrm{o})$ is identifiable for any trajectory of p.*

CTLPVA3 *$u(t_k)$ is not correlated to $e_\mathrm{o}(t_k)$.*

CTLPVA4 *Two different models inside* $\mathcal{M}$ *produce two different prediction errors given* $\mathcal{D}_N$.

CTLPVA5 $\mathcal{S}_\mathrm{o}$ *is globally BIBO stable, i.e. for any trajectory of* $p : \mathbb{R} \mapsto \mathbb{P}$ *and any bounded input signal* u*, the output of* $\mathcal{S}_\mathrm{o}$ *is bounded [Tót10].*

3.4.5 Optimal instrumental variable for LPV CT systems

Based on the MISO-LTI formulation (3.91), it becomes possible in theory to achieve optimal PEM using linear regression. This allows to extend the *Refined Instrumental Variable* (RIV) approach of the LTI identification framework to provide an efficient way of identifying hybrid LPV-BJ models. Using the LTI model (3.79), $y(t_k)$ can be written in the regression form :

$$y^{(n_a)}(t_k) = \varphi^\top(t_k)\rho + \tilde{v}(t_k) \tag{3.96}$$

where,

$$\begin{aligned}
\varphi(t_k) &= [\, -y^{(n_a-1)}(t_k) \;\; \dots \;\; -y(t_k) \;\; -\chi_{1,1}(t_k) \;\; \dots \;\; -\chi_{n_\mathrm{a},n_\alpha}(t_k) \;\; u_{0,0}(t_k) \;\; \dots \;\; u_{n_\mathrm{b},n_\beta}(t_k) \,]^\top \\
\rho &= [\, a_{1,0} \;\; \dots \;\; a_{n_\mathrm{a},0} \;\; a_{1,1} \;\; \dots \;\; a_{n_\mathrm{a},n_\alpha} \;\; b_{0,0} \;\; \dots \;\; b_{n_\mathrm{b},n_\beta} \,]^\top
\end{aligned}$$

and

$$\tilde{v}(t_k) = F(\not{d}\,, \rho)v(t_k).$$

The extended regressor in (3.96) contains the noise-free output terms $\{\chi_{i,k}\}$. Therefore, by momentary assuming that $\{\chi_{i,l}(t_k)\}_{i=1,l=0}^{n_\mathrm{a},n_\alpha}$ are known *a priori*, the conventional PEM approach on (3.96) leads to the prediction error $\varepsilon_\theta(t_k)$ given as :

$$\varepsilon_\theta(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)} \left[\frac{1}{F(\not{d}\,,\rho)} \left(F(\not{d}\,,\rho)y(t_k) - \Big[-\sum_{i=1}^{n_\mathrm{a}} \sum_{l=1}^{n_\alpha} a_{i,l}\chi_{i,l}(t_k) + \sum_{j=0}^{n_\mathrm{b}} \sum_{l=0}^{n_\mathrm{b}} b_{j,l}u_{j,l}(t_k) \Big] \right) \right]. \tag{3.97}$$

In the given context, the filters $\frac{D(q^{-1},\eta)}{C(q^{-1},\eta)}$ and $F(\not{d}\,,\rho)$ in (3.97) commute (see Section 1.6).
The later allows us to rewrite *the one-step ahead prediction error* (3.97) associated to (3.96) as

$$\varepsilon_\theta(t_k) = \left(F(\not{d}\,,\rho)y_\mathrm{f}\right)(t_k) - \sum_{i=1}^{n_\mathrm{a}} \sum_{l=1}^{n_\alpha} a_{i,l}\chi_{i,l}^\mathrm{f}(t_k) + \sum_{j=0}^{n_\mathrm{b}} \sum_{l=0}^{n_\beta} b_{j,l}u_{k,j}^\mathrm{f}(t_k) \tag{3.98}$$

where $y_\mathrm{f}(t_k)$, $u_{j,l}^\mathrm{f}(t_k)$ and $\chi_{i,l}^\mathrm{f}(t_k)$ represent the outputs of the complete hybrid prefiltering operation, involving the continuous-time filtering operation using the filter (see [YGG08]) :

$$Q_\mathrm{c}(\not{d}\,,\rho) = \frac{1}{F(\not{d}\,,\rho)}, \tag{3.99}$$

and the discrete-time filtering operation using the filter :

$$Q_\mathrm{d}(q,\eta) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)}. \tag{3.100}$$

In other words :

$$y_\mathrm{f}(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)} \left[\left(\frac{1}{F(\not{d}\,,\rho)} y \right)(t_k) \right], \tag{3.101}$$

Based on (3.98), the associated linear-in-the-parameters model takes the form [YGG08] :

$$y_{\mathrm{f}}^{(n_a)}(t_k) = \varphi_{\mathrm{f}}^{\top}(t_k)\rho + \tilde{v}_{\mathrm{f}}(t_k), \tag{3.102}$$

where

$$\begin{aligned}
\varphi_{\mathrm{f}}(t_k) &= [\ -y_{\mathrm{f}}^{(n_a-1)}(t_k)\ \ \dots\ \ -y_{\mathrm{f}}(t_k)\ \ -\chi_{1,1}^{\mathrm{f}}(t_k)\ \ \dots\ \ -\chi_{n_{\mathrm{a}},n_\alpha}^{\mathrm{f}}(t_k)\ \ u_{0,0}^{\mathrm{f}}(t_k)\ \ \dots\ \ u_{n_{\mathrm{b}},n_\beta}^{\mathrm{f}}(t_k)\ ]^{\top} \\
\tilde{v}_{\mathrm{f}}(t_k) &= Q_{\mathrm{d}}(q^{-1},\eta)Q_{\mathrm{c}}(\mathscr{d},\rho)\tilde{v}(t_k).
\end{aligned}$$

Consequently when the data is generated from $\mathcal{S}_{\mathrm{o}}$, if $\theta = \theta_{\mathrm{o}}$, then $\tilde{v}_{\mathrm{f}}(t_k) = e_{\mathrm{o}}(t_k)$ which is a white noise.
Therefore, according to the conditions for optimal estimates as given in Section 1.5.5, C4 defines the optimal filtered instrument as the noise-free version of the regressor :

$$\zeta^{\mathrm{opt}}(t_k) = \left[\ -\chi_{\mathrm{o}}^{(n_a-1)}(t_{k-1}) \dots -\chi_{\mathrm{o}}(t_k) - \chi_{1,1}^{\mathrm{o}}(t_k) \dots -\chi_{n_{\mathrm{a}},n_\alpha}^{\mathrm{o}}(t_k)\quad u_{0,0}(t_k) \dots u_{n_{\mathrm{b}},n_\beta}(t_k)\ \right]^{\top} \tag{3.103}$$

while the optimal filter is given as the following chain involving the continuous-time filtering operation (see [YGG08]) :

$$L_{\mathrm{c}}^{\mathrm{opt}}(\mathscr{d}) = Q_{\mathrm{c}}(\mathscr{d},\rho_{\mathrm{o}}) = \frac{1}{F_{\mathrm{o}}(\mathscr{d})}, \tag{3.104}$$

and the discrete-time filtering operation using the filter :

$$L_{\mathrm{d}}^{\mathrm{opt}}(q) = Q_{\mathrm{d}}(q,\eta_{\mathrm{o}}) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})}. \tag{3.105}$$

In other words :

$$y_{\mathrm{f}}(t_k) = \frac{D_{\mathrm{o}}(q^{-1})}{C_{\mathrm{o}}(q^{-1})}\left[\left(\frac{1}{F_{\mathrm{o}}(\mathscr{d})}y\right)(t_k)\right], \tag{3.106}$$

Remarks

Under the assumption that both the inverse noise model $Q_{\mathrm{d}}(q^{-1},\eta)$ and the CT filter $Q_{\mathrm{c}}(\mathscr{d},\rho)$ and consequently that $\{\chi_{i,l}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$ are known *a priori*, traditional parametric estimation methods from the LTI framework could hand out efficient estimates of ρ and η. However these variable are unknown, and only some estimates will be available.
The final concern is that even if all these conditions are fulfilled, assuming that the LTI properties apply to this LPV case would mean that the noise-free output terms are *a priori* known. Therefore, even if the presented method considerably lowers the variance in the estimated parameters, the optimality cannot be guaranteed.

3.4.6 The `LPV-RIVC` algorithm for BJ models

Based on the previous considerations, the iterative scheme of the LPV-*Refined Instrumental Variable for Continuous-time models* (`LPV-RIVC`) as well as the simplified version (`LPV-SRIVC`) can be given in the considered hybrid LPV framework. The following algorithm is designed for hybrid LPV BJ models.

Algorithm 8 (LPV-RIVC)

Step 1 *Generate an initial estimate* $\hat{\rho}^{(0)}$ *using the* SRIVC *algorithm (by simply firstly forcing an LTI structure). Set* $C(\mathscr{d}, \hat{\eta}^{(0)}) = D(\mathscr{d}, \hat{\eta}^{(0)}) = 1$. *Set* $\tau = 0$.

Step 2 *Compute an estimate of* $\chi(t_k)$ *via numerical approximation of*

$$A(p_t, \mathscr{d}, \hat{\rho}^{(\tau)})\hat{\chi}(t) = B(p_t, \mathscr{d}, \hat{\rho}^{(\tau)})u(t),$$

where $\hat{\rho}^{(\tau)}$ *is the estimate obtained at the previous iteration. Based on* $\mathcal{M}_{\hat{\theta}^{(\tau)}}$, *deduce* $\hat{\chi}(t_k)$ *which is bounded according to Assumption CTLPVA5.*

Step 3 *Compute the estimated continuous-time filter using* (3.99) *:*

$$L_{\mathrm{c}}(\mathscr{d}, \hat{\rho}^{(\tau)}) = \frac{1}{F(\mathscr{d}, \hat{\rho}^{(\tau)})} \tag{3.107}$$

where $F(\mathscr{d}, \hat{\rho}^{(\tau)})$ *is as given in* (3.79).

Step 4 *Use the CT filter* $L_{\mathrm{c}}(\mathscr{d}, \hat{\rho}^{(\tau)})$ *as well as* $\hat{\chi}(t_k)$ *in order to generate the estimates of the derivatives which are needed :*

$$\begin{aligned}
&\{L_{\mathrm{c}}(\mathscr{d}, \hat{\rho}^{(\tau)})\hat{\chi}^{(i)}(t_k)\}_{i=0}^{n_{\mathrm{a}}-1} \\
&\{L_{\mathrm{c}}(\mathscr{d}, \hat{\rho}^{(\tau)})\hat{y}^{(i)}(t_k)\}_{i=0}^{n_{\mathrm{a}}-1} \\
&\{L_{\mathrm{c}}(\mathscr{d}, \hat{\rho}^{(\tau)})u_{j,l}(t_k)\}_{j=0,l=0}^{n_{\mathrm{b}},n_\beta} \\
&\{L_{\mathrm{c}}(\mathscr{d}, \hat{\rho}^{(\tau)})\hat{\chi}_{i,l}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}
\end{aligned}$$

Step 5 *Compute the estimated discrete-time filter as in* (3.100) *:*

$$L_{\mathrm{d}}(q, \hat{\eta}^{(\tau)}) = \frac{D(q^{-1}, \hat{\eta}^{(\tau)})}{C(q^{-1}, \hat{\eta}^{(\tau)})}$$

Step 6 *The needed filtered signals* $\{u_{j,l}^{\mathrm{f}}(t_k)\}_{j=0,l=0}^{n_{\mathrm{b}},n_\beta}$, $y_{\mathrm{f}}(t_k)$ *and* $\{\chi_{i,l}^{\mathrm{f}}(t_k)\}_{i=1,l=0}^{n_{\mathrm{a}},n_\alpha}$ *are computed using the DT filter on the estimated derivatives obtained from Step 4.*

Step 7 *Build the filtered regressor* $\hat{\varphi}_{\mathrm{f}}(t_k)$ *and, in terms of C4, the filtered instrument* $\hat{\zeta}_{\mathrm{f}}(t_k)$ *as :*

$$\begin{aligned}
\hat{\varphi}_{\mathrm{f}}(t_k) = \Big[& -y_{\mathrm{f}}^{(n_a-1)}(t_k) \quad \dots \quad -y_{\mathrm{f}}(t_k) \\
& -\hat{\chi}_{1,1}^{\mathrm{f}}(t_k) \quad \dots \quad -\hat{\chi}_{n_{\mathrm{a}},n_\alpha}^{\mathrm{f}}(t_k) \quad u_{0,0}^{\mathrm{f}}(t_k) \quad \dots \quad u_{n_{\mathrm{b}},n_\beta}^{\mathrm{f}}(t_k) \Big]^\top
\end{aligned}$$

$$\begin{aligned}
\hat{\zeta}_{\mathrm{f}}(t_k) = \Big[& -\hat{\chi}_{\mathrm{f}}^{(n_a-1)}(t_{k-1}) \quad \dots \quad -\hat{\chi}_{\mathrm{f}}(t_k) \\
& -\hat{\chi}_{1,1}^{\mathrm{f}}(t_k) \quad \dots \quad -\hat{\chi}_{n_{\mathrm{a}},n_\alpha}^{\mathrm{f}}(t_k) \quad u_{0,0}^{\mathrm{f}}(t_k) \quad \dots \quad u_{n_{\mathrm{b}},n_\beta}^{\mathrm{f}}(t_k) \Big]^\top
\end{aligned}$$

Step 8 *The IV optimization problem can now be stated in the form*

$$\hat{\rho}^{(\tau+1)}(N) = \arg\min_{\rho \in \mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\hat{\varphi}_{\mathrm{f}}^\top(t_k) \right] \rho - \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}^{(n_a)}(t_k) \right] \right\|^2 \tag{3.108}$$

where the solution is obtained as

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\hat{\varphi}_{\mathrm{f}}^{\top}(t_k)\right]^{-1} \sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k) y_{\mathrm{f}}^{(n_a)}(t_k).$$

The resulting $\hat{\rho}^{(\tau+1)}(N)$ is the IV estimate of the process model associated parameter vector at iteration $\tau+1$ based on the prefiltered input/output data.

Step 9 *An estimate of the noise signal v is obtained as*

$$\hat{v}(t_k) = y(t_k) - \hat{\chi}(t_k, \hat{\rho}^{(\tau)}). \tag{3.109}$$

Based on $\hat{v}$, the estimation of the noise model parameter vector $\hat{\eta}^{(\tau+1)}$ follows, using in this case the `ARMA` *estimation algorithm of the* `MATLAB` *identification toolbox (an IV approach can also be used for this purpose, see [You08]).*

Step 10 *If $\theta^{(\tau+1)}$ has converged or the maximum number of iterations is reached, then stop, else increase τ by 1 and go to Step 2.*

Moreover, after the convergence is complete, it is possible to compute the estimated parametric error covariance matrix $\hat{\mathbf{P}}_{\rho}$ from the expression (see Section 1.5.4) :

$$\hat{\mathbf{P}}_{\rho} = \hat{\sigma}_e^2 \left(\sum_{k=1}^{N} \hat{\zeta}_{\mathrm{f}}(t_k)\hat{\zeta}_{\mathrm{f}}^{\top}(t_k)\right)^{-1} \tag{3.110}$$

where $\hat{\zeta}$ is the IV vector obtained at convergence and $\hat{\sigma}_e^2$ is the estimated residual variance.

3.4.7 The `LPV-SRIVC` algorithm for OE models

Based on a similar concept, the so-called *simplified* `LPV-RIVC` (`LPV-SRIVC`) method, can also be developed for the estimation of CT LPV-OE models. This method is based on a model structure (3.77) with $C(q^{-1}, \eta) = D(q^{-1}, \eta) = 1$ and consequently in this case $Q_{\mathrm{d}}(q^{-1}, \eta) = 1$. Therefore the `LPV-SRIVC` algorithm remains the same as the `LPV-RIVC` algorithm except for Step 5, 6 and 10 of Algorithm 8 which are skipped. Naturally, the `LPV-SRIVC` does not lead to the statistically optimal PEM for hybrid LPV-BJ models, however it still has a certain degree of robustness as it is shown in Section 3.4.8. It can be noticed as well that a CT LPV-OE models does not involve any DT filtering and consequently, their structure is fully CT in opposition to the hybrid BJ LPV model.

3.4.8 Performance of the `LPV-RIVC` and `LPV-SRIVC` methods

As a next step, the performance of the proposed algorithms are presented on a representative simulation example. It is important to note that to the best of the author's knowledge, the presented method is the first able to handle the case of coloured output measurement noise for CT LPV models. Therefore, the results obtained for the presented algorithm cannot be compared to other algorithms. The system taken into consideration is inspired by a benchmark example proposed by Rao and Garnier in [RG02]. It has been widely used since then to demonstrate the performance of direct continuous-time identification methods [Lju03], [CSS08], [LGGY08], [RG02], [RG04]. In order to

create a CT LPV system on which the strength of direct CT identification can be demonstrated, a "moving pole" is considered. A particular feature of LPV systems is that they have an LTI representation for every constant trajectory of p. Such an LTI representation describes the so-called frozen behavior of the system and can be expressed in a transfer function form. In terms of the frozen concept, the "moving pole" means that a particular pole of these frozen transfer functions of $\mathcal{S}_\mathrm{o}$ is a function of p. This phenomenom often occurs in mechatronic applications such as for instance, wafer scanners [Was02]. In our case, the Rao-Garnier benchmark inspired "moving pole" LPV system is a fourth order with non-minimum phase frozen dynamics and a p-dependent complex pole pair. It is defined as follows :

$$\mathcal{S}_\mathrm{o}\begin{cases} A_\mathrm{o}(\mathscr{d},p) &= \mathscr{d}^4 \quad + \quad (2\zeta_2\omega_2(p)+2\zeta_1\omega_1)\,\mathscr{d}^3 \quad + \quad \left(\omega_1^2+\omega_2^2(p)+4\zeta_2\zeta_1\omega_2(p)\omega_1\right)\mathscr{d}^2 \\ & \quad + \left(2\zeta_2\omega_2(p)\omega_1^2+2\zeta_1\omega_2^2(p)\omega_1\right)\mathscr{d} \quad + \quad \omega_2^2(p)\omega_1^2 \\ B_\mathrm{o}(\mathscr{d},p) &= -T\omega_2^2(p)\omega_1^2\mathscr{d}+\omega_2^2(p)\omega_1^2 \\ H_\mathrm{o}(q) &= \dfrac{1}{1-q^{-1}+0.2q^{-2}} \end{cases} \tag{3.111}$$

where $T=4$ [sec], $\omega_1=20$ [rad/s], $\zeta_1=0.1$, $\zeta_2=0.5$. The slow frozen mode ω_2 is p-dependant and chosen as : $\omega_2=2+0.5p$. Notice that the frozen behavior (p is fixed to a constant trajectory) of $\mathcal{S}_\mathrm{o}$ for $p=0$ corresponds exactly to the Rao-Garnier benchmark defined as

$$G_\mathrm{RG}(\mathscr{d}) = \frac{-T\mathscr{d}+1}{(\frac{\mathscr{d}^2}{\omega_1^2}+2\zeta_1\frac{\mathscr{d}}{\omega_1}+1)(\frac{\mathscr{d}^2}{\omega_2^2(0)}+2\zeta_2\frac{\mathscr{d}}{\omega_2(0)}+1)}. \tag{3.112}$$

Using the given numerical values, $\mathcal{S}_\mathrm{o}$ takes the following form

$$\mathcal{S}_\mathrm{o}\begin{cases} A_\mathrm{o}(\mathscr{d},p) = \mathscr{d}^4+a_1^\mathrm{o}(p)\mathscr{d}^3+a_2^\mathrm{o}(p)\mathscr{d}^2+a_3^\mathrm{o}(p)\mathscr{d}+a_4^\mathrm{o}(p) \\ B_\mathrm{o}(\mathscr{d},p) = b_0^\mathrm{o}(p)\mathscr{d}+b_1^\mathrm{o}(p) \\ H_\mathrm{o}(q) = \dfrac{1}{1-q^{-1}+0.2q^{-2}} \end{cases} \tag{3.113}$$

where

$$a_1^\mathrm{o}(p) = 5+0.25p, \tag{3.114a}$$
$$a_2^\mathrm{o}(p) = 408+3p+0.25p^2, \tag{3.114b}$$
$$a_3^\mathrm{o}(p) = 416+108p+p^2, \tag{3.114c}$$
$$a_4^\mathrm{o}(p) = 1600+800p+100p^2, \tag{3.114d}$$
$$b_0^\mathrm{o}(p) = -6400-3200p-400p^2, \tag{3.114e}$$
$$b_1^\mathrm{o}(p) = 1600+800p+100p^2. \tag{3.114f}$$

The Bode plot of 20 frozen behaviors of $\mathcal{S}_\mathrm{o}$ is depicted in Figure 3.8 for 20 fixed values of the scheduling variable p equally distributed from -1 to 1 where the consequence of the moving low frequency mode can be clearly observed.

To obtain data records for identification purposes, the input signal u is chosen as a uniformly distributed sequence $\mathcal{U}(-1,1)$, while the scheduling variable is chosen as $p(t)=\sin(\pi t)$. Furthermore, the sampling period is chosen as $T_\mathrm{s}=1$ms, and the simulation time is $T_\mathrm{max}=10$s which, considering the currently available acquisition possibilities, is a fair assumption.

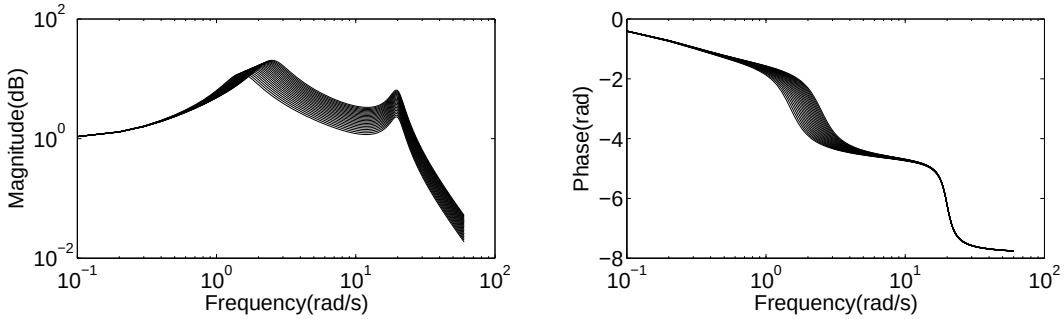


FIG. 3.8 – Bode plot of the frozen behaviors of the true LPV system for 20 values of the scheduling variable p (between -1 to 1)

In the sequel both the `LPV-RIVC` and the `LPV-SRIVC` algorithms are studied for the identification of the data generating system $\mathcal{S}_o$. The proposed `LPV-RIVC` method is applicable to hybrid LPV-BJ model and assumes the following model structure :

$$\mathcal{M}_{\texttt{LPV-RIVC}} \begin{cases} A(\not{d}, p) = \not{d}^4 + a_1(p)\not{d}^3 + a_2(p)\not{d}^2 + a_3(p)\not{d} + a_4(p) \\ B(\not{d}, p) = b_0(p)\not{d} + b_1(p) \\ H(q) = \dfrac{1}{1 + d_1 q^{-1} + d_2 q^{-2}} \end{cases}$$

where

$$a_1(p) = a_{1,0} + a_{1,1}p, \tag{3.115a}$$
$$a_2(p) = a_{2,0} + a_{2,1}p + a_{2,2}p^2, \tag{3.115b}$$
$$a_3(p) = a_{3,0} + a_{3,1}p + a_{3,2}p^2, \tag{3.115c}$$
$$a_4(p) = a_{4,0} + a_{4,1}p + a_{4,2}p^2, \tag{3.115d}$$
$$b_0(p) = b_{0,0} + b_{0,1}p + b_{0,2}p^2, \tag{3.115e}$$
$$b_1(p) = b_{1,0} + b_{1,1}p + b_{1,2}p^2. \tag{3.115f}$$

while the `LPV-SRIVC` method is applicable to CT LPV-OE models and assumes the following model structure :

$$\mathcal{M}_{\texttt{LPV-SRIVC}} \begin{cases} A(\not{d}, p) = \not{d}^4 + a_1(p)\not{d}^3 + a_2(p)\not{d}^2 + a_3(p)\not{d} + a_4(p) \\ B(\not{d}, p) = b_0(p)\not{d} + b_1(p) \\ H(q) = 1 \end{cases} \tag{3.116}$$

with $a_1(p)$, $a_2(p)$, $a_3(p)$, $a_4(p)$, $b_0(p)$, $b_1(p)$ as given in (3.115a-f).
In terms of identification, the model $\mathcal{M}_{\texttt{LPV-RIVC}}$ corresponds to the case $\mathcal{S}_o \in \mathcal{M}$ while the model $\mathcal{M}_{\texttt{LPV-SRIVC}}$ corresponds to the more realistic practical assumption $\mathcal{G}_o \in \mathcal{G}$ and $\mathcal{H}_o \notin \mathcal{H}$. Therefore, 17 parameters are to be estimated by the `LPV-SRIVC` algorithm and 19 by the `LPV-RIVC` algorithm. The results of a *Monte Carlo* simulation are presented, using $N_{\text{run}} = 200$ random realizations with a *Signal-to-Noise Ratio* (SNR) of 20dB. The MCs results obtained using both algorithms are presented in Table 3.8. It can be clearly seen that the estimates are unbiased which is conform to the theory. The standard deviation for the nominal part of the coefficients ($a_{*,0}(p)$ and $b_{*,0}(p)$) remains low whereas it raises considerably for coefficients $a_{*,2}(p)$ and $b_{*,2}(p)$.

TAB. 3.8 – Monte Carlo simulation results with additive coloured measurement noise for $SNR = 20$ dB

		LPV-RIVC		LPV-SRIVC	
Name	True Value	mean	st. dev.	mean	st. dev.
$a_{1,0}$	5	4.99	0.058	4.99	0.059
$a_{1,1}$	0.25	0.249	0.081	0.249	0.082
$a_{2,0}$	408	407.94	1.27	407.95	1.29
$a_{2,1}$	3	2.93	1.47	2.92	1.47
$a_{2,2}$	0.25	0.268	2.35	0.251	2.37
$a_{3,0}$	416	415.64	14.22	415.61	14.19
$a_{3,1}$	108	107.72	11.20	107.76	11.13
$a_{3,2}$	1	1.14	30.62	1.17	30.40
$a_{4,0}$	1600	1598.9	44.58	1599.1	44.67
$a_{4,1}$	800	799.31	27.22	799.38	27.13
$a_{4,2}$	100	101.31	79.85	101.08	80.01
$b_{0,0}$	-6400	-6396.4	47.95	-6396.4	48.03
$b_{0,1}$	-3200	-3195.2	65.72	-3195.2	65.76
$b_{0,2}$	-400	-398.92	70.40	-399.14	70.30
$b_{1,0}$	1600	1593.8	366.9	1593.8	365.7
$b_{1,1}$	800	805.02	219.77	802.43	221.7
$b_{1,2}$	100	104.86	749.61	102.95	743.9
d_1	-1	-0.999	0.0096	X	X
d_2	0.2	0.202	0.0094	X	X

In Figure 3.9, the 200 simulated model outputs are plotted together with some validation data (under the same excitation conditions as the one used for estimation). It can be seen that despite the large variance in the estimated parameters $a_{*,2}(p)$ and $b_{*,2}(p)$, the simulated outputs remain close to the true noise-free output signal (considering the level of noise corresponding to a SNR = 20dB). It appears therefore, that these parameter values have a low contribution to the observed output signal under these excitation circumstances. Thus, in terms of minimization of the squared prediction error, their role is less significant which results in a relatively large variance of their estimates under noisy conditions. This fact underlines that experiment design is needed to minimize the variance of the parameters estimates. However, there is a lack of input design methods suited for the identification of LPV systems (see [LGTG10b], [Tót10] and references therein) as the concept of persistency of excitation is not well understood yet for LPV models. A procedure for input design for LTI CT models such as in [CSS08] might be suitable to solve the estimation problem of the $a_{*,2}(p)$ and $b_{*,2}(p)$ parameters but optimal input design has not been yet investigated for the considered system.

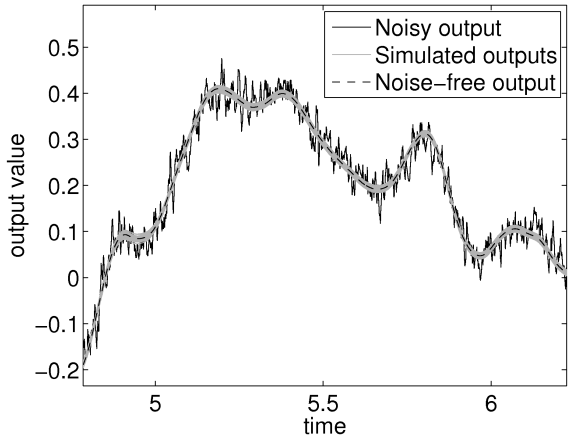


FIG. 3.9 – 200 simulated model outputs together with a noise-free and a single MCs output

Consequently, the quality of the estimated model cannot only be only judged from the parameter values. Therefore, the Bode plot of each estimated model ($N_{\mathrm{MC}} = 200$) at 20 frozen values of p from -1 to 1 are depicted in Figure 3.10. It can be clearly observed from the Bode plot that the large variance of the parameters $a_{*,2}(p)$ and $b_{*,2}(p)$ plays an important role in the low frequency area only whereas it does not affect the quality of the estimated models for frequencies above 0.5 rad/s. Therefore in order to analyze the distribution of responses at low frequencies, the density of curves at frequency $\omega = 0.1$ rad/s is displayed on the left hand-side of Figure 3.10 both for both magnitude and phase diagrams. Furthermore, this distribution density is used as color coding for drawing the frequency responses in the Bode plot : the darker a given line, the higher the number of estimated models having the same response at $\omega = 0.1$ rad/s. By using this display, it becomes clear that the estimated models are normally distributed and centered on the true model. Moreover, it can be seen from Table 3.8 that the `LPV-RIVC` and `LPV-SRIVC` give very similar variances in the estimated parameters. This reinforces the hypothesis of a sub-optimal excitation.

Finally, the most interesting advantage of the direct CT estimation in LPV framework is shown in Figure 3.11. In this figure, the poles of the 200 estimated models at 20 fixed values of p between

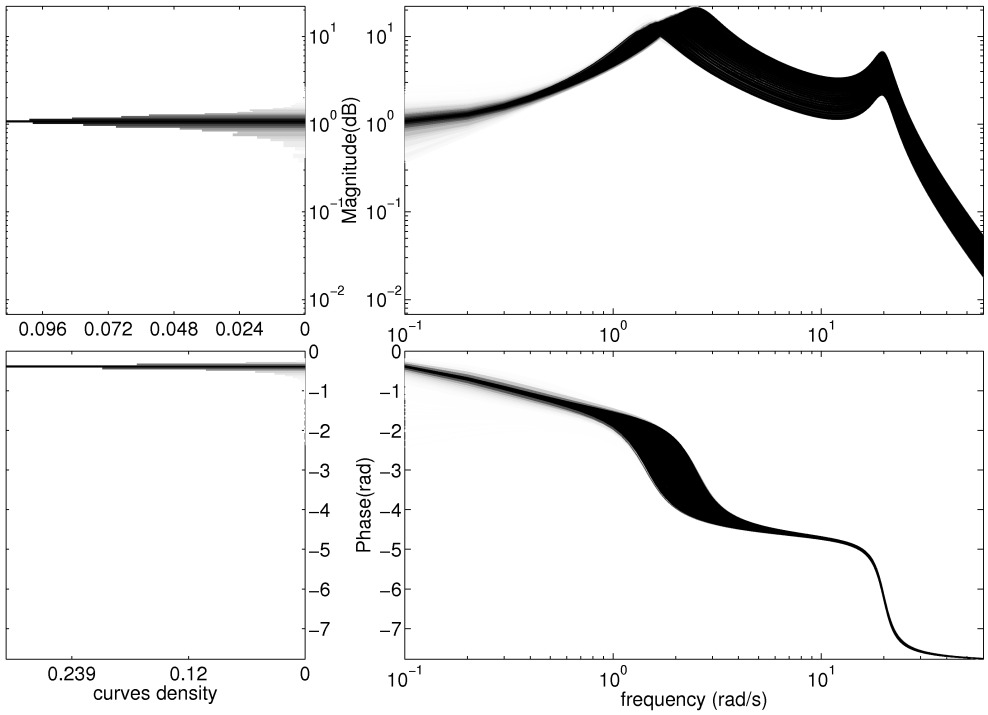


FIG. 3.10 – Bode plots of the estimated models

-1 and 1 are plot along with the true model poles. This figure also displays in the top part the density of the pole real parts. Using the same idea as in Figure 3.10, the intensity of each displayed pole is related to the number of poles among all MCs which have the same real part. Using this representation, it can clearly be seen that the pole distribution around the "fixed pole" (around $-2 \pm 20\mathrm{i}$) has a gaussian distribution while there is a sparse repartition around the "moving pole" (around $(-0.5 \pm 0.16) \pm (2 + 0.5p)\mathrm{i}$). Furthermore, the imaginary part of all estimated poles is in the close neighborhood of the "fixed pole" imaginary part. This means that by using a parsimonious CT model, the estimated models perfectly capture the time-varying nature of $\mathcal{S}_\mathrm{o}$ as well as the transient behavior from one operating condition to the other, which is a relevant achievement considering the complexity of the studied system and the level of additive noise.

Comments

It must be noticed that trying to identify a discrete-time model for such a system is quite a tedious task and two critical issues would have to be considered :

– In the LTI case, it was shown that direct CT identification methods are better suited for the identification of the Rao-Garnier benchmark [RG04]. In the presented LPV context, this means that any frozen behavior of $\mathcal{S}_\mathrm{o}$ (for any fixed trajectory of p) is better identified using a direct CT identification method.

– Using the trapezoidal integration method [Atk89], the system described in (3.3) is discretized into :

$$y(t_k) = \frac{2 - \mathtt{T_s} p(t_{k-1})}{2 + \mathtt{T_s} p(t_k)} y(t_{k-1}) + \frac{\mathtt{T_s} b}{2 + \mathtt{T_s} p(t_k)} u(t_k) + \frac{\mathtt{T_s} b}{2 + \mathtt{T_s} p(t_k)} u(t_{k-1}). \tag{3.117}$$

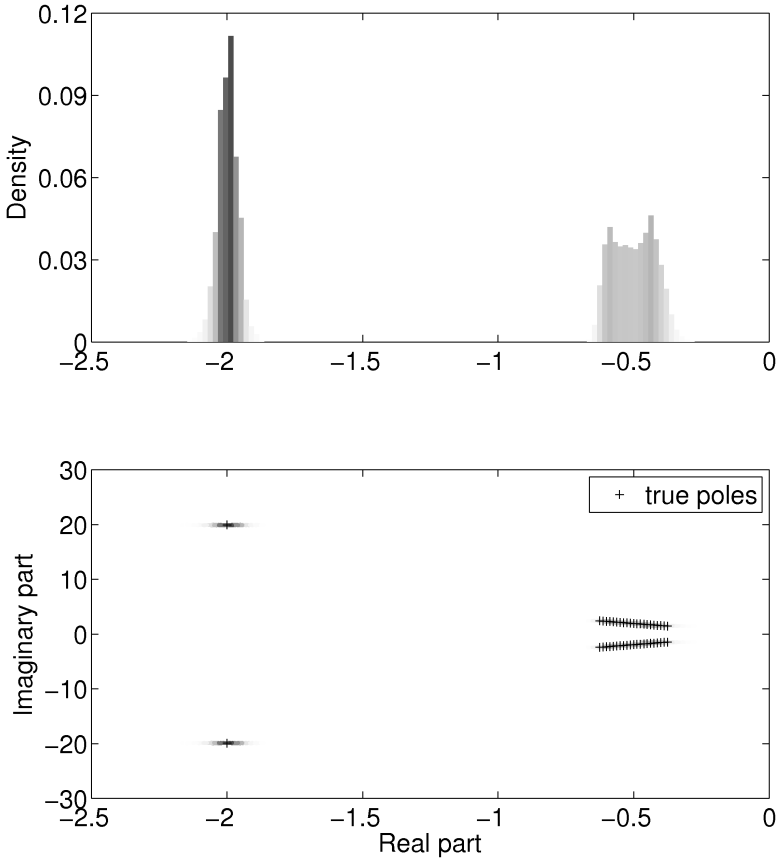


FIG. 3.11 – Poles of the estimated models and of the true system

By analyzing this simple discretized system, it is clear that discretizing the system (3.111) would result in i) the augmentation of the number of parameters to be estimated, ii) the non-linear-in-the-parameter dependence on p iii) some dynamic dependence on p (appearance of $p(t_{k-1}), p(t_{k-2}) \ldots$ terms).

Estimating such a DT model, where the dependences on p are non-linear-in-the-parameter, is hardly feasible using the existing parametric methods. The only possible way to estimate the DT model is either to use non-parametric methods, or to approximate the model, which is equivalent to relax the assumption $\mathcal{G}_o \in \mathcal{G}$: in other words, the time-varying property of the LPV system (in this case, a "moving pole") would not anymore be directly linked to the coefficients of the considered DT model and consequently, could not be clearly identified.

3.5 Conclusion

This chapter highlighted the lack of efficient methods in the literature to handle the estimation of LPV Box-Jenkins models. It has been shown that the conventional formulation of least squares estimation cannot lead to statistically optimal parameter estimates. As a solution, the LPV identification problem was reformulated and a method to estimate efficiently LPV-BJ models with p-independent noise process was proposed both for closed-loop and open-loop systems. The introduced method has been compared to the existing methods of the literature both in terms of theoretical analysis and in terms of a representative numerical example. The presented example has shown that the proposed procedure is robust to noise and outperforms the existing methods. It was shown that according to the theory, RIV based methods are unbiased for p-dependant noise models, even if the proposed method optimality cannot be proven in this case.

Moreover, due to the lack of methods dedicated to the case of direct continuous-time identification of LPV models, a novel method has been proposed in this chapter for the identification of hybrid LPV-BJ models and CT LPV-OE models with a p-independent noise process. The presented method is based on a particular MISO-LTI reformulation of the data equations which enables the use of Refined IV-based methods for LPV IO models in the error-prediction-minimization framework. The proposed algorithm has been tested on a representative numerical simulation example inspired by the Rao-Garnier benchmark. The presented example has shown that the proposed procedure is robust to noise and can reasonably well estimate the system in case of an imperfect noise model. Furthermore, it was motivated that in the given LPV framework and for relatively complicated systems, a direct CT estimation method is an attractive approach for capturing the true time-varying nature of the studied system.

A part of this work has been published in [LGTG10b], [LGTG10a] and [LTGG10].

Chapitre 4

Data-based modelling of the rainfall/runoff relationship

4.1 Hydrology and modelling

The identification of rainfall/runoff relationship is a challenging issue, mainly because of the complexity to find a suitable model for a whole given catchment [Bev00a]. However, the need for such relationship grows with the size of drainage networks in urban catchments or with the water pollution increase in agricultural regions. The inherent difficulties to define correct models are different whether it concerns urban or rural catchments. In urban catchments, the nonlinearities involved in the model are mainly caused by the water processing infrastructures [BGM01],[PL09] but most part of raw rainfall is channeled in the outlet. In rural catchments however, there is a high spatio-temporal variability of the soil property whether it lies in the vegetation, in the soil type or evapotranspiration [YG06] and there is a large difference between the total rainfall (measured rainfall) and the efficient rainfall (proportion of rainfall reaching the outlet).

Hydrology is an experimental science. In this sense, this field of research is focused on designing the experiments, acquiring the data but the interpretation of these data was mainly realised until the 1980's by using a so-called "bottom-up" process : from the knowledge of physical processes to the explanation of the data. The drawbacks of these conceptual models is that despite a good understanding at the local scale, they are hardly extendible to the catchment scale. They imply too many parameters which cannot be identified. This problem was early acknowledged in [Kle83] and more recently in [JH93], [Bev00b]. The problem associated with this type of models has been summed up in [You03] as follows :
"...such over-parameterization and the associated identifiability problems are quite often acknowledged, often implicitly, by the reductionist simulation model-builder. For example, the modeller sometimes constrains the values of certain "better known" parameters and seeks to fit the model by optimizing the chosen cost function in relation to the remaining parameters, which are normally few in number. However, the model then has a degree of "surplus content" not estimated from the available data, but based on a somewhat ad hoc evaluation of all available prior knowledge of the system and coloured by the analyst's preconceived notions of its behavioural mechanisms. "

In order to face these "reductionist philosophy" problems, "top-down" approaches were proposed : the model is determined using the measured data and despite some slight vocabulary divergences, these approaches link directly to the field of system identification. Among the models used in the literature, the linear transfer function models are widely used [You86], [Doo59], [TDV91]. Even though the simplicity of these linear models, they are still used nowadays and offer some good solution in urban context for example [PL09], [KLG+09].

Naturally, some of the rainfall/runoff relationship models were developed using nonlinear models. Just like in the general system identification framework, data-based modelling approaches can be categorized in some parametric and non-parametric approaches. Among the non-parametric approaches, the neural networks [FJ98] [TJ99] present some good forecasting possibilities. Unfortunately, they reveal a very little of their internal structure and cannot propose any physical insight of the studied system. Consequently, they present a low interest for hydrologists in case the physical interpretation of the obtained model is compulsory.

Among the parametric model estimation approaches, the first estimation method using Hammerstein model (to the best of author's knowledge) was proposed in [LG74] and has been successfully used until now [BMD09]. In this solution, the nonlinear static part is designed using physical knowledge based model. The advantage of such model is the very good forecasting capability for the studied catchment while the drawback is the need for another model when applied to data issued from other geographical regions. Some hybrid solutions were proposed more recently in [YL93] where the static nonlinearity is firstly estimated non-parametrically. Based on the resulting nonlinear function, a parsimonious parametric model inside a possible conceptual model structure is built. It is referred to in the literature as the data-based mechanistic framework. The resulting model is therefore optimized stochastically in a conceptual possible model structure. These models show also good interpretability and prediction performances. Of course, the most challenging part of this identification scheme, is to be able to propose a model structure which i) has a physical meaning, ii) is parsimonious and, iii) is able to fit the data. It can be stated therefore that the mostly used structure so far, for answering these requests is the Hammerstein model such as presented in Figure 1.4 [BMD09], [LG74], [You03], [JLW90]. In [YB94], P. Young introduced the concept of *State Dependent Parameters* (SDP) models which are in many aspects similar to the LPV models : nonetheless, those SDP models have an Hammerstein structure in which the nonlinearity is assumed to be dependent on some states of the system. Recently, the identification of full LPV models in urban catchments has been successfully applied [PL09], but the estimation method used is based on the `LS` algorithm and the determination of the scheduling variables is carried out in a non-parametrical way.

This chapter aims at presenting some generic scheme for the identification of the rainfall/runoff relationship using LPV models. Section 4.2 presents the case study as well as the inherent difficulties linked to this type of data. Section 4.3, 4.4 and 4.5 present the results of the data-based linear modelling, data-based Hammerstein modelling and data-based LPV modelling respectively.

4.2 The case study

The studied Hohrain catchment area is located in the Alsatian vineyard (Eastern part of France, altitude 284 m). The area of the catchment is 42 hectares. Exceptional annual precipitations maximally reached 867 mm (1999) and minimally 361 mm (1953). The average annual rainfall calculated since 1946 is 600 mm. The mean slope of the catchment is 15%. Geologically, Würm loamy loess and Oligocene clayey conglomerates and marls, as well as compact calcareous substrate largely dominate in the upper and lower parts of the catchment, respectively. The main soil type is mostly calcareous clay loams with medium infiltration capacity. 68% of the hydraulic catchment are covered by vineyards (see Figure 4.1). The land use shows a gradient from mostly forested areas and partly orchard at the upstream of the basin to agricultural and vineyard areas nearer to the outlet. With more than 120 farming plots, it should be noted that the road network is dense, mostly impervious and represents about 6% of the area of catchment (see Figure 4.2(d)). The catchment can be qualified as a dry catchment with no permanent flow. The hydrological functioning can be summarized in three steps : i) first of all no river network is observed and no discharge occurs without rainfall , ii) then, from a total rainfall depth of 4 mm only the road network contributes to the discharge, iii) finally, over a total rainfall depth of 8 mm, the number of fields contributing to the discharge increases with both intensity and total rainfall depth.

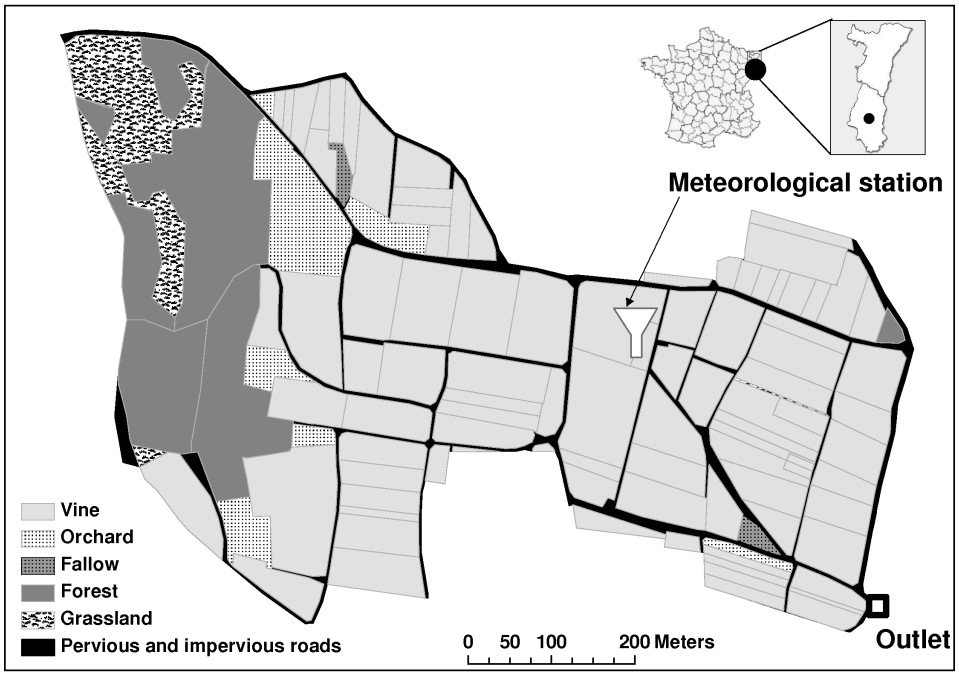


FIG. 4.1 – Hohrain catchment

At the bottom of the catchment lies the Venturi channel (Figure 4.2(a)) measuring the outlet flow where a drainage area (Figure 4.2(c)) was built. The purpose of this area is to collect the water drained from recent rain events and store it in order to i) analyse the water in term of polluting particles issued from the agriculture (See Figure 4.2(b)) ii) purify the water using plants and sand barriers in the drainage area(Figure 4.2(c)).

(a) Outlet flow measurements

(b) Pollution measurement samples

(c) Drainage area

(d) Vineyards

FIG. 4.2 – Considered catchment photos

The far end goal of the presented catchment, is the understanding of the rainfall/pesticides relationship in agricultural regions. It can be however stated that the drainage of polluting components depends on the pesticides spread by farmers and the outlet flow. Therefore, the runoff forecasting is needed and consequently the identification of the rainfall/runoff relationship is intended as a first step to reach the final intended rainfall/pesticides modelling.

Therefore, the identification problem in the present case is the following :
Given the raw rainfall data u and the outlet flow data y sampled at times t_k $k = 1..N$, the goal is to estimate the rainfall/runoff relationship. In the given case, the sample period is 6 minutes, the flow unit is expressed in l/s and the rainfall is expressed in mm. The data acquired during the year 2008 is shown in Figure 4.3.

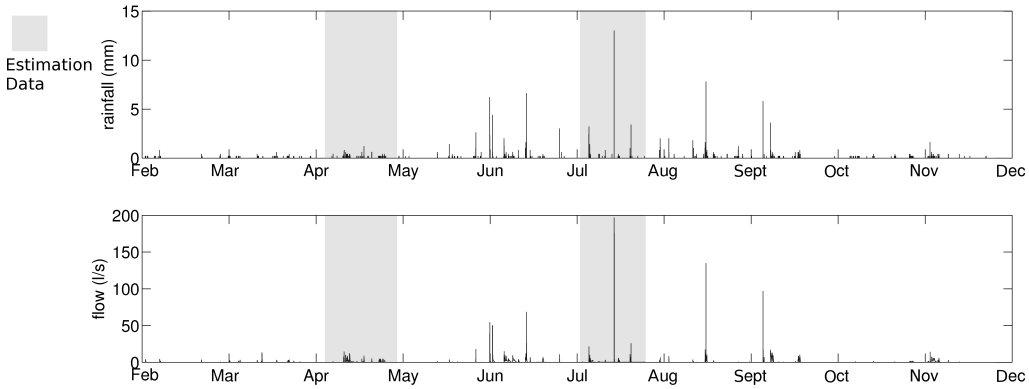


FIG. 4.3 – Rainfall and flow data for the Hohrain catchment during 2008

Over the 70000 samples acquired, only 5000 are over 0.3 l/s and are relevant for the identification process. The data has to be split into some identification data on which the identification algorithm will be run and some validation data to validate the model studied. The identification data needs to include the widest possible dynamic range and therefore needs to incorporate both low rainfall events and strong rainfall events. Moreover, both the identification and validation data sets need to contain enough informative data. In order to fulfill both characteristics and for all the presented models, the identification data set is exposed in grey in Figure 4.3 and the validation data is chosen as being the whole data set. The accuracy of a model is computed using the NASH coefficient defined as [NS70] :

$$\text{NASH} = 1 - \frac{\frac{1}{N}\sum_{k=1}^{N}\left[\hat{y}(t_k) - y(t_k)\right]^2}{\text{var}(y(t_k))}. \tag{4.1}$$

The data acquired from this experimental site present many challenging problems in terms of identification :

– As previously said, over the 70000 samples acquired, only 5000 are over 0.3 l/s and are relevant for the identification process and there is therefore a lack of excitation.
– The uncommon measurement noise characteristics. The presented catchment is dry and when it does not rain, there is no outlet flow and therefore the noise variance is null. When the outlet flow is rising however, the error in the Venturi channel captor measuring the flow rises accordingly. It

can be consequently concluded that the variance of the noise varies in time according to the flow intensity.

– During long or strong rainfall events, some particles get stuck in the Venturi channel and lure it into false measurements until it is cleaned as shown in Figure 4.4. This property prevents from using any statistical distribution as a representation for the noise.
– The difficulty to characterize the nonlinearity for rural catchments : in rural catchments, there is a high spatio-temporal variability of the soil property whether it lies in the vegetation, in the soil type or evapotranspiration [YG06] and there is a high difference between the raw and efficient rainfall. In this case, linear models most often completely fail in delivering a satisfying rainfall/flow relationship. It might be added that in the considered catchment, the evapotranspiration can be neglected due to the fast reaction time of the system (a few minutes).

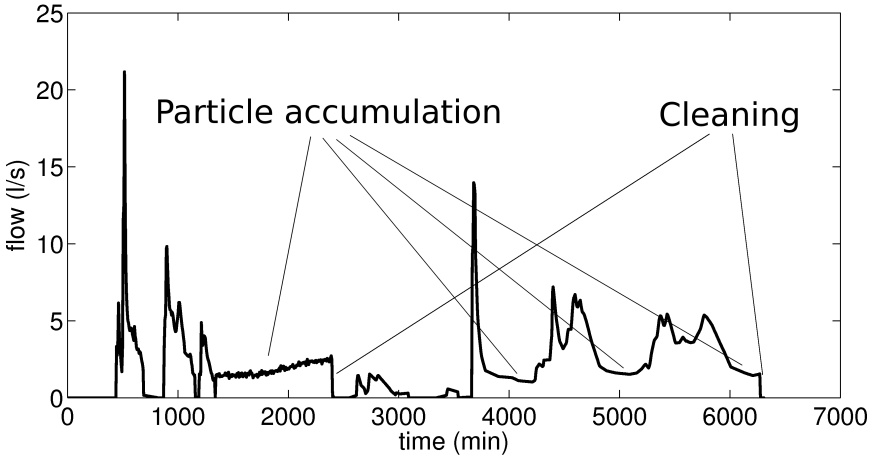


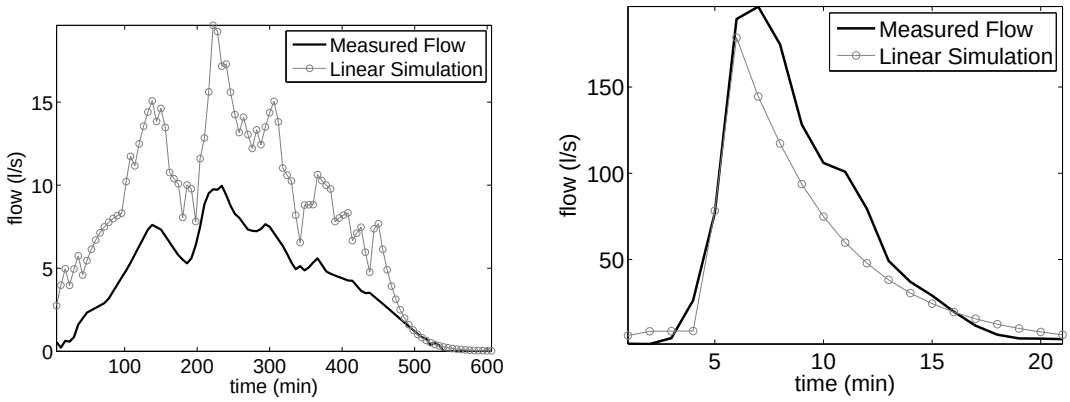
FIG. 4.4 – Effect of the particles accumulation on the measured data

4.3 Data-based linear modelling

Linear models are not expected to deliver suitable results for the estimation of rainfall/runoff relationship in rural catchments but they highlight the modelling problems for the considered catchment. Up to this section, the identification problem was addressed under the hypothesis that the different orders of the system are *a priori* known. Naturally it is not the case in this application and therefore an order search procedure has to be used. Both the orders n_b and n_a and the delay d of the model are computed in an automated manner by identifying linear OE type of models and minimizing the Young's Information Criterion (YIC) defined as [YJ80] :

$$\mathrm{YIC} = \log \frac{\mathrm{var}(\hat{y}(t_k) - y(t_k))}{\mathrm{var}(y(t_k))} + \log \frac{1}{n_\theta} \sum_{i=1}^{n_\theta} \frac{\hat{p}_{ii}}{\hat{\rho}_i} \tag{4.2}$$

where t_k belongs to the validation set, p_{ii} is the i^{th} diagonal element of the covariance matrix $\hat{\mathbf{P}}_\rho$ associated to ρ. The estimation method used is the `SRIV` algorithm [GW08]. Minimizing the YIC criterion is equivalent to jointly maximize the *NASH* coefficient and encourage the parsimony of the model. The results obtained for this catchment are $n_b = 1$, $n_a = 1$ and $d = 1$. According to the



(a) Linear model results for a light rainfall event *(b)* Linear model results for a strong rainfall event

FIG. 4.5 – Linear modelling results

experts' knowledge, this model structure is highly probable due to the small size of the catchment.

The linear model considered in this section can therefore be written as the following OE model :

$$\mathcal{L}\left\{y(t_k) = \frac{b_0 q^{-1}}{1+a_1 q^{-1}} u(t_k) + e(t_k), \right. \tag{4.3}$$

which is estimated using the `SRIV` algorithm.
The NASH coefficient obtained for the proposed model $\mathcal{L}$ is $\text{NASH}_{\mathcal{L}} = 0.63$. It clearly indicates that the model is unable to accurately fit the validation data as it cannot model the change of dynamics and gain in the system according to the amount of rain. The same conclusion can be driven by looking at the data from Figure 4.5 : the linear model response is too low for strong rainfall events and too strong for light rainfall events.

It has to be nonetheless pointed out that such a simple linear structure can deliver results accurate enough in urban catchments. The results of the data-based linear rainfall/flow modelling using the `SRIV` method have been published in [KLG+09] for an urban catchment situated in Brittany, France.

4.4 Data-based Hammerstein modelling

As a next step, the Hammerstein model structure is considered as it remains the mostly used model structure for rural catchments in the data-based mechanistic framework. In this nonlinear context, not only the orders of the system have to be determined, but also the nonlinearity parameters. The nonlinearity is here chosen as a polynomial function as it is mostly done in the literature. It has to be pointed out that usually, one of the weakness of the polynomial approximation is its weak extrapolation capability outside the domain of study. This motivates again the importance of including the strongest rainfall event in the identification process. A first assumption made here is that the linear model is an approximation of a more complex Hammerstein model. Consequently, the orders n_b, n_a and delay d of the linear dynamic part of the Hammerstein model are considered

to be the same as for the linear model $\mathcal{L}$. The order of the polynomial nonlinearity is afterwards optimized empirically by choosing the model offering the best complexity/accuracy tradeoff based on the NASH coefficient value.
For the considered catchment, the model obtained is given as the Hammerstein OE model :

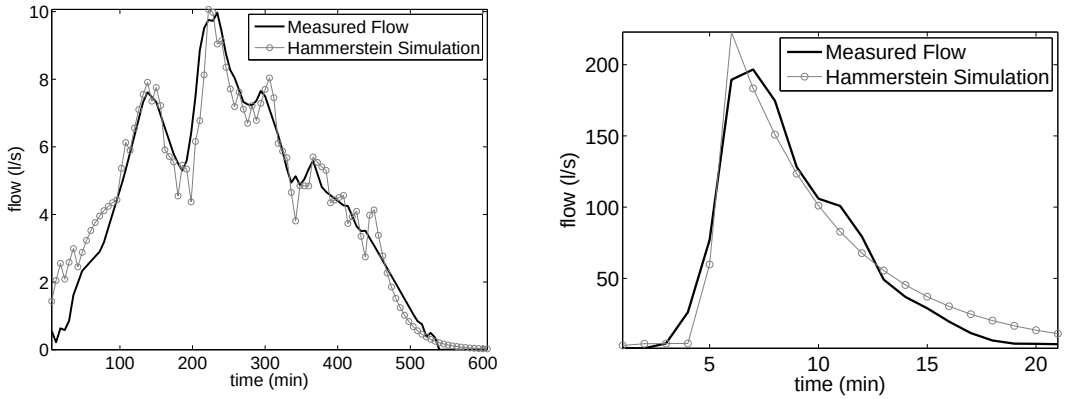
$$\mathcal{H}\begin{cases}\bar{u}(t_k) = u(t) + \alpha_1 u^2(t_k) + \alpha_2 u^3(t_k)\\ \chi(t_k) = \frac{b_0 q^{-1}}{1+a_1 q^{-1}}\bar{u}(t_k)\\ y(t_k) = \chi(t_k) + e(t_k)\end{cases} \tag{4.4}$$

which is estimated using the `HSRIV` method such as presented in Section 2.2.
The NASH coefficient obtained for the proposed model $\mathcal{H}$ is $\text{NASH}_{\mathcal{H}} = 0.81$. It clearly indicates the improvement of such model over a linear model. The model is now able to cope with the behavior changes in the system, whether a strong or a weak rainfall event is considered. This actually corroborates the accepted view of the catchment as a sponge of constant capacity which retains the water according to the moisture of the field. If the rain event is strong, and therefore if the ground is soaked, a larger amount of water reaches the outlet whereas if the soil is dry, the rain is mostly retained by the soil.
Figure 4.6 represents the output of the Hammerstein model. It can be seen from Figure 4.6(a) that for the weak rainfall event from the validation set, the model fits very well the measured data. For the strong rainfall event used in the estimation data set in Figure 4.6(b), it can be seen that the output of the Hammerstein model fits much better than the output of the linear model as expected. Nonetheless, one of the drawback of the Hammerstein models can be pointed out from this example. Due to the Hammerstein structure, only the gain can vary in time while the dynamic part is fixed. Naturally, one can argue that on this example, there is some measurement problem and that the output of the Hammerstein model is closer to the truth than the measured data. Nonetheless, from a general point of view, if the dynamic of the system varies over time, then the Hammerstein structure cannot cope with these changes. Therefore, the next section shows the application results of a generic scheme for the identification of the rainfall/runoff relationship using full LPV models which can deal with a wider range of system behaviours.

4.5 Data-based LPV modelling

As previously mentioned, the use of LPV models is pretty recent in the field of identification. There are still many open issues for the identification of LPV models, where most of the time, all the needed orders and scheduling variables are *a priori* known. However, these assumptions are violated here. In the linear case, only the orders and delay have to be searched for. In the Hammerstein case, the nonlinearity parameters have to be defined on top of the linear block parameters. In the LPV case, the problem becomes even harder. Naturally, both the nonlinear dependence on the scheduling variables and the system orders have to be defined. However, the most crucial choice concerns the scheduling variable definition : what do the model parameters depend on ?

When considering that none of these are *a priori* known like in the present application, the joint search for dependency, nonlinearity and orders might become impossible. Consequently, the main



(a) Hammerstein model results for a light rainfall event

(b) Hammerstein model results for a strong rainfall event

FIG. 4.6 – Hammerstein modelling results

goal is to propose a whole generic identification scheme for the problem of rainfall/runoff modelling. Before starting the whole process, the first assumption is that the orders n_a, n_b and the delay d are the same as for the linear model using the same justification as in the Hammerstein case.

4.5.1 Choice of the scheduling variables

The choice of the scheduling variables is crucial, and the possibilities for the scheduling parameters are theoretically endless. Nonetheless, their choice is constrained by :
– The physical understanding of the system studied.
– The available data.
Choosing some scheduling dependence based on specialist knowledge is the solution which most probably leads to some interpretable model and which therefore fits the best the data-based mechanistic framework considered here. The experts' knowledge states that the system behavior changes with the moisture of the catchment. Nonetheless, such a measure is unaccessible in practice and in the considered catchment, the only accessible data are the total rainfall and the outlet flow. For the determination of the scheduling variables, the work presented here is inspired by the recent work of P. Young. In [You02], the author proposes an Hammerstein structure (see Fig. 1.4) where the non-linearity depends on the measured flow in a forecasting context : this hypothesis reflects the fair assumption that the outlet flow is well correlated with the moisture of the catchment. More recently (see [You03] and the prior references therein), P. Young proposed a model for simulation in which the nonlinearity depends on output of a linear tank model obtained from the data with some successful interpretation capability.

As a main contribution for the considered application, the use of two scheduling variables is proposed. The first scheduling variable is chosen as in [You03] as the simulated outlet flow obtained from the linear model $\mathcal{L}$ presented in Section 4.3 and is denoted as $\hat{y}_{\mathrm{m}}(t_k)$. This variable physically represents the moisture at the bottom of the field where the flow is measured.

The second scheduling variable is computed by claiming the fair statement that "the more it has rained, the moister the soil is". This measure physically represents the moisture of the field on top of the catchment, where the rain is measured. In order to incorporate some memory information, the average of the past rainfall is considered :

$$\bar{u}(t_k) = \frac{1}{\Delta} \sum_{i=0}^{\Delta} u(t_{k-i}). \tag{4.5}$$

The window size for the sum is represented by the parameter Δ which has to be defined. The next section describes how to choose this user parameter.

4.5.2 Determination of the window size Δ

In order to find a suitable window size Δ, the dependence on the scheduling variables is momentarily considered to be linear. The orders of the system are considered to be the same as for the linear model and therefore $n_a = 1$, $n_b = 1$ and the delay $d = 1$. Consequently, Δ is chosen as the minimum number of samples maximizing the NASH coefficient for the LPV model :

$$\mathcal{LPV}_\Delta \begin{cases} A(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1})\chi(t_k) = B(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1})u(t_k) \\ y(t_k) = \chi(t_k) + e(t_k) \\ \text{with} \\ A(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1}) = 1 + \big(a_{1,0} + a_{1,1}\bar{u}(t_k)\big)q^{-1} \\ B(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1}) = \big(b_{1,0} + b_{1,1}\bar{u}(t_k)\big)q^{-1}, \\ \bar{u}(t_k) = \dfrac{1}{\Delta} \displaystyle\sum_{i=0}^{\Delta} u(t_{k-i}). \end{cases} \tag{4.6}$$

By searching Δ in this way, the implicit assumption is made that "the better the scheduling variable, the better the output of the model fits the measured data", which is a fair assumption. In this catchment, Δ was estimated to 20 samples (*e.g.* 2 hours) which is reasonable according to the experts' knowledge for a catchment of this size.

4.5.3 Determination of the nonlinear scheduling dependence order

At this stage, it is considered that the system orders are known and $n_a = 1$, $n_b = 1$, $d = 1$. The summing window size has been chosen as $\Delta = 20$ and therefore both scheduling variables $\bar{u}$ and $\hat{y}_{\mathrm{m}}$ are accessible. The remaining challenge is to determine the nonlinear dependence on the scheduling variables. A polynomial dependence on the scheduling variables is considered and therefore, one is finally interested in finding the orders n_α and n_β maximizing the NASH coefficient for the following

model :

$$\mathcal{LPV}_{n_\alpha,n_\beta}\begin{cases} A(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1})\chi(t_k) = B(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1})u(t_k) \\ y(t_k) = \chi(t_k) + e(t_k)\text{with} \\ A(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1}) = 1 + \big(a_{1,0} + \sum_{i=1}^{n_\alpha} a_{1,i}\hat{y}_{\mathrm{m}}^i(t_k) + \sum_{j=1}^{n_\beta} \mathrm{a}_{1,j}\bar{u}^j(t_k)\big)q^{-1} \\ B(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1}) = \big(b_{1,0} + \sum_{i=1}^{n_\alpha} b_{1,i}\hat{y}_{\mathrm{m}}^i(t_k) + \sum_{j=1}^{n_\beta} \mathrm{b}_{1,j}\bar{u}^j(t_k)\big)q^{-1}, \end{cases} \tag{4.7}$$

It must be pointed out, that for this particular catchment, raising the values of n_α and n_β does not bring any significant improvement in the estimated model. Therefore, the most parsimonious model is chosen and n_α and n_β are set to 1.

4.5.4 Final LPV model

The identification scheme introduced here for the data-based LPV modelling of the rainfall/runoff relationship is summarized in Figure 4.7.

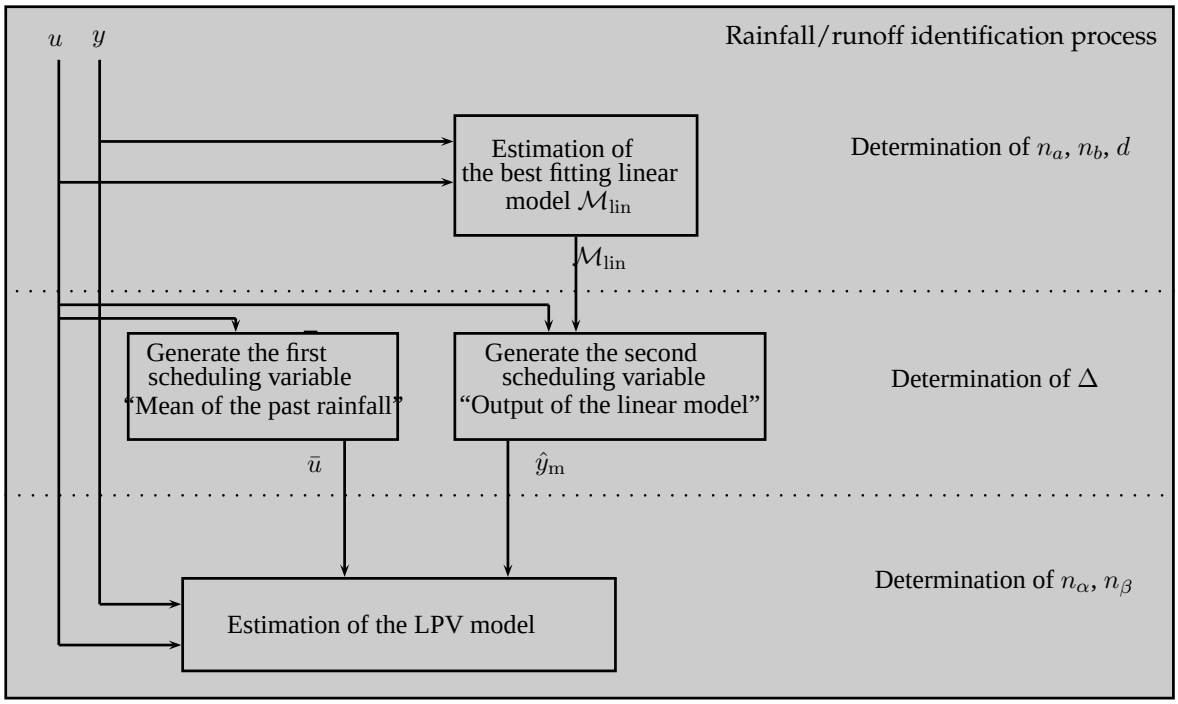
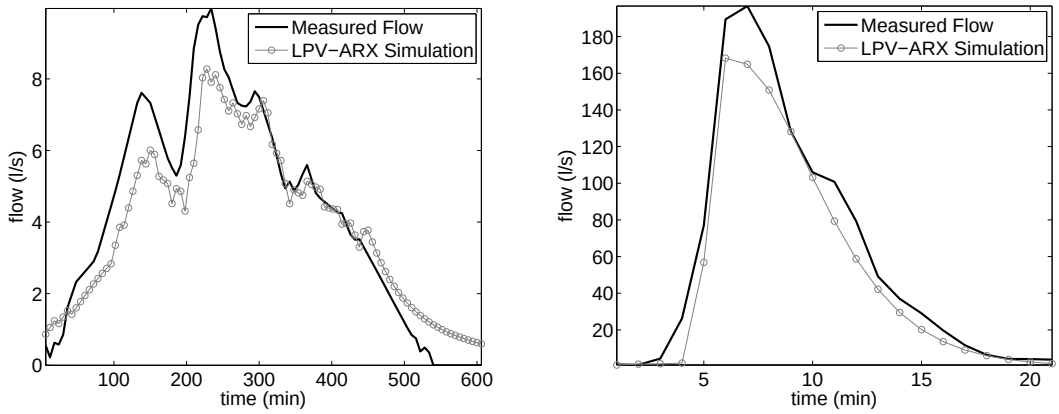


FIG. 4.7 – Proposed data-based LPV modelling scheme for the rainfall/runoff relationship

This proposed identification scheme leads to the DT LPV-OE model considered as :

$$\mathcal{LPV}_{\mathrm{OE}}\begin{cases} A(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1})\chi(t_k) = B(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1})u(t_k) \\ y(t_k) = \chi(t_k) + e(t_k)\text{with} \\ A(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1}) = 1 + \big(a_{1,0} + a_{1,1}\hat{y}_{\mathrm{m}}(t_k) + a_{1,2}\bar{u}(t_k)\big)q^{-1} \\ B(\hat{y}_{\mathrm{m}}(t_k),\bar{u}(t_k),q^{-1}) = \big(b_{1,0} + b_{1,1}\hat{y}_{\mathrm{m}}(t_k) + b_{1,2}\bar{u}(t_k)\big)q^{-1}, \end{cases} \tag{4.8}$$



(a) LPV-ARX model results for a light rainfall event

(b) LPV-ARX model results for a strong rainfall event

FIG. 4.8 – LPV-ARX modelling results

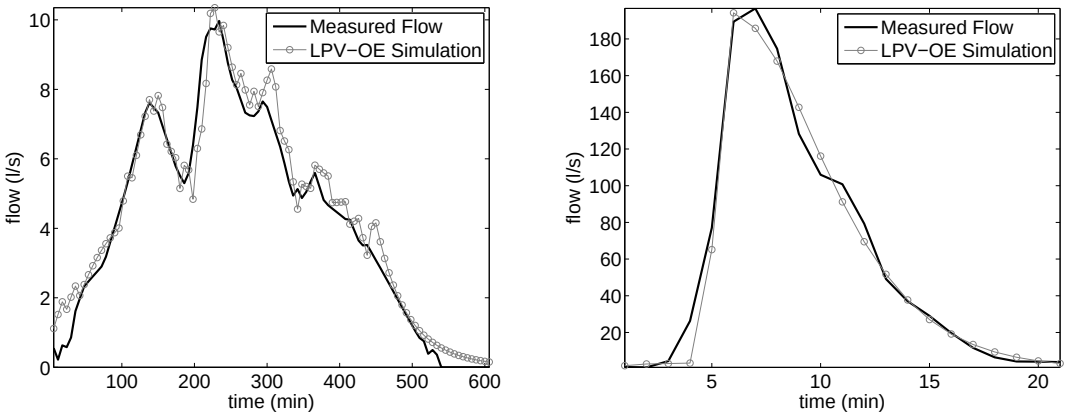
which is estimated using the `LPV-SRIV` method such as presented in Section 3.2.
In order to compare the proposed algorithm `LPV-SRIV` to the mostly used method `LPV-LS` method the following LPV-ARX model is considered for comparison :

$$\mathcal{LPV}_{\mathrm{ARX}} \begin{cases} A(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1})y(t_k) = B(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1})u(t_k) + e(t_k) \\ \text{with} \\ A(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1}) = 1 + \big(a_{1,0} + a_{1,1}\hat{y}_{\mathrm{m}}(t_k) + a_{1,2}\bar{u}(t_k)\big)q^{-1} \\ B(\hat{y}_{\mathrm{m}}(t_k), \bar{u}(t_k), q^{-1}) = \big(b_{1,0} + b_{1,1}\hat{y}_{\mathrm{m}}(t_k) + b_{1,2}\bar{u}(t_k)\big)q^{-1}, \end{cases} \tag{4.9}$$

and is estimated using the `LPV-LS` method such as presented in [BG02b]. It must be pointed out that both methods use the same process model, the same estimation data set. The NASH coefficients obtained for both models are :

– $\mathrm{NASH}_{\mathcal{LPV}_{\mathrm{OE}}} = 0.84$
– $\mathrm{NASH}_{\mathcal{LPV}_{\mathrm{ARX}}} = 0.81$.

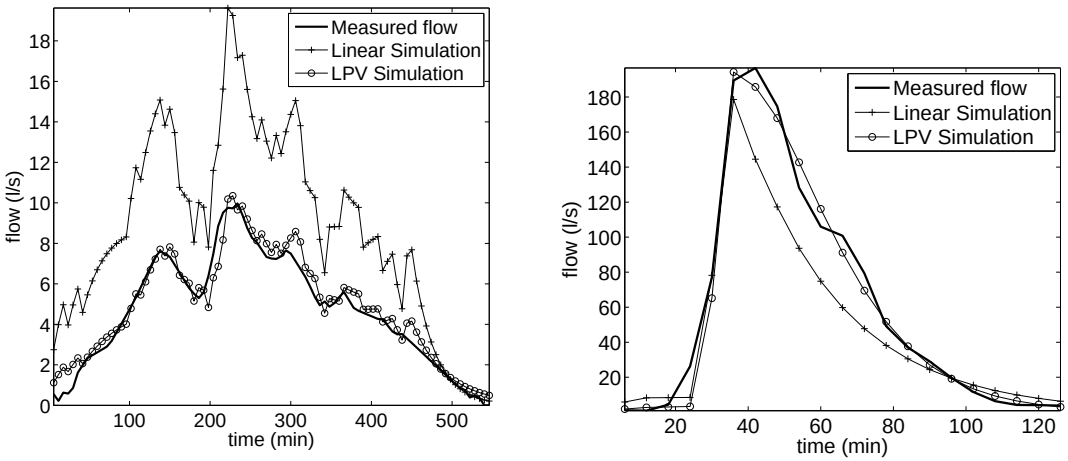
Naturally, both LPV models allowing dynamic and gain variations, the estimated models $\mathcal{LPV}_{\mathrm{ARX}}$ and $\mathcal{LPV}_{\mathrm{OE}}$ perform much better than their linear truncation. The NASH coefficient shows a slight superiority of the `LPV-SRIV` algorithm over the `LPV-LS` method on this catchment. The difference between both model responses can be better appreciated when looking at Figure 4.8 and Figure 4.9 and the improvement when using the SRIV-based algorithm can be clearly noticed on the strong rainfall event. For both LPV models, the assumed noise model is wrong. Nevertheless, it seems the LPV-SRIV algorithm copes better with this problem as expected from the theory.
It can be seen from the strong event results that both LPV models can cope with a dynamical change of the system. Again, this strong event alone does not justify the use of an LPV over an Hammerstein structure for this given catchment. Nonetheless, from a general point of view, it shows that the LPV model achieves a finer approximation. Just like in the Hammerstein case, using an LPV model approximates the catchment as being a sponge in which the efficient rainfall is correlated to the moisture of the field. In the LPV case however (in opposition to the Hammerstein case), the



(a) LPV-OE model results for a light rainfall event

(b) LPV-OE model results for a strong rainfall event

FIG. 4.9 – LPV-OE modelling results



(a) LPV and linear model results for a light rainfall event

(b) LPV and linear model results for a strong rainfall event

FIG. 4.10 – Comparison of linear and LPV OE models for light and strong rainfall events

TAB. 4.1 – NASH coefficients for the different models

Model	NASH coefficient
$\mathcal{L}$	0.63
$\mathcal{H}$	0.81
$\mathcal{LPV}_{\text{ARX}}$	0.81
$\mathcal{SLPV}_{\text{OE}}$	0.84

speed at which the water reaches the outlet is allowed to change over time as well which can be interpreted as a higher degree of freedom. Finally and for comparison purposes, the outputs of the linear model $\mathcal{L}$ and LPV model $\mathcal{LPV}_{\text{OE}}$ are shown simultaneously on Figure 4.10. The better fit of the LPV model output can be clearly observed.

4.6 Conclusion

In this chapter, the results of the proposed methods have been analysed in the intended application field : the identification of the rainfall/runoff relationship in a rural catchment. The presented catchment usually poses some problems to the hydrologist conceptual models according to the experts. Moreover, the measured signals are challenging for identification purposes as the excitation is very low and the noise cannot be expressed in a stochastic way. The NASH coefficients for all proposed models are summarized in Table 4.1.

It can be seen that in the given context, as expected from the theory, the linear models are too limited for the modelling of rural catchments. From the NASH coefficients, it can be concluded that all nonlinear models performs similarly. Nonetheless, when looking at the data series, it was shown that the fixed dynamic structure of the Hammerstein models can be a drawback in such an application. The SRIV-based algorithm outperforms the existing method based on the least squares, which was up to now the most used algorithm for the estimation of LPV models.

The main contribution to this application field is the introduction of a global identification scheme for the data-based LPV modelling of rainfall/runoff relationship at the catchment scale. Moreover, some scheduling variables have been introduced and their correlation with the moisture of the field seems to be verified from the data. Furthermore, the LPV model structure is a good candidate in this data-based mechanistic framework :

– It can easily integrate some *a priori* knowledge (some scheduling variables issued from expert's knowledge or logic).
– It can easily be interpreted in terms of common features of the time varying linear model obtained (varying static gain or varying time constant for example).

It has to be pointed out that the same scheme can be applied using some continuous-time models. Nonetheless, the results obtained using CT models have not been presented for the following reasons :

– The aim of this chapter was to present an applicable identification framework for the rainfall/runoff relationship modelling and not to put discrete-time and continuous-time models in competition.

– The results are very similar to the one obtained from the discrete-time models and displaying both results would be redundant.

The main advantage of the CT model would be the possible comparison with physical coefficients used by hydrologists. Nonetheless, the *a posteriori* interpretation of the model has not been realized so far and is intended as future research. Moreover, the final aim of the considered catchment is the understanding of pesticides transports and the present work constitutes a good basis for the far end goal of rainfall/pesticides relationship modelling.

Chapitre 5

Conclusion and perspectives

5.1 Conclusion

Nonlinear models represent a wide class of models which can largely vary in terms of complexity and representation capability. Therefore in order to make some substantial contributions to nonlinear system identification, the work described in this thesis was developed following the general process for system identification. By analysing some collected data and the main features of the system under study, two model structures of interest were considered : the Hammerstein model structure and the LPV model structure. Nonetheless, the contributions have to be well separated due to the clear difference in the advancement of the system identification framework with respect to these structures. While the Hammerstein structure has been widely studied in the system identification framework, the LPV model structure has not received much attention in the described framework. The third stage of the identification process is the choice of the minimization criterion in order to find the best model among the model set defined. The focus was put on the commonly accepted prediction error framework and more precisely on the refined instrumental variable method for its interesting properties issued from the LTI framework : i) the obtention of optimal estimates if the system belongs to the model set ii) the consistency of the estimates with reasonably low variance even if the assumed noise model is wrong iii) the included prefiltering of the data allowing some close estimates of the derivatives in a CT context.

In chapter 2, the main contribution lies in the development of an estimation method handling Hammerstein Box–Jenkins models. Moreover, the performance of the proposed method has been investigated in different frameworks such as the open-loop, the closed-loop or the continuous-time case. It appeared that the LTI features of the RIV straight forwardly applied to the Hammerstein case for the proposed modelling which lead to a LTI MISO representation of the Hammerstein structure. The drawbacks for choosing the proposed model, namely i) the overparameterisation of the system and ii) the approximation of the true nonlinear function by a sum of basis functions, were also investigated. On the one hand, the results of blind estimation showed the robustness of the proposed estimation method with respect to plant modelling error. On the other hand, the variances obtained at low SNR showed the robustness of the method under harsh noise conditions and when the noise is not correctly modelled. Finally, the advantages of direct continuous-time model

estimation were depicted through an example based on the Rao-Garnier Benchmark.

In chapter 3, the contribution lies in the development of identification methods for LPV models in presence of colored noise. It had to be firstly stated that the identification of LPV model is at a very immature state. It appeared as well that even if a unified system theoretical framework of LPV systems has recently been developed, LPV models still present a lot of open issues. Among them, a crucial issue is that most methods assume DT models while the usual nature of systems is expressed in CT. While the CT/DT transformation is well understood for LTI models, it has been shown that the same transformation can lead to the high complexification of the estimation process in the LPV framework. Moreover, it has been shown that the existing methods in the LPV I/O literature could not face the time varying nature of the optimal prefilter. It has been shown that the effect of these negligence can have critical consequences on the quality of the estimates. Therefore, the main contribution of this part is to propose some robust identification methods for realistic colored noise conditions. The outperformance of the presented method over the existing methods has been depicted through a set of examples. Finally, a method able to deal with the identification of CT I/O LPV models has been proposed and has been tested on a LPV benchmark deriving from the Rao-Garnier benchmark in which a "moving pole" has been introduced. Not only this method answers the needs exposed in the introduction part, but the introduction of this method opens a wide track about the possibility of fulfilling the needs for controller optimal design in a CT form.

In chapter 4, the relevance of the studied models and of the associated estimation methods has been investigated for the considered application field : the identification of the rainfall/runoff relationship in a rural catchment. On the one hand it has been shown that even if the Hammerstein model structure considerably improves the results in the forecasting of the rainfall/runoff relationship of a rural catchment, it has also a restrained representation capability due the fix dynamic of the model structure. On the other hand, LPV models are a very interesting model set candidate in the considered application. They exposed a very simple linear structure, and therefore can be interpreted using the usual criteria (poles, zeros, static gain) while they can model strong nonlinearities. The proposed estimation methods have exposed robustness when used on challenging data (poor excitation, non-stationarity of the noise, error-in-variable context). It can be therefore concluded that the combination LPV/RIV makes a good pair for practical applications. Moreover, it has to be pointed out that the data based modelling methods capable of capturing the behavior of natural environments while keeping a very parsimonious structure are a very innovating technique in the hydrology community. Consequently the contribution for this part is the introduction of relevant scheduling variable choice and some general identification framework for the rainfall/runoff relationship modelling.

Therefore, the contributions of this thesis can be summed up in the following way :

- The proposition of a refined instrumental method for the identification of discrete-time Hammerstein Box–Jenkins models
- The proposition of a refined instrumental method for the identification of discrete-time Hammerstein Box–Jenkins models operating in closed loop
- The proposition of a refined instrumental method for the identification of Hybrid Hammerstein

Box–Jenkins models and continuous-time Hammerstein Output error models
- The proposition of a refined instrumental method for the identification of discrete-time LPV Box–Jenkins models
- The proposition of a refined instrumental method for the identification of discrete-time LPV Box–Jenkins models operating in closed loop
- The proposition of a refined instrumental method for the identification of Hybrid LPV Box–Jenkins models and continuous-time LPV Output error models
- The proposition of a novel approach for the identification of the rainfall/runoff relationship in rural catchments

5.2 Perspectives

This thesis focused on the development of estimation methods dedicated to given types of nonlinear structures justified by the application field. Nevertheless, the presented work constitutes a basis for the development of estimation method in the following contexts :
- The estimation of other block structures.
- The estimation of LPV models with dynamic dependency on the scheduling parameters : it was shown that the discretisation of CT LPV models implies a certain number of complexities related to the dynamic dependency on the scheduling parameters which have not been addressed yet in the literature.
- The estimation of CT LPV models operating in closed-loop : such a method could open an interesting field in the development of optimal I/O LPV control design.
- The estimation of "full" LPV models including LPV noise model. The presented method can handle this case in a suboptimal way but no method from the literature can deal efficiently with this case.
- The estimation of MIMO I/O LPV models. Such a method would open a larger application field to the presented method in industrial context.

Moreover, many open issues well investigated in the LTI framework remain unexplored in the LPV framework such as :
- In the LPV domain, the questions concerning the persistency of excitation for such systems remain completely unanswered.
- The presented method performances were analysed from Monte Carlo simulations. Nonetheless it is worth to investigate the statistical properties of the proposed estimators from a theoretical point of view : for example, is `LPV-RIV` statistically optimal ?

Finally, in terms of application, data based modelling such as presented here is a nearly unexplored field for complex systems in experimental sciences such as environmental sciences. The presented application is a preliminary work for the estimation of the rainfall/pesticide relationship in rural catchments which needs the same features as the presented models :
- Good forecasting possibilities.
- Simple structure.

– The inclusion of *a priori* physical knowledge.

This work is intended for further research. Moreover, and in a more general way, a tremendous work is needed in order to expose to the environmental community the advantages of the proposed methods. This can only be done by jointly analysing the impact and the physical meaning of the estimated coefficients in order to be able to build hybrid environmental theories. The application of LPV models is endless in the environmental world, whether it deals with climatology, hydrology or geology and other type of data should be investigated.

Bibliographie

[AG95] P. Apkarian and P. Gahinet. A convex characterization of gain-scheduled $\mathcal{H}_\infty$ controllers. *IEEE Trans. on Automatic Control*, 40(5) :853–864, 1995.

[Atk89] K. E. Atkinson. *An Introduction to Numerical Analysis*. John Wiley and Sons, 1989.

[AW09] H. Abbas and H. Werner. An instrumental variable technique for open-loop and closed-loop identification of input-output LPV models. In *Proc. of the European Control Conf.*, pages 2646–2651, Budapest, Hungary, 2009.

[Bai98] E.W. Bai. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, 34, Issue 3 :333–338, March 1998.

[Bai02a] E-W. Bai. A blind approach to the Hammerstein-Wiener model identification. *Automatica*, 38, Issue 6 :967–979, 2002.

[Bai02b] E-W. Bai. Identification of linear systems with hard input nonlinearities of known structure. *Automatica*, 38, Issue 5 :853–860, May 2002.

[BC08] E-W. Bai and K-S. Chan. Identification of an additive nonlinear system and its applications in generalized Hammerstein models. *Automatica*, 44, Issue 2 :430–436, February 2008.

[Bev00a] K. J. Beven. *Rainfall-Runoff Modelling : The Primer*. John Wiley, Hoboken, NJ, 2000.

[Bev00b] K. J. Beven. Uniqueness of place and process representations in hydrological modelling. *Hydrology and earth system sciences*, 4 :203–213, 2000.

[BG99] B. Bamieh and L. Giarré. Identification of linear parameter varying models. In *Proceedings of the 38th IEEE Conference on Decision and Control*, pages 1505–1510, Phoenix, USA, Dec 1999.

[BG00] B. Bamieh and L. Giarré. Identification for a general class of LPV models. In *Proceedings of the 11th IFAC Symposium on System Identification*, Santa Barbara, California, USA, June 2000.

[BG02a] B. Bamieh and L. Giarré. Identification of linear parameter-varying models. *International Journal of Robust and Nonlinear Control*, 12 :841–853, 2002.

[BG02b] B. Bamieh and L. Giarré. Identification of linear parameter varying models. *Int. Journal of Robust and Nonlinear Control*, 12 :841–853, 2002.

[BGM01] A. Boukhris, S. Giuliani, and G. Mourot. Rainfall-runoff multi-modelling for sensor fault diagnosis. *Control Engineering Practice*, 9, Issue 6 :659–671, June 2001.

[BKL08] M. Butcher, A. Karimi, and R. Longchamp. On the consistency of certain identification methods for linear parameter varying systems. In *Proceedings of the 17th IFAC World Congress*, pages 4018–4023, Seoul, Korea, July 2008.

[BMD09] G. Bastin, L. Moens, and P. Dierickx. Online river flow forecasting with hydromax : successes and challenges after twelve years of experience. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 6-8 2009.

[BW08] S. Boonto and H. Werner. Closed-loop system identification of lpv input-output models - application to an arm-driven inverted pendulum. In *Proceedings of the 47th IEEE conference on Decision and Control*, Cancun, Mexico, Dec 9-11 2008.

[CR03] V. Cerone and D. Regruto. Parameter bounds for discrete-time Hammerstein models with bounded errors. *IEEE Transaction on Automatic Control*, 48, Issue 10 :1855–1860, 2003.

[CSS08] M. C. Campi, T. Sugie, and F. Sakai. An iterative identification method for linear continuous-time systems. *IEEE transactions on automatic control*, 53, Issue 7 :1661–1669, august 2008.

[DAL02] F. De Bruyne, B. D. O. Anderson, and I. D. Landau. Recursive identification of nonlinear plants operating in closed loop using kernel representations. *Automatica*, 38, Issue 11 :2021–2027, November 2002.

[DC05] F. Ding and T. Chen. Identification of Hammerstein nonlinear ARMAX systems. *Automatica*, 41, Issue 9 :1479–1489, September 2005.

[DL80] L. Dugard and I. D. Landau. Recursive Output Error identification algorithms theory and evaluation. *Automatica*, 16, Issue 5 :443–462, 1980.

[Doo59] J. C. I Dooge. A general theory of the unit hydrograph. *Journal of geophysical research*, 64 :241–256, 1959.

[DS01] M. Dettori and C. W. Scherer. LPV design for a CD player : An experimental evaluation of performance. In *Proc. of the 40th IEEE Conf. on Decision and Control*, pages 4711–4716, Orlando, Florida, USA, Dec. 2001.

[DSC07] F. Ding, Y. Shi, and T. Chen. Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. *Systems & Control Letters*, 56, Issue 5 :373–380, 2007.

[dSRdC07] P. L. dos Santos, J. A. Ramos, and J. L. M. de Carvalho. Identification of linear parameter varying systems using an iterative deterministic-stochastic subspace approach. In *Proc. of the European Control Conference*, pages 4867–4873, Kos, Greece, July 2007.

[FAD01] K. Fujimoto, B. D. O. Anderson, and F. De Bruyne. A parametrization for closed-loop identification of nonlinear systems based on differentially coprime kernel representations. *Automatica*, 37, Issue 12 :1893–1907, December 2001.

[FE10] F. Giri and E-W. Bai (Eds). *Block-oriented nonlinear system identification*. Springer-Verlag, London, 2010.

[FJ98] D. A. K. Fernando and A. W. Jayawardena. Runoff forecasting using rbf networks with OLS algorithm. *ASCE Journal of hydrological engineering*, 3 :203–209, 1998.

[FL99] U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35(7) :1215–1241, 1999.

[GBBM09] M. Gevers, A.S Bazanella, X. Bombois, and L. Mišković. Identification and the information matrix : how to get just sufficiently rich ? *IEEE Transactions on Automatic Control*, 54, Issue 12 :2828–2840, December 2009.

[GBFB06] L. Giarré, D. Bauso, P. Falugi, and B. Bamieh. LPV model identification for gain scheduling control : An application to rotating stall and surge control problem. *Control Engineering Practice*, 14 :351–361, 2006.

[GdH05] M. Gilson and P. Van den Hof. Instrumental variable methods for closed-loop system identification. *Automatica*, 41, Issue 2 :241–249, 2005.

[Gev93] M. Gevers. *Towards a joint design of identification and control ?*, chapter 5, pages 111–151. H.L. Trentelman and J.C. Willems, Birkhäuser - Boston, 1993. Essay on Control : perspectives in the theory and its applications.

[Gev05] M. Gevers. Identification for control : from the early achievements to the revival of experiment design. *European Journal of Control*, 11(4-5) :335–352, 2005.

[Gev06] M. Gevers. A personal view of the development of system identification. *IEEE Control systems magazine*, 26, Issue 6 :93–105, December 2006.

[GGS00] G. C. Goodwin, S. F. Graebe, and M. E. Salgado. *Control System Design*. Prentice Hall, 2000.

[GGYdH09] M. Gilson, H. Garnier, P. C. Young, and P. Van den Hof. Refined instrumental variable methods for closed-loop system identification methos. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, June 2009.

[GL09] J. Gillberg and L. Ljung. Frequency domain identification of continuous-time ARMA models from sampled data. *Automatica*, 45, Issue 6 :1371–1378, June 2009.

[GL10a] J. Gillberg and L. Ljung. Frequency domain identification of continuous-time output error models, part I : Uniformly sampled data and frequency function approximation. *Automatica*, 46, Issue 1 :1–10, January 2010.

[GL10b] J. Gillberg and L. Ljung. Frequency domain identification of continuous-time output error models, part II : Non-uniformly sampled data and B–spline output approximation. *Automatica*, 46, Issue 1 :11–18, January 2010.

[GMR03] H. Garnier, M. Mensler, and A. Richard. Continuous-time model identification from sampled data : implementation issues and performance evaluation. *International Journal of Control*, 76, 2003.

[Gre00] W. Greblicki. Continuous-time hammerstein system identification. *IEEE Transactions on Automatic Control*, 45, Issue 6 :1232–1236, June 2000.

[GS01] G. B. Giannakis and E. Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81, Issue 3 :533–580, March 2001.

[GVdH05] M. Gilson and P. Van den Hof. Instrumental variable methods for closed-loop system identification. *Automatica*, 41(2) :241–249, 2005.

[GW08] H. Garnier and L. Wang (Editors). *Identification of Continuous-time Models from Sampled Data*. Springer-Verlag, London, March 2008.

[JH93] A. J. Jakeman and G. M. Hornberger. How much complexity is warranted in a rainfall-runoff model ? *Water Resources Research*, 29 :2637–2649, 1993.

[JLW90] A. J. Jakeman, I. G. Littlewood, and P. G. Whitehead. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of hydrology*, 117 :275–300, 1990.

[Joh94] R. Johansson. Identification of Continuous–Time models. *IEEE Transactions on Signal Processing*, 42, 1994.

[Kle83] V. Klemes. Conceptualization and scale in hydrology. *Journal of hydrology*, 65 :1–23, 1983.

[KLG+09] D. Kuss, V. Laurain, H. Garnier, M. Zug, and J. Vazquez. Data-based mechanistic rainfall-runoff continuous-time modelling in urban context. In *proceedings of the 15th IFAC Symposium on System Identification*, pages 1780–1785, Saint-Malo, France, 3-6 July 2009.

[LAD01] I. D. Landau, B. D. O. Anderson, and F. De Bruyne. Recursive identification algorithms for continuous-time nonlinear plants operating in closed loop. *Automatica*, 37, Issue 3 :469–475, March 2001.

[LB07] Y. Liu and E. W. Bai. Iterative identification of Hammerstein systems. *Automatica*, 43, Issue 2 :346–354, February 2007.

[LG74] B. Lorent and M. Gevers. Identification of rainfall/runoff processes. *Proceedings of the 4th IFAC Symposium on Identification and parameter Estimation*, pages 735–744, 1974.

[LGG09] V. Laurain, M. Gilson, and H. Garnier. Refined instrumental variable methods for identifying hammerstein models operating in closed loop. In *Proceedings of the 48th IEEE Conference on Decision and Control*, Shanghai, China, December 2009.

[LGGY08] V. Laurain, M. Gilson, H. Garnier, and P. C. Young. Refined instrumental variable methods for identification of Hammerstein continuous time Box-Jenkins models. In *proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, December 2008.

[LGTG10a] V. Laurain, M. Gilson, R. Tóth, and H. Garnier. Identification of lpv box-jenkins models via optimal refined instrumental variable methods. In *Proceedings of the American Control Conference 2010*, page invited session "Identification of LPV models", Baltimore, Maryland, USA, June 2010.

[LGTG10b] V. Laurain, M. Gilson, R. Tóth, and H. Garnier. Refined instrumental variable methods for identification of LPV output-error and Box-Jenkins models. *Automatica*, 46, Issue 6 :959–967, June 2010.

[Lju99] L. Ljung. *System identification : theory for the user - Second Edition*. Prentice-Hall, 1999.

[Lju03] L. Ljung. Initialisation aspects for subspace and output-error identification methods. In *European Control Conference (ECC'2003)*, Cambridge (U.K.), September 2003.

[Lju09] L. Ljung. Experiments with identification of continuous time models. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

[LK97] I. D. Landau and A. Karimi. An output error recursive algorithm for unbiased identification in closed loop. *Automatica*, 33, Issue 5 :933–938, May 1997.

[LM07] M. Lovera and G. Mercère. Identification for gain-scheduling : a balanced subspace approach. In *Proc. of the American Control Conference*, pages 858–863, New York City, USA, July 2007.

[LTGG10] V. Laurain, R. Tóth, M. Gilson, and H. Garnier. Direct identification of continuous-time lpv input/ouput models. *Submitted to special issue "Continuous-time model Identification", IET Control Theory & Applications*, april 2010.

[MB04] A. Marcos and G. J. Balas. Development of linear-parameter-varying models for aircraft. *Journal of Guidance, Control and Dynamics*, 27(2) :218–228, 2004.

[MP96] S. Mambretti and A. Paoletti. A new approach in overland flow simulation in urban catchments. In *proceedings of the 7th International Conference on Urban Storm Drainage*, Hannover, Germany, 1996.

[MSJS99] R. Murray-Smith, T. A. Johansen, and R. Shorten. On the interpretation of local models in blended multiple model structures. *Int. Journal of Control*, 72(7-8) :620–628, 1999.

[Nel01] O. Nelles. *Nonlinear system identification*. Springer-Verlag, Berlin, 2001.

[NS70] J. E. Nash and J. V. Stucliffe. River flow forecasting through conceptuel models. part 1-a discussion of principles. *Journal of Hydrology*, 10, Issue 3 :282–290, 1970.

[ØSB08] K. Z. Østergaard, J. Stousturp, and P. Barth. Rate bounded linear parameter varying control of a wind turbine in full load operation. In *Proc. of the 17th IFAC World Congress*, Seoul, Korea, July 2008.

[PL09] F. Previdi and M. Lovera. Identification of parametrically- varying models for the rainfall-runoff relationship in urban drainage networks. In *proceedings of the 15th IFAC Symposium on System Identification*, pages 1768–1773, Saint-Malo, France, 3-6 July 2009.

[PLM99] F. Previdi, M. Lovera, and S. Mambretti. Identification of the rainfall-runoff relationship in urban drainage networks. *Control Engineering Practice*, 7, Issue 12 :1489–1504, 1999.

[PMEB+07] H. J. Palanthandalam-Madapusi, B. Edamana, D. S. Bernstein, W. Manchester, and A. J. Ridley. Narmax identification for space weather prediction using polynomial radial basis functions. *46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 2007.

[PSR00] R. Pintelon, J. Schoukens, and Y. Rolain. Box–Jenkins continuous–time modeling. *Automatica*, 36, 2000.

[PU95] A. Patra and H. Unbehauen. Identification of a class of nonlinear continuous-time systems using Hartley modulating functions. *International Journal of Control*, 62, Issue 6 :1431–1451, December 1995.

[RG02] G. P. Rao and H. Garnier. Numerical illustrations of the relevance of direct continuous-time model identification. In *15th Triennial IFAC World Congress on Automatic Control*, Barcelona, Spain, 2002.

[RG04] G. P. Rao and H. Garnier. Identification of continuous-time systems : direct or indirect ? *Systems Science*, 30, Issue 3 :25–50, 2004.

[RH06] G.P Rao and H.Unbehauen. Identification of continuous-time systems. *IEE Proceedings on Control Theory Appl.*, 153(2), March 2006.

[SÖ7] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43, Issue 6 :939–958, June 2007.

[SA92] J. S. Shamma and M. Athans. Gain scheduling : potential hazards and possible remedies. *IEEE Control Systems Magazine*, 12, Issue 3 :101–107, 1992.

[Sch96] C. W. Scherer. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control for time-varying and linear parametrically-varying systems. *Int. Journal of Robust and Nonlinear Control*, 6(9-10) :929–952, 1996.

[SJ00] P. Stoica and M. Jansson. MIMO system identification : state-space and subspace approximations versus transfer function and instrumental variables. *IEEE Transaction on Signal Processing*, 48(11) :3087 – 3099, 2000.

[SS83] T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification.* Springer-Verlag, New York, 1983.

[SS89] T. Söderström and P. Stoica. *System identification.* Prentice-Hall, 1989.

[SWGP07] J. Schoukens, W.D. Widanage, K.R. Godfrey, and R. Pintelon. Initial estimates for the dynamics of a Hammerstein system. *Automatica*, 43, Issue 7 :1296–1301, July 2007.

[TÓ8] R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems : An Orthonormal Basis Function Approach.* PhD thesis, Delft University of Technology, 2008.

[TDV91] P. A. Troch, F. P. De Troch, and J. Van Hyfte. *Modelling the time dependant natrue of the rainfall-runoff relationship using on-line identification.* In *Hydrological applications of weather radar,* I. D. Cluckie, C. G. Collier (eds), pages 519–530. E. Horewood : New York, 1991.

[TFHV07] R. Tóth, F. Felici, P. S. C. Heuberger, and P. M. J. Van den Hof. Discrete time LPV I/O and state space representations, differences of behavior and pitfalls of interpolation. In *Proc. of the European Control Conf.*, pages 5418–5425, Kos, Greece, July 2007.

[THV08] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof. Flexible model structures for LPV identification with static scheduling dependency. In *Proc. of the 47th IEEE Conf. on Decision and Control*, pages 4522–4527, Cancun, Mexico, Dec. 2008.

[THV09] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof. An LPV identification framework based on orthonormal basis functions. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

[THV10] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof. On the discretization of LPV state-space representations. *Accepted to IET Control Theory & Applications*, 2010.

[TJ99] A. S. Tokar and P. A. Johnson. Rainfall-runoff modeling using artficial neural networks. *ASCE Journal of hydrological engineering*, 4 :232–239, 1999.

[TLHV09] R. Tóth, M. Lovera, P. S. C. Heuberger, and P. M. J. Van den Hof. Discretization of linear fractional representations of LPV systems. In *Proc. of the 48th IEEE Conf. on Decision and Control*, pages 7424–7429, Shanghai, China, Dec. 2009.

[Tót10] R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems*. Lecture Notes in Control and Information Sciences, Vol 403. Springer-Germany, 2010.

[TZGG09] S. Thil, W. X. Zheng, M. Gilson, and H. Garnier. Unifying some higher-order statistic-based methods for errors-in-variables model identification. *Automatica*, 45, Issue 8 :1937–1942, August 2009.

[VdH98] P. Van den Hof. Closed-loop issues in system identification. *Annual Reviews in Control*, 22 :173–186, 1998.

[VPS08] V. Vanbeylen, R. Pintelon, and J. Schoukens. Blind maximum likelihood identification of hammerstein systems. *Automatica*, 44, Issue 11 :3139–3146, 2008.

[vWV09] J. W. van Wingerden and M. Verhaegen. Subspace identification of bilinear and LPV systems for open- and closed-loop data. *Automatica*, 45 :372–381, 2009.

[Was02] M. G. Wassink. LPV control for wafer stages. Msc thesis, CTB595-02-2097, Philips CFT, May 2002.

[WD06a] X. Wei and L. Del Re. On persistent excitation for parameter estimation of quasi-LPV systems and its application in modeling of diesel engine torque. In *Proceedings of the 14th IFAC Symposium on System Identification*, pages 517–522, Newcastle, Australia, March 2006.

[WD06b] F. Wu and K. Dong. Gain-scheduling control of LFT systems using parameter-dependent Lyapunov functions. *Automatica*, 42 :39–50, 2006.

[WJC07] H. Wang, K. Ju, and K. H. Chon. Closed-loop nonlinear system identification via the vector optimal parameter search algorithm : Application to heart rate baroreflex control. *Medical Engineering & Physics*, 29, Issue 4 :505–515, May 2007.

[WNP+06] F. Wijnheijmer, G. Naus, W. Post, M. Steinbuch, and P. Teerhuis. Modeling and LPV control of an electro-hydraulic servo system. In *Proc. of the IEEE International Conf. on Control Applications*, pages 3116–3120, Munich, Germany, Oct. 2006.

[WZL09] J. Wang, Q. Zhang, and L. Ljung. Revisiting hammerstein system identification through the two-stage algorithm for bilinear parameter estimation. *Automatica*, 45, Issue 9 :2627–2633, September 2009.

[YB94] P. C. Young and K. J. Beven. Data-based mechanistic modelling and the rainfall-flow non-linearity. *Environmetrics*, 5 :335–363, 1994.

[YG06] P. C. Young and H. Garnier. Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling & Software*, 21, Issue 8 :1055–1072, August 2006.

[YGG08] P. C. Young, H. Garnier, and M. Gilson. *Identification of continuous-time models from sampled data*, chapter Refined instrumental variable identification of continous-time hybrid Box-Jenkins models. Springer-Verlag, London, 2008.

[YJ80] P. C. Young and A. Jakeman. Refined instrumental variable methods of recursive time-series analysis - part III. extensions. *International Journal of Control*, 31, Issue 4 :741–764, 1980.

[YL93] P. C. Young and M. J. Lees. The active mixing volume : a new concept in modelling environmental systems. In *Statistics for the Environment, Barnett V., Turkman K. F. (eds)*, pages 3–43, John Wiley, Chichester, 1993.

[You76] P. C. Young. Some observations on instrumental variable methods of time-series analysis. *International Journal of Control*, 23 :593–612, 1976.

[You86] P. C. Young. *Time series methods and recursive estimation in hydrological systems analysis.* In *River flow modelling and forecasting,* D. A. Kraijenhoff, J. R. Moll (eds), pages 129–180. D. Reidel : Dordrecht, 1986.

[You02] P. C. Young. Advances on real-time flood forecasting. *Phil. Trans. of the Royal Society*, 360, no. 1796 :1433–1450, 2002.

[You03] P. C. Young. Top-down ad data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological processes*, 17 :2195–2217, 2003.

[You08] P. C. Young. The refined instrumental variable method : Unified estimation of discrete and continuous-time transfer function models. *Journal Européen des Systèmes Automatisés*, 42 :149–179, 2008.

[ZD98] K. Zhou and J. C. Doyle. *Essentials of Robust Control*. Prentice-Hall, 1998.

[Zhe03] W. X. Zheng. Application of bels based methods in direct identification of linear systems from closed loop data. In *42nd IEEE Conference on Decision and Control*, pages 4539–4544, Hawaï - USA, December 2003.

[ZRBW06] D. Zhai, D. K. Rollins, N. Bhandari, and H. Wu. Continuous-time Hammerstein and Wiener modeling under second-order static nonlinearity for periodic process signals. *Computers and Chemical Engineering*, 31 :1–12, May 2006.

Résumé

La procédure d'identification consiste à rechercher un modèle mathématique adéquat pour un système dynamique donné à partir de données expérimentales. Alors que l'identification de système est orientée majoritairement pour répondre aux problèmes de commande depuis les années 90, l'identification de systèmes naturels reste cruciale pour une meilleure compréhension de notre environnement. Cette thèse vise à apporter une solution au problème de modélisation de la relation pluie/débit dans un bassin versant rural. À cet effet, deux structures de modèles non-linéaires sont étudiées : les modèles Hammerstein et les modèles *Linéaires à Paramètres variants* (LPV). La contribution principale de cette thèse réside dans le développement de méthodes dédiées à l'estimation de ces modèles, à temps discret ou continu, opérant en boucle ouverte ou fermée, en se concentrant sur le cas réaliste où le bruit de sortie est coloré et indépendant du processus étudié : le cas Box–Jenkins (BJ). De plus, les méthodes proposées ont été conçues spécialement pour fournir des résultats utiles dans le cas réel où le modèle de bruit est inconnu ou mal évalué.
Le premier chapitre constitue une introduction à la problématique de l'identification de systèmes naturels et motive les développements théoriques impliqués. Le deuxième chapitre présente une méthode sous optimale de variable instrumentale pour l'estimation des modèles Hammerstein BJ grâce à l'augmentation du modèle considéré. Le troisième chapitre se concentre sur l'identification de modèles LPV-BJ, soulève les problèmes rencontrés par les méthodes existantes, et propose une solution *via* une reformulation du modèle en un modèle multi-entrées-mono-sortie invariant dans le temps. Enfin, le dernier chapitre est dédié à l'application de ces méthodes sur des données acquises sur un bassin versant rural situé à Rouffach, Alsace, France et propose un processus d'identification innovant pour la modélisation de la relation pluie/débit.

Summary

System identification is an established field in the aera of system analysis and control. It aims at determining mathematical models for dynamical systems based measured data. Even if system identification has been mainly control oriented since the 90's, the identification of natural systems is crucial for a better understanding of our environment. This work aims at solving the modelling problem of the rainfall/flow relationship in rural catchments. In order to achieve this goal, two non-linear model structures are studied : the Hammerstein and the Linear Parameter Varying (LPV) models. The contribution of this work lies in the development of identification methods dealing robustly with estimation problem of such models, both in discrete-time and continuous-time, in open-loop and closed-loop configuration, focusing on the most realistic case where the noise added on the output is colored and independent on the studied process : the Box–Jenkins (BJ) case. Moreover, the methods are especially designed to result in relevant estimates in case the noise model is unknown or wrong which is the case in most (and in the presented) practical applications.
The first chapter is an introduction defining the problems encountered with natural systems and motivating the theoretical work induced. The second chapter presents a suboptimal *Refined Instrumental Variable* based method for Hammerstein BJ models using an augmented linear model. The third chapter focuses on the identification of LPV-BJ models, highlights the problems encountered by the existing methods and proposes a solution via a multi-input-single-output linear time invariant reformulation of the model. Finally, the last chapter is dedicated to the application of the presented methods on some real rainfall/flow data set acquired from a rural catchment situated in Rouffach, Alsace, France and proposes an innovating identification process for the identification of the rainfall/runoff relationship.

Mots-clés : Identification de systèmes, variable instrumentale, modèles non-linéaires, modèles Hammerstein, modèles LPV, systèmes environnementaux