GRAY-BOX NONLINEAR SYSTEM IDENTIFICATION USING POLYNOMIAL NARMAX MODELS

Allyne Machado dos Santos

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Química, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Química.

Orientadores: Argimiro Resende Secchi
Maurício Bezerra de Souza Júnior

Rio de Janeiro
Fevereiro de 2019

GRAY-BOX NONLINEAR SYSTEM IDENTIFICATION USING POLYNOMIAL NARMAX MODELS

Allyne Machado dos Santos

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA QUÍMICA.

Examinada por:

Prof. Argimiro Resende Secchi, D.Sc.

Prof. Maurício Bezerra de Souza Júnior, D.Sc.

Prof. Bruno Didier Oliver Capron, D.Sc.

Prof. Luiz Augusto da Cruz Meleiro, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
FEVEREIRO DE 2019

Santos, Allyne Machado dos

Gray-Box Nonlinear System Identification using polynomial NARMAX Models/Allyne Machado dos Santos. – Rio de Janeiro: UFRJ/COPPE, 2019.

XXI, 105 p.: il.; 29, 7cm.

Orientadores: Argimiro Resende Secchi

Maurício Bezerra de Souza Júnior

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Química, 2019.

Referências Bibliográficas: p. 80 – 84.

1. NARMAX model. 2. Nonlinear systems. 3. Gray-box. I. Secchi, Argimiro Resende *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Química. III. Título.

Dedico este trabalho aos meus pais, pelo apoio e pelo amor incessantes.

Acknowledgements

I am very grateful to my advisors Argimiro Resende Secchi and Maurício Bezerra de Souza Júnior, for guidance when I needed most and making my problems seem not that big on every meeting. Also, thanks to them I had a great opportunity to participate on the Brazilian-Norwegian Subsea Operations Consortium (BN-SOC), going on exchange to Trondheim, Norway.

I would like to express my appreciation to Professor Sigurd Skogestad, for welcoming me in the group, and Dinesh Krishnamoorthy, for assistance and all the shared knowledge.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

This study was also financed in part by the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

My acknowledgement to the International Partnerships for Excellence Education, Research and Innovation (INTPART) for funding this research during my stay in Trondheim.

To my friends from the Chemical Engineering Program (PEQ), Fabiana Coelho, Maycou Zamprognio, Mellyssa de Sousa and Otávio Ivo, for the deep and wise knowledge sharing talks in our endless lunch group. Without emotional support, a huge part of this would have been my worst nightmare, but you made it disappear on every despair I showed.

To my friends from the Norwegian University of Science and Technology (NTNU), for the conversations during lunch and external activities. My stay in Trondheim went smoothly and happy because of you.

To the Federal University of Rio de Janeiro.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

IDENTIFICAÇÃO CAIXA-CINZA DE SISTEMAS NÃO LINEARES USANDO MODELOS NARMAX POLINOMIAIS

Allyne Machado dos Santos

Fevereiro/2019

Orientadores: Argimiro Resende Secchi
Maurício Bezerra de Souza Júnior

Programa: Engenharia Química

O uso de um conjunto de modelos lineares para descrever um sistema não-linear tem muitas desvantagens. Para superar essas desvantagens, modelos não-lineares foram aprimorados. O modelo não-linear usado neste trabalho é o modelo de média móvel auto-regressiva não-linear com entradas exógenas, do inglês "Nonlinear AutoRegressive Moving Average models with eXogenous inputs" (NARMAX) do tipo polinomial. Esse tipo de modelo é linear nos parâmetros e considera, no modelo, o ruído, inerente a uma medição em uma planta industrial. Em geral, existem dois tipos de identificação: a identificação caixa-preta, que é um método típico de entrada e saída, ou seja, requer apenas dados para identificar o processo; e a identificação da caixa-cinza, que requer algumas informações sobre o sistema, além de dados. No presente trabalho, um tipo caixa-cinza é comparado com o tipo caixa-preta para fins de otimização e controle. A identificação é realizada usando o algoritmo de mínimos quadrados ortogonais e método de validação cruzada de k passos a frente. A otimização dinâmica em tempo real foi definida com base no modelo fenomenológico e em modelos estimados, e comparadas, para avaliar a melhoria na aplicação de modelos não lineares identificados. A identificação do tipo caixa-cinza se mostrou mais representativa em relação à não linearidade do sistema. A aplicação em otimização e controle gerou instabilidade do algoritmo. Isso pode ser devido ao fato de que o algoritmo de otimização usado na otimização dinâmica em tempo real tinha o mesmo valor para horizonte de controle e horizonte de predição. Apesar das oscilações de um estudo de caso, o algoritmo de identificação caixa-cinza mostrou sua capacidade de melhorar o modelo.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

GRAY-BOX NONLINEAR SYSTEM IDENTIFICATION USING POLYNOMIAL NARMAX MODELS

Allyne Machado dos Santos

February/2019

Advisors: Argimiro Resende Secchi
Maurício Bezerra de Souza Júnior

Department: Chemical Engineering

The usage of a collection of linear models to describe a nonlinear system has many disadvantages. In order to overcome these disadvantages, nonlinear models have been improved. The nonlinear model used in this work is the Nonlinear AutoRegressive Moving Average models with eXogenous inputs (NARMAX) of polynomial type. This type of model is linear on the parameters and accounts, in the model, for the existent noise, that is inherent of a measurement on a industrial plant. Broadly, there are two types of identification: the black-box identification, which is a typical input-output method, *i.e.*, only requires data in order to identify the process; and the gray-box identification, which requires some system information, besides data. In the present work, a gray-box identification is compared with the black-box one for optimization and control purposes. The identification is performed using the Orthogonal Least Square algorithm and validation is made using k-step-ahead cross-validation method. Dynamic real-time optimization was set based on both first principle models and estimated models, and compared, in order to evaluate improvement on the application of identified nonlinear models. The gray-box identification was more representative in relation to the nonlinearity of the system. The application in optimization and control generated instability of the algorithm. It can be due to the fact that the optimization algorithm used in dynamic real-time optimization had the same value for control horizon and prediction horizon. Despite the oscillations of one case study, the gray-box identification algorithm showed its capacity to improve the model.

Contents

List of Figures

List of Tables

List of Symbols

A_R	Surface area of the reactor, p. 20
A_a	Cross-sectional area of the annulus, p. 24
A_{bh}	Cross-sectional area of the well below the injection point, p. 23
A_r	Cross-sectional area of the riser, p. 24
A_r	Cross-sectional area of the riser manifold, p. 23
A_w	Cross-sectional area of the well above the injection point, p. 23
C_a	Concentration of component A in the reactor, p. 19
C_b	Concentration of component B in the reactor, p. 19
C_{a_0}	Concentration of component A at the reactor entrance, p. 19
C_{iv}	Valve flow coefficient for the downhole injection valve, p. 24
C_{pc}	Valve flow coefficient for the production choke, p. 24
C_{rh}	Valve flow coefficient for the riser head valve, p. 24
c_p	Specific heat capacity of the liquid, p. 20
d	Time delay, p. 10
E	Activation energy, p. 20
$\boldsymbol{e}$	Noise vector, p. 10
F	Flow rate through the reactor, p. 19
G	Polynomial function of the NARX part of the model, p. 13
GOR	Gas-oil ratio, p. 24

g	Acceleration of gravity constant, p. 24
g	Inequality constrains, p. 17
H	Polynomial function of the MA part of the model, p. 13
H_b	Vertical height of well tubing below the injection point, p. 24
H_w	Vertical height of well tubing above the injection point, p. 24
H_r	Vertical height of the riser, p. 24
h	Equality constrains, p. 17
J_{OLS}	Objective function of OLS algorithm, p. 13
$J_{n_{p_{MA}}}$	Objective function of MA part of the model with $n_{p_{MA}}$ features, p. 16
$J_{n_{p_{NARX}}}$	Objective function of NARX part of the model with $n_{p_{NARX}}$ features, p. 16
k_w	Heat transfer coefficient, p. 20
k_0	Pre-exponential constants, p. 20
L_a	Length of the annulus, p. 24
L_{bh}	Length of the well below the injection point, p. 23
L_r	Length of the riser, p. 24
L_r	Length of the riser manifold, p. 23
L_w	Length of the well above the injection point, p. 23
M	Process model, p. 18
M_w	Molecular weight of the gas, p. 23
m_{ga}	Mass of gas in the annulus, p. 22
m_{gr}	Mass of gas in the riser, p. 22
m_{gr}	Mass of oil in the riser, p. 22
m_{gt}	Mass of gas in the well tubing, p. 22
m_{ot}	Mass of oil in the well tubing, p. 22

N	Number of samples, p. 10
n	Order of the Pseudo-Random Binary Sequence, p. 6
n_θ	Number of candidates to be regressors, p. 11
n_e	Maximum lag of noise, p. 10
n_p	Number of features of the estimated model, p. 10
n_y	Maximum lag of system output, p. 10
$n_{p_{MA}}$	Number of features of MA part of the model, p. 14
$n_{p_{NARX}}$	Number of features of NARX part of the model, p. 14
n_{u_i}	Maximum lag of system inputs, p. 10
P	Polynomial function, p. 10
PI	Reservoir production index, p. 24
p_a	Pressure on the annulus, p. 24
p_m	Manifold pressure, p. 24
p_r	Reservoir pressure, p. 24
p_s	Separator pressure, p. 24
p_{bh}	Bottom hole pressure, p. 24
p_{rh}	Riser head pressure, p. 24
p_{wh}	Well-head pressure, p. 24
p_{wi}	Well injection point pressure, p. 24
R	Gas constant, p. 20
R^2	Determination coefficient, p. 14
T	Temperature in the reactor, p. 19
T_K	Cooling jacket temperature, p. 20
T_a	Temperature in the annulus, p. 23
T_w	Temperature in the well tubing, p. 24

T_0	Inlet temperature, p. 19
T_r	Average temperature in the riser, p. 24
$\boldsymbol{t}$	Time sample vector, p. 7
$\boldsymbol{u}$	Input variable vector, p. 10
V	Reactor volume, p. 20
V_a	Volume of the annulus, p. 24
$\overline{\text{var}}$	Variable after normalization, p. 9
var_{min}	Minimum value of a generic variable, p. 9
var	Measured data of a generic variable, p. 9
w_{gl}	Mass rate of gas lift injection, p. 22
w_{iv}	Flow through the downhole gas lift injection valve, p. 24
w_{iv}	Gas flow rate from the annulus into the tubing, p. 22
w_{pc}	Total flow through the production choke, p. 24
w_{pg}	Flow rate of produced gas, p. 22
w_{po}	Produced oil flow rate, p. 22
w_{rg}	Gas flow rate from the reservoir, p. 22
w_{rh}	Flow through the riser head choke, p. 24
w_{ro}	Oil flow rate from the reservoir, p. 22
w_{tg}	Total gas flow rate, p. 22
w_{tg}	Total oil flow rate, p. 22
$\bar{y}$	Output average, p. 14
$\boldsymbol{y}$	Output variable vector, p. 10
$\hat{y}(k)$	Predicted output variable at instant k, p. 10
Inf	Infinite, p. 43
NaN	Not-a-Number, p. 43

$(-\Delta H)$	Heat of reaction, p. 20
Δp_{fric}^{bh}	Frictional pressure drop in the well tubing below the injection point, p. 24
Δp_{fric}^{t}	Frictional pressure drop above the injection point, p. 24
Δp_{fric}^{t}	Frictional pressure drop in the riser, p. 24
Δvar	Operating range of a generic variable, p. 9
$\boldsymbol{\Psi}$	Matrix of regressors, p. 11
$\boldsymbol{\Psi}^*$	Matrix of chosen regressors, p. 11
$\boldsymbol{\psi_j}$	Vector of a regressor candidate, p. 11
$\boldsymbol{\theta}$	Estimated parameter vector, p. 10
$\tilde{\boldsymbol{\Psi}}^{(\boldsymbol{k})}$	k Householder transformations made into $\tilde{\boldsymbol{\Psi}}$, p. 12
$\tilde{\boldsymbol{\Psi}}$	Extended matrix, p. 12
ℓ	Nonlinearity degree, p. 10
ω	Input signal frequency of multisine signal, p. 7
ρ	Liquid density, p. 20
ρ_a	Density of gas in the annulus, p. 23
ρ_o	Density of oil in the reservoir, p. 23
ρ_w	Fluid mixture density in the tubing, p. 23
ε	Prediction error, p. 10

List of Abbreviations

AIC	Akaike's Information Criterion, p. 4
ANN	Artificial Neural Networks, p. 28
ANN-MS	Artificial Neural Network identified from Multisine disturbance, p. 29
ANN-PRBS	Artificial Neural Network identified from PRBS disturbance, p. 29
ANN-RRSS	Artificial Neural Network identified from RRSS disturbance, p. 29
BIC	Bayes Information Criterion, p. 4
BSD	Bootstrap to Structure Detection, p. 4
CSTR	Continuous Stirred-Tank Reactor, p. 18
DRTO	Dynamic Real-Time Optimization, p. 2
DRTO-ID	DRTO based on first principle model, p. 42
DRTO-NARMAX	DRTO based on NARMAX model, p. 42
ELS	Extended Least Square algorithm, p. 3
ERR	Error Reduction Rate, p. 4
FOLS	Forward Orthogonal Least Square, p. 4
FO	Fast Orthogonal algorithm, p. 3
FPE	Final Prediction Error, p. 4
GH	Golub-Householder algorithm, p. 10
MA	Moving Average, p. xvi

MIMO	Multiple Input Multiple Output, p. 9
MISO	Multiple Input Single Output, p. 9
MLP	Multi-Layer Perceptron, p. 28
MPC	Model-based Predictive Controller, p. 16
NARMAX	Nonlinear AutoRegressive Moving Average models with eXogenous inputs, p. 2
NARX	Nonlinear AutoRegressive model with eXogenous inputs, p. 13
OH	Optimization Horizon, p. 16
OLS	Orthogonal Least Square algorithm, p. 3
PH	Prediction Horizon, p. 8
PRBS	Pseudo-Random Binary Sequence, p. 6
PSO	Particle Swarm Optimization, p. 4
RRSS	Random Range Step Sequence, p. 6
SQE	Sum of Quadratic Errors, p. 28

Chapter 1

Introduction

1.1 Motivation and Objectives

In most industrial chemical processes, the nonlinear behavior of the plant is evident and relevant. However, physical models are usually complex or unavailable, demanding a high effort to be developed or to be solved. These characteristics make the usage of these models in process control and real-time optimization disadvantageous, leading to a tendency of substituting them by an identified and simpler one. Thus, system identification is being highlighted, when using input-output type.

High usage frequency of linear mathematical models to represent nonlinear systems has been reported in the 80's. However, problems emerged when such systems were highly nonlinear, generating relevant loss or even instability of the plant. Such degeneracy is due to the fact that linear models used in controllers can not handle dynamic behaviors, such as gain signal inversion or large gain variation, *etc.* Problems with lack of representativeness of nonlinear behavior demand the development of simpler nonlinear model structures (in order to be preferred over local linear models), when compared with physical models.

There are several types of nonlinear identification models that can be divided into three groups:

- Models in frequency domain;
- Non-parametric models;
- Parametric models.

Models in frequency domain use data that are subjected to Fourier transformation or data that are naturally in frequency domain. Non-parametric models are not described explicitly by a finite number of parameters and its structure is not pre-selected. It can represent time or frequency-domain data (LJUNG, 1999). Parametric models use parametric mathematical structures aiming to represent the dynamic behavior of a nonlinear system in time domain. The parametric mathematical structure considered in this study is the polynomial NARMAX model (Nonlinear AutoRegressive Moving Average models with eXogenous inputs). This model uses measured discrete data of inputs and outputs of a system, from industrial or computational source (by means of simulations, as this study).

NARMAX models can be used not only for black-box system identification (system information is unknown), but also for gray-box type (system information is used in the determination of the model structure and/or parameter estimation). The application of black-box identification is considered all-embracing as it does not require prior information of the process. Whereas, for the gray-box approach, the number of terms, computational cost and possibility of error in capturing systems dynamics are reduced, due to user interaction that can be used to improve the NARMAX models.

This work has the main objective of developing a methodology for identifying nonlinear systems with prior knowledge using NARMAX models (gray-box identification), focusing on the determination of the model structure. The specific objectives are to analyze the sensitivity of the output when disturbing the inputs, perform the identification of a benchmark nonlinear process (non-isothermal Van de Vusse reactor) and an oil production system with two gas-lift wells, compare with the results of a black-box approach that uses only polynomial NARMAX structure, validate the models to its purpose, use it on Dynamic Real-Time Optimization (DRTO) of the studied cases and compare it with an ideal DRTO (that is constituted by first principle models and discretized using direct collocation method).

1.2 Dissertation Structure

A bibliographic review of the concepts and tools most used on this matter is made in Chapter 2, emphasizing necessary points for better understanding.

A methodology is proposed in Chapter 3 to identify the models of two case studies using a gray-box procedure with NARMAX model.

Chapter 4 presents the results and its discussion divided in two sections exposing and discussing the application results of black-box and gray-box methods.

Final conclusions and suggestions for future work are presented in Chapter 5.

Chapter 2

Literature Review

The idea of NARMAX model emerged from a model that is based on Volterra series, whose main issue is the huge quantity of terms (more than 100) (LEONTARITIS & BILLINGS, 1985). Its concept was developed by BILLINGS & LEONTARITIS (1981, 1982), when they studied input-output models that could represent a major class of nonlinear problems. The authors also developed two methods based on the least square algorithm in order to overcome one of the inherent difficulties of parameter estimation of NARMAX models (known as polarized results).

BILLINGS & VOON (1983) showed the inefficiency of traditional covariance tests on nonlinear problems and proposed methods to detect all terms on residuals. In the following year, the same authors continued their investigation on modified parameter estimation techniques for nonlinear systems and also discussed about the importance of choosing data with representative nonlinearity, the system sensitivity in relation to its inputs, and the selection of input variables (BILLINGS & VOON, 1984). With that, gray-box identification of nonlinear systems using NARMAX structure emerged. It consists in using available information before obtaining the final model (not to be mistaken with *a priori* knowledge, that is knowledge coming from physical modeling) (CORRÊA & AGUIRRE, 2004).

Specific classes of NARMAX models were given special names regarding its structure, such as polynomial NARMAX and rational NARMAX (BILLINGS & CHEN, 1989). The difference is evident in the parameter estimation method; polynomial NARMAX is linear on the parameters, while rational NARMAX is nonlinear on the parameters.

The parameter estimation step of identification using polynomial NARMAX received a lot of contributions, as the Extended Least Square algorithm (ELS) (BILLINGS & VOON, 1984), Prediction error with stepwise regression algorithm (BILLINGS & VOON, 1986), Fast Orthogonal algorithm (FO) (KORENBERG *et al.*, 1988), Orthogonal Least Square algorithm (OLS) (CHEN *et al.*, 1989), OLS with forward subset selection (FOLS - Forward Orthogonal Least Square)

(BILLINGS & CHEN, 1998), Bootstrap to Structure Detection algorithm (BSD) (KUKREJA *et al.*, 2004), Genetic algorithm (MARIUS & NICOLAE, 2015), Particle Swarm Optimization (PSO) (ABDULLAH *et al.*, 2015). The method used in this work is based on the OLS algorithm due to its simplicity of calculation and its capacity to mitigate ill-conditioning problems, which are very common in nonlinear identification.

In order to define the optimal number of terms of the model, there are some information criteria: Akaike's Information Criterion (AIC) (AKAIKE, 1974), Final Prediction Error (FPE) , that is equivalent to AIC in some way (LEONTARITIS & BILLINGS, 1987b), Bayes Information Criterion (BIC). The most used information criteria is AIC (AGUIRRE *et al.*, 1998), although it looses effectiveness when it comes to nonlinear identification (some regressors with low AIC but actually with high importance to the model are wrongly disregarded) (AGUIRRE, 2000). In this work, the optimal number of terms was found by doing a wide search and comparing the objective function value (trade off between accuracy and model simplicity).

The detection of terms of the model can be executed by several methods, some of them are described in AGUIRRE *et al.* (1998). Besides those cited methods, some authors use statistic models to choose its structure and validate the model through the usage of the confidence interval method. Although most articles showed its efficiency, it also can be an exhaustive method, as the number of possible terms can increase a lot depending on the nonlinear constants that are inherent to the model (BILLINGS & FADZIL, 1985). Besides, adding or removing terms do not affect estimated parameters in some of estimation methods (as for the OLS algorithm). Therefore, criteria of structure detection have been developed, such as Error Reduction Rate (ERR) (CHEN & BILLINGS, 1989), algorithms of forward regression (BILLINGS *et al.*, 1988), backward regression (DRAPER & SMITH, 1998), stepwise regression (BILLINGS & VOON, 1986). It is necessary to highlight that AGUIRRE *et al.* (1998) used a detection method called term cluster, which has been defined in AGUIRRE & BILLINGS (1995). It reduces the quantity of candidate terms to the final model by eliminating clusters with much smaller coefficient than effective clusters' coefficients. As the parameter estimation algorithm used in this work was based on the OLS algorithm, the chosen criteria of structure detection was the ERR as in THOMSON *et al.* (1996), where the authors presented an out of the ordinary methodology, identifying the model of a parallel-tube heat exchanger. This methodology uses an algorithm that is based on orthogonal estimator from BILLINGS *et al.* (1988) and validates the model by one-step-ahead prediction, 95% confidence interval of all normalized function correlations and step response (also called dynamic simulation method). Unlike most of other authors, they do not use AIC, because the algorithm has already an objective function, that is the quadratic error, to optimize the number of terms.

CORRÊA & AGUIRRE (2004) made an extended review about system identification with gray-box nonlinear identification. They described how to use auxiliary information on structure detection of identification using polynomial NARMAX models (based on static gain, number of stationary states on the output variables, qualitative characteristics with respect to dynamic behavior of the system) and on parameter estimation, both using term cluster method.

The gray-box identification was highlighted in JOHANSEN (1996). The author used different types of knowledge of the system, like a basic model that represents the system within operating conditions; noise with linear model; mass balance on stationary state; stability of the system. The author applied each of these types to a pH neutralization tank and compared them in order to observe how the type of information affects the number of parameters of identified polynomial NARMAX model.

In JÁCOME (1996), the author identified the model using a gray-box type identification with OLS algorithm, Householder transformations and term cluster. The auxiliary information helped in selecting polynomial structure.

When the type of auxiliary information arises from the static behavior of the system, gray-box nonlinear identification methods use a multi-objective algorithm, because it searches a mid term between dynamic and static modes of a system so it can be represented by the identified model in any of these situations (BARBOSA *et al.*, 2011). This type of identification has advantages over the black-box type only when the data set of the system do not represent all desired system information (TEIXEIRA & AGUIRRE, 2011).

A gray-box problem can be classified into several shades of gray (KARPLUS, 1977). The model is labeled as light gray when its structure is defined and has physical meaning. Nevertheless, its parameters still need to be estimated. The dark gray is when some auxiliary information, such as based on static gain, number of stationary states on the output variables, mass balance, energy balance, are used to select the structure of the polynomial model. Finally, there is a middle gray, that is not much investigated. It uses auxiliary information, such as qualitative characteristics with respect to dynamic behavior of the system, to form non-polynomial-based structures, which is the case of this work.

Chapter 3

Proposed Methodology

System identification is executed in five main steps (AGUIRRE, 2000):

- Dynamic tests and data acquisition;
- Choice of mathematical structure to be used;
- Model structure determination;
- Parameter estimation;
- Model validation.

This work focused on the model structure determination, presented in Section 3.3.

3.1 Type of Disturbances

Many studies on linear identification have used Pseudo-Random Binary Sequence (PRBS) to generate input signals. However, when this type of disturbance is used for nonlinear identification, inaccurate models are generated (LEONTARITIS & BILLINGS, 1987a).

A prior study of types of disturbances was necessary because of divergence in literature on what would be the best one to perform a nonlinear identification. Three types of signals were tested in the identification of the non-isothermal Van de Vusse reactor using neural networks: PRBS, multisine and random range step sequence (RRSS). The first one is mostly used in linear identification, the second one is used in linear and nonlinear cases and the third one is proposed in this study. They were applied to both input variables of the Van de Vusse reactor, which is described in Section 3.8.1.

PRBS, as the name suggests and Figure 3.1 shows, has only two values. It is generated by choosing range and order (n, which is an integer number). The order n defines the maximum period, which is given by $2^n - 1$.

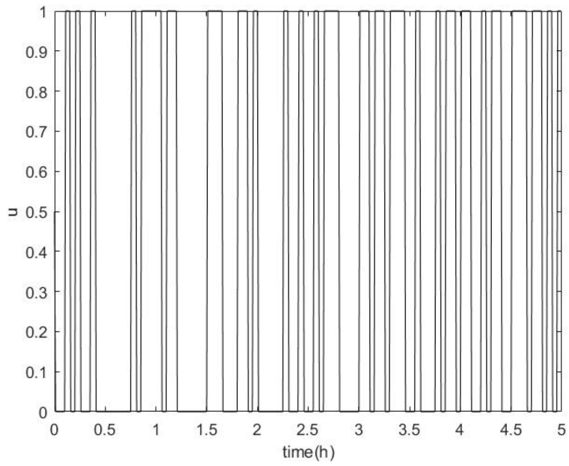


Figure 3.1: Pseudo-Random Binary Sequence.

Multisine signal (SCHMITZ & GREEN, 2012), in Figure 3.2, was generated by a linear combination of sines with random argument between 0 and $2\pi\omega t$, where ω is the input signal frequency, $\boldsymbol{t}$ is the vector of the time samples.

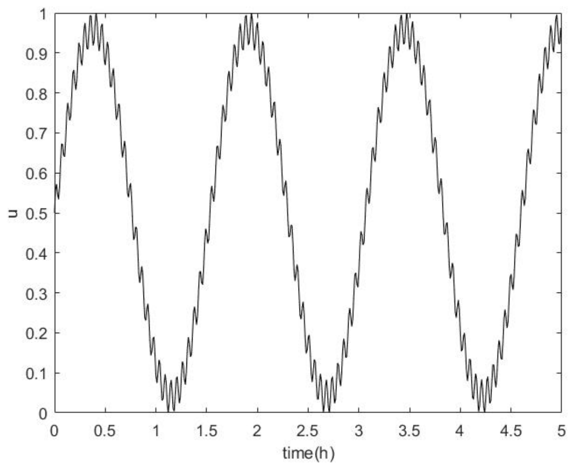


Figure 3.2: Multisine signal.

RRSS, as shown in Figure 3.3, is generated by setting random numbers between 0 and 1 to the magnitude of each step and specifying how long the step value is kept constant.

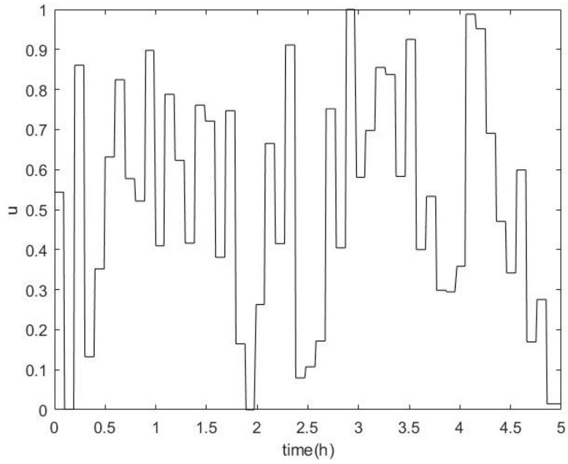


Figure 3.3: Random Range Step Sequence.

3.2 Data Acquisition

Data acquisition is an important task as it can determine the accuracy of a model. The data must have the same nonlinearity degree of the process, enough time on each excitation to a stable dynamic response and a large enough range to represent a larger operating region.

In order to know all this information about the process, a study of the process was made. A pulse was applied on each input, once at a time, and its response was recorded. The time constant can be taken and the prediction horizon (PH) can be calculated, as 5 times the time constant.

The real system was emulated by using first principle model. Simulations were performed in the software MATLAB version 2016b. To simulate the data acquisition procedure, each input variable was disturbed. Due to measurement uncertainty, white noise was added to the measured variables. In the case of flow rate, uncertainty was generally around 2%, while for temperature, uncertainty was between 0.5 and 1 ℃.

An additional procedure is the normalization of data. This is important to mitigate ill-conditioning problems, which are characteristic of NARMAX model identification. Normalization was performed in terms of Equation 3.1:

$$\overline{\text{var}} = \frac{\text{var} - \text{var}_{min}}{\Delta\text{var}} \tag{3.1}$$

where var represents any of the acquired data (measured data) to be normalized, $\overline{\text{var}}$ is the variable vector after normalization, var_{min} is the minimum value of the original values and Δvar is the operating range of var.

3.3 Model Structure

NARMAX models can be multiple input multiple output (MIMO), when an estimated output variable depends on other output variables. However, the NARMAX model that is used in this work has multiple input and single output (MISO), with the predicted output variable depending on its past values, past values of input variables and noise, generally represented by Equation 3.2.

$$y(k) = P^{\ell}[y(k-1), ..., y(k-n_y), u_i(k-d), ..., u_i(k-d-n_{u_i}), e(k-1), ..., e(k-n_e)] + \varepsilon(k) \tag{3.2}$$

where P is a polynomial function with nonlinearity degree ℓ in relation to all variables (inputs - u_i, with i referring to the number of input variables, output - y, noise - e), with $k = 1, ..., N$, N is the number of samples, n_y, n_{u_i}, n_e are maximum lags of system output, inputs and noise, respectively, d is the time delay of the model, ε is the prediction error, or residual, that is defined in Equation 3.3.

$$\varepsilon(k) = y(k) - \hat{y}(k) \tag{3.3}$$

where $y(k)$ is the output variable at instant k and $\hat{y}(k)$ is the predicted value of the same variable at the same instant.

For polynomial NARMAX models, Equation 3.2 is expanded into Equation 3.4:

$$\begin{aligned} y(k) = \sum_{m_1=1}^{n} \theta_{m_1} x_{m_1}(k) + \sum_{m_1=1}^{n} \sum_{m_2=m_1}^{n} \theta_{m_1 m_2} x_{m_1}(k) x_{m_2}(k) \, + \\ \sum_{m_1=1}^{n} \cdots \sum_{m_\ell=m_{\ell-1}}^{n} \theta_{m_1 \ldots m_\ell} x_{m_1}(k) \ldots x_{m_\ell}(k) \; + \; \varepsilon(k) \end{aligned} \tag{3.4}$$

Equation 3.4 can be rewritten in a matrix form, as in Equation 3.5:

$$\boldsymbol{y} = \boldsymbol{\Psi}^{*} \boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{3.5}$$

where θ is the estimated parameter vector, with n_p terms.

$$\boldsymbol{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \quad \boldsymbol{\Psi}^{*} = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_{n_p} \end{bmatrix} \quad \boldsymbol{\psi_j} = \begin{bmatrix} \psi_{1j} \\ \psi_{2j} \\ \vdots \\ \psi_{Nj} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \vdots \\ \varepsilon(N) \end{bmatrix}$$

Some definitions of specific nomenclature are described below.

Definition 1: Candidates to be regressors are all possible combinations between variables (inputs, output and noise lags) and it generates the matrix of regressors, $\mathbf{\Psi}$, of dimension $N \times n_\theta$. In $\mathbf{\Psi}^*$, the asterisk refers to the matrix of chosen regressors that compose the estimated model (with n_p features).

Definition 2: Features are the candidates to be regressors, $\boldsymbol{\psi_j}$, multiplied by the estimated parameter, θ_j, with $j = 1, 2, ..., n_p$.

The polynomial regressors are linear and nonlinear combinations among all the variables in a way that the maximum nonlinearity degree is ℓ. Meanwhile, change in coordinates does not have this limitation of degree nor has to be polynomial at all, as it can be exponential, logarithmic, sinusoidal, *etc.* It does not affect the linearity on the parameters, so it does not affect the solving algorithm. Also, it makes the user experience with the system very important in the application of this type of methodology.

3.4 Parameter Estimation

There are three types of parameter estimation. One is known as batch estimation. It uses all the data at once to identify the system. It is an off-line technique and the estimated parameters are time-invariant. A second type is a recursive estimation, which is on-line and the parameters are time-variant (ZHU & BILLINGS, 1991). The third one is a mixture between the batch and recursive approaches. It is called moving horizon estimation (MHE), which uses batch estimation in a moving window of data (JØRGENSEN, 2004). in the present work, the batch estimation is used.

Some classic parameter estimation algorithms face ill-conditioning problems, if the system presents high nonlinearity. There are modifications for such cases described in AGUIRRE (2000). One of the methods is the Golub-Householder algorithm (GH), which is an orthogonal least square algorithm with Householder transformations and error reduction rate (ERR), aiding the regressors selection. This one is the implemented method in this part of identification; a brief description follows.

3.4.1 Golub-Householder Algorithm with ERR

The number of candidates to be regressors, n_θ, is given by the equations below:

$$n_\theta = M + 1 \tag{3.6}$$

$$M = \sum_{i=1}^{\ell} n_i \tag{3.7}$$

$$n_i = \frac{n_{i-1}(n_y + n_{u_1} + n_{u_2} + ... + n_{u_j} + n_e + i - 1)}{i}, n_0 = 1 \tag{3.8}$$

An extended matrix is set (matrix $\tilde{\boldsymbol{\Psi}}$, of dimension $N \times (n_\theta + 1)$) with all the candidates (matrix $\boldsymbol{\Psi}$, of dimension $N \times n_\theta$) and the vector of output data ($\boldsymbol{y}$), represented in Equation 3.9:

$$\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}^{(0)} = [\boldsymbol{\Psi}\ \boldsymbol{y}] \tag{3.9}$$

where the index (0) makes reference to the number of Householder transformations made into the extended matrix.

After $(k-1)$ Householder transformations, the extended matrix is shown in Equation 3.10.

$$\tilde{\boldsymbol{\Psi}}^{(k-1)} = \begin{bmatrix} \boldsymbol{V}_{k-1} & \tilde{\boldsymbol{\psi}}_{\boldsymbol{j}}^{(k-1)} \cdots & \tilde{\boldsymbol{\psi}}_{\boldsymbol{n_\theta}}^{(k-1)} & \boldsymbol{y}^{*,(k-1)} \\ \boldsymbol{0} & & & \end{bmatrix} \tag{3.10}$$

where $\boldsymbol{V}_{k-1}$ is an upper triangular matrix of dimension $(k-1)\times(k-1)$, the superscript $*$ in $\boldsymbol{y}^*$ refers to the vector $\boldsymbol{y}$ with some lines changed by the $(k-1)$ householder transformations.

The ERR is calculated using the equations below:

$$a_j^{(k)} = \sum_{i=k}^{N} (\tilde{\psi}_{ij}^{(k-1)})^2, j = k, ..., n_\theta \tag{3.11}$$

$$b_j^{(k)} = \sum_{i=k}^{N} \tilde{\psi}_{ij}^{(k-1)} y_i^{(k-1)}, j = k, ..., n_\theta \tag{3.12}$$

$$ERR_j^{(k)} = \frac{\left(b_j^{(k)}\right)^2}{a_j^{(k)} \langle y, y \rangle} \tag{3.13}$$

where $\langle ., . \rangle$ indicates the inner product, k refers to transformation number k and j refers to term number j.

The GH with ERR algorithm has four main steps, which are described below.

`Algorithm:`

1. `Calculate ERR of the other candidates (for all of them, at the beginning);`

2. `Determine the next candidate with highest ERR and add it to the model, recording its position on the extended matrix;`

3. `Make the Householder transformation` n_θ `times;`

4. `Repeat the cycle until` n_p `features have been chosen to the model (number of terms of the model, determined by the user or by an external determination algorithm).`

The result of making the Householder transformations n_θ times is an orthogonal matrix Q, as shown in Equation 3.15. The Householder transformation is detailed in the Appendix A.

$$\tilde{\mathbf{\Psi}}^{(n_\theta)} = \begin{bmatrix} \boldsymbol{V}_{n_\theta} & \boldsymbol{y_1^*} \\ \mathbf{0} & \boldsymbol{y_2^*} \end{bmatrix} \tag{3.14}$$

$$\boldsymbol{Q}\mathbf{\Psi} = \begin{bmatrix} \boldsymbol{V}_{n_\theta} \\ \mathbf{0} \end{bmatrix} \tag{3.15}$$

where $\boldsymbol{V}_{n_\theta}$ is an upper triangular matrix of dimension $n_\theta \times n_\theta$ and the null matrix has dimension of $(N - n_\theta) \times n_\theta$.

The estimated parameters are given by Equation 3.16 with objective function given by Equation 3.17.

$$\boldsymbol{\theta}_{OLS} = \boldsymbol{V}_{n_\theta}^{-1} \boldsymbol{y_1^*} \tag{3.16}$$

$$J_{OLS} = \boldsymbol{y_2^{*,T}} \boldsymbol{y_2^*} \tag{3.17}$$

3.4.2 NARX and MA Parameter Estimation

The GH with ERR algorithm is used in two parts of the parameter estimation. The decomposition of the original problem, Equation 3.2, is a way of reducing the complexity of NARMAX estimation, by transforming that equation into Equation 3.18. In BILLINGS (2013), this decomposition is made with a different estimation algorithm (FOLS) and uses a different stopping criterion (ERR), but here, it is used with OLS, that chooses the iteration by comparing the objective function value with the previous one, when a tolerance is reached.

$$y(k) = F^{\ell}[y(k-1), ..., y(k-n_y), u_i(k-d), ..., u_i(k-d-n_{u_i})] + \\ G^{\ell}[y(k-1), ..., y(k-n_y), u_i(k-d), ..., u_i(k-d-n_{u_i}), e(k-1), ..., e(k-n_e)] + \varepsilon(k) \tag{3.18}$$

where G is a polynomial function that contains only the combinations of input and output variables, defined here as NARX part of the model; and H is a polynomial function that contains all combinations of input and output variables with noise, defined here as the MA part.

First, it estimates the parameters of the NARX part, using a wide search of the model order parameters (n_{u_i} and n_y), which have maximum value given by the user. Within each combination of these values, it chooses the optimal number of features $n_{p_{NARX}}$ considering the objective function value (J_{OLS}), and varying from 1 to 15, which is an average value for nonlinear chemical process identification. After that, the combination with minimum objective function value is chosen. This procedure is done with constant nonlinearity degree (ℓ), which varies from 1 to 3, which is also an average value for nonlinear chemical process identification. After choosing the model order parameters for each ℓ, the choice of the nonlinearity degree is carried out with the cross-validation method using the R-squared as comparison criterion.

Second, the noise model is identified using the residuals. The parameters are estimated changing n_e. If the objective function is reduced, the MA part with $n_{p_{MA}}$ features is added to the model.

3.5 Validation of the Model

There are a variety of model validations, such as correlation-based validity tests, cross-validation, step-response testing THOMSON *et al.* (1996). In this present work, cross-validation was performed with different data set than the one used for parameter estimation, and evaluated by calculating the determination coefficient (R-squared) given by Equation 3.19. It needs to be pointed out that the R-squared may happen to not be between 0 and 1, in the case of a nonlinear identification. This happens when the identification is very poor and do not represent the system at all. The cross-validation was made by comparing the new data set with the predicted output generated in each prediction horizon.

$$R^2 = 1 - \frac{\sum_{k=1}^{N} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N} (y(k) - \bar{y}(k))^2} \tag{3.19}$$

where $\bar{y}$ is the output average and N is the number of samples.

The validation can be also divided into three types regarding to the length of the prediction trajectory.

Definition 3: A one-step-ahead validation consists in the prediction values of one step ahead calculated using past original data, *i.e.*, each prediction trajectory has length of one point.

Definition 4: A k-step-ahead validation consists in a prediction trajectory calculated using a few points of the past original data in a way that predicted values are calculated with predicted data, but after a number of points (prediction horizon) it uses past original data again to restart the prediction trajectory, so each prediction trajectory has length of k points.

Definition 5: An infinite-horizon validation consists in a prediction trajectory calculated using only predicted values, except the starting points, that use past original data, *i. e.*, it is a recursive calculation with only the starting point depending on the original data. The prediction trajectory, in this case, has infinite length.

The chosen validation method was the cross-validation with k steps ahead, because the proposed identification methodology is directed to optimization and control purposes, which needs the identified model to represent the process within a certain prediction horizon.

3.6 Identification Summary

The identification algorithm used in this work can be summarized by Figure 3.4.

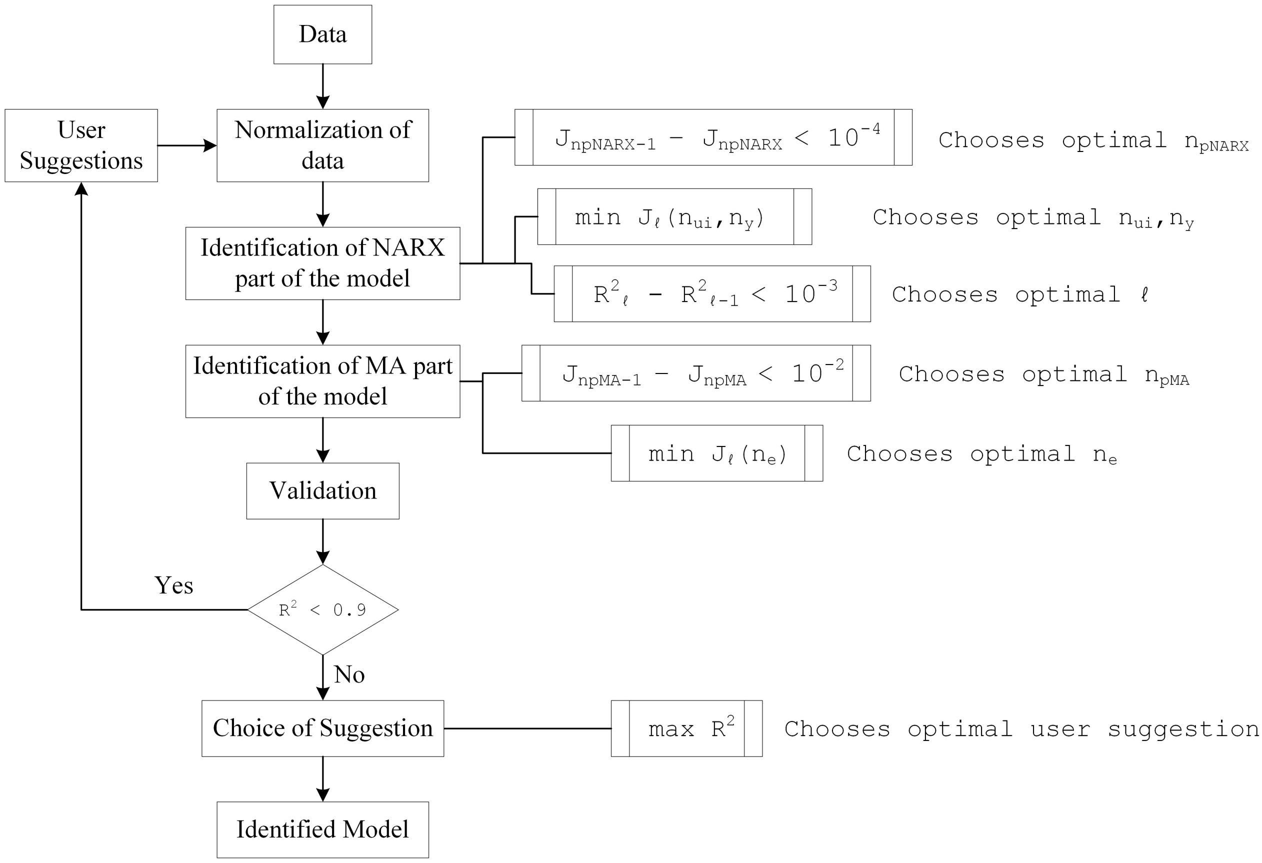


Figure 3.4: Block diagram of batch estimation.

where $J_{n_{p_{NARX}}}$ is the OLS objective function of $n_{p_{NARX}}$-term model (NARX part) and $J_{n_{p_{MA}}}$ is the OLS objective function of $n_{p_{MA}}$-term model (MA part).

The identification algorithm can be divided into seven steps, which are described below.

Algorithm:

1. The user starts by suggesting some change on the coordinates or not.
2. The user specifies a range of values for the model orders $(n_y, n_{ui}, n_e$ and $d)$.
3. Synthetic data is acquired from first principle models.
4. Normalization of data is done.
5. Off-line identification algorithm chooses regressors, estimates parameters and optimizes the number of features for each type of model (varying the nonlinearity degree, ℓ) and for each suggestion on coordinate change.

6. Validation step compares the R-squared value of all types of models and chooses the model that has the highest one.

7. The suggestion on coordinate change is also chosen comparing the R-squared values and the one of the black-box identification.

The user interaction is the usage of user knowledge about the process. Some suggestions were made testing simple nonlinear combinations, such as $\sqrt{\boldsymbol{u_1}}$, $\boldsymbol{u_1}/\boldsymbol{u_2}$, $\boldsymbol{u_1^2}/\boldsymbol{u_2}$, *etc.*, and others based on energy or mass balance.

3.7 Dynamic Real-Time Optimization

Dynamic real-time optimization (DRTO), as the name suggests, is a real-time optimization, but using dynamic models to compute trajectories for the decision variables and using an economic objective function (JAMALUDIN & SWARTZ, 2016).

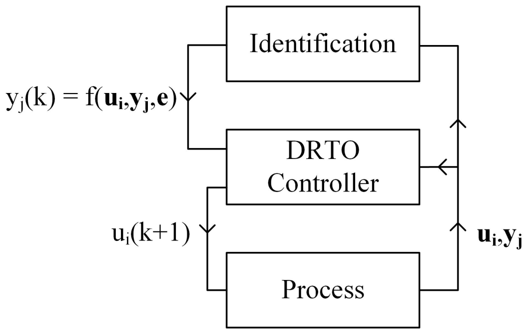


Figure 3.5: Optimizing control hierarchy.

The control strategy used in this work is a one-layer architecture (as shown in Figure 3.5), that sends the control actions directly to the plant, although it can have a different architecture, as in WÜRTH *et al.* (2011), which used a two-layer architecture composed by a lower layer with a model-based predictive controller (MPC) based on valid linear models when near the operating point; and a upper layer with a DRTO based on rigorous nonlinear models.

The closed loop consists in starting with as many normalized samples as the NARMAX model requires. DRTO controller unnormalizes prediction output and calculates the control actions aiming to minimize the economic objective function subject to constrains g and h, as in Equations 3.20-3.22. The optimization horizon (OH) is set by the user. The objective function is the integral of the profit function, f_1, summed to the cost function, f_2.

$$J_{DRTO} = \int_0^{OH} (-f_1(\boldsymbol{y_j}) + f_2(\boldsymbol{u_i}))dt \tag{3.20}$$

$$\min_{\boldsymbol{u_i}} \ J_{DRTO}(\boldsymbol{u_i}, \boldsymbol{y_j}) \tag{3.21}$$

subject to

$$\begin{aligned} y_j(k) = M(\boldsymbol{u_i}, \boldsymbol{y_j}) \text{ or } \dot{y}_j = M(\boldsymbol{u_i}, \boldsymbol{y_j}) \\ g(\boldsymbol{u_i}) \geqslant 0 \\ h(\boldsymbol{u_i}) = 0 \end{aligned} \tag{3.22}$$

where M is the model of the process. It can be the NARMAX model or the first principle model.

After minimizing J_{DRTO} for OH steps ahead, the first control action is sent to the system, which responds to it. The response and the input variables are normalized and sent back to the DRTO controller and the loop continues.

The most common way to solve a nonlinear optimal control problem (e.g. MPC, DRTO) is by discretizing the infinite dimensional control problem into a nonlinear programming problem (NLP). This can be performed by using single shooting, multiple shooting or direct collocation methods. Another way is to use directly discrete models, such as an identified NARMAX model.

A comparison was made between a DRTO composed with first principle model, discretized with direct collocation method, and a DRTO composed with the identified NARMAX models.

The closed loop DRTO routine was developed by Dinesh Krishnamoorthy in CasADi v3.4.5, which is a Matlab Front-end developed at the Optimization in Engineering Center, in K.U.Leuven, Belgium (ANDERSSON, 2013). The one based on first principle model uses the third order direct collocation method to set the problem up and IPOPT was used to solve it, as in KRISHNAMOORTHY *et al.* (2018).

3.8 Case Studies

Two processes were chosen to be cases of study. The Van de Vusse reactor is a well-known nonlinear process and it was used to compile the identification code from the scratch. An oil production system with two gas-lift wells is a large and complex nonlinear system that is recently being studied with implementation of dynamic real-time optimization (KRISHNAMOORTHY *et al.*, 2018), so the need of a good nonlinear discrete model appears. The Van de Vusse reactor is described in Section 3.8.1, and oil production system with two gas-lift wells, in Section 3.8.2.

3.8.1 Van de Vusse Reactor

This case study is the identification of a non-isothermal Van de Vusse continuous stirred-tank reactor (CSTR) model (TRIERWEILER, 1997), presented in Figure 3.6. The involved kinetics are shown in Equation 3.23:

$$\begin{gathered} \mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C} \\ 2\mathrm{A} \xrightarrow{k_3} \mathrm{D} \end{gathered} \tag{3.23}$$

where A, B, C and D represent the components cyclopentadiene, cyclopentenol, cyclopentanediol and dicyclopentadiene, respectively.

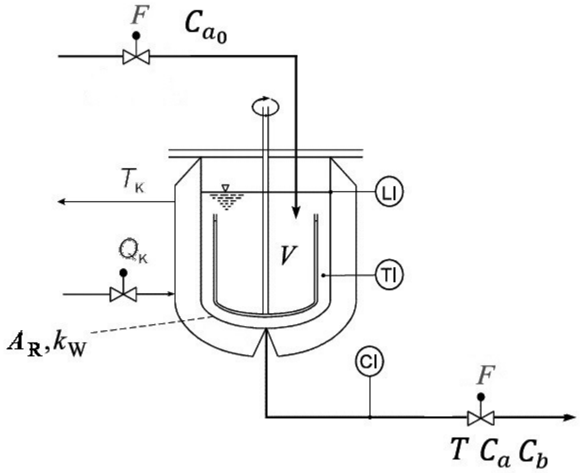


Figure 3.6: Van de Vusse CSTR. Adapted from TRIERWEILER (1997).

The component mass balance and the energy balance are given by the following ordinary differential equations:

$$\frac{dC_a}{dt} = \frac{F}{V}(C_{a_0} - C_a) - k_1 C_a - k_3 C_a^2 \tag{3.24}$$

$$\frac{dC_b}{dt} = \frac{F}{V} C_b + k_1 C_a - k_2 C_b \tag{3.25}$$

$$\begin{aligned} \frac{dT}{dt} = \frac{1}{\rho c_p}[k_1 C_a(-\Delta H_{AB}) + k_2 C_b(-\Delta H_{BC}) + k_3 C_a^2(-\Delta H_{AD})] + \\ \frac{F}{V}(T_0 - T) + \frac{k_w A_R}{\rho c_p V}(T_K - T) \end{aligned} \tag{3.26}$$

where C_{a_0} is the concentration of component A at the reactor entrance, T_0 is the inlet temperature, C_a and C_b are the concentrations in the reactor of components A and B, respectively, T is the temperature in the reactor, F is the flow rate through the reactor, V is the reactor volume, ρ is the liquid density, c_p is the specific heat capacity of the liquid, k_w is the heat transfer coefficient, A_R is the surface area of the reactor, $(-\Delta H_{AB})$, $(-\Delta H_{BC})$ and $(-\Delta H_{AD})$ are the heat of each reaction, and T_K is the temperature in the cooling jacket. The parameter values are shown in Table 3.1.

Modeling assumptions: perfect mixture in the reactor; constant specific density and calorific capacity of the liquid; constant volume V; the dynamic of the cooling jacket is neglected; reaction turning A into B as being of second order with respect to A; reaction turning B into C as being of first order with respect to B; reaction turning A into D as being of first order with respect to A; specific reaction rates are temperature dependent, according to Arrhenius' Equation 3.27, with T in degrees Celsius:

$$k_i = k_{i_0}\, exp\left(\frac{-E_i/R}{T + 273.15}\right) \tag{3.27}$$

where E_i, with $i = 1, 2, 3$, are the activation energy of the three different reactions; k_{i_0}, with $i = 1, 2, 3$, are the pre-exponential constants of the three different reactions; and R is the gas constant.

The chosen input variables of the system are the cooling jacket temperature, T_K, and the ratio F/V. The chosen output variables of the system are the concentration of component A in the reactor, C_a, the concentration of component B, also in the reactor, C_b, and the temperature of the reactor, T. The operating intervals were chosen in order to contain the intervals where the system has nonlinear behavior. They are described below:

$$12\ h^{-1} \leqslant F/V \leqslant 132\ h^{-1} \tag{3.28}$$

$$68\ ℃ \leqslant T_K \leqslant 188\ ℃$$

Table 3.1: Reactor parameters and their values (TRIERWEILER, 1997):

Parameter	Value	Unit
k_{1_0}	1.287x10^{12}	h^{-1}
k_{2_0}	1.287x10^{12}	h^{-1}
k_{3_0}	9.043x10^{9}	L mol A^{-1} h^{-1}
$-E_1/R$	-9,758.3	K
$-E_2/R$	-9,758.3	K
$-E_3/R$	-8,560.0	K
$(-\Delta H_{AB})$	-4.20	kJ mol A^{-1}
$(-\Delta H_{BC})$	11.00	kJ mol B^{-1}
$(-\Delta H_{AD})$	41.85	kJ mol A^{-1}
ρ	0.9342	kg L^{-1}
c_p	3.01	kJ kg^{-1} K^{-1}
k_w	4,032.0	kJ h^{-1} K^{-1} m^{-2}
A_R	0.215	m^2
V	10	L
T_0	130	℃
C_{a_0}	5.1	mol A L^{-1}

3.8.2 Oil Production System with Two Gas-Lift Wells

In oil production, it is desired that the natural pressure inside the reservoir is sufficient to lift the oil upwards the well to the topside facility. When it is not, artificial ways can be employed (e.g. boosting, water injection, gas lift). One technology widely employed is the gas-lift method, represented in Figure 3.7. It consists in the injection of compressed gas at the bottom of the well. The fluid from the reservoir enters the tube from the bottom, mixes with the lift gas, then flows through the common riser manifold and goes to the topside processing facility, so it can separate the oil and gas phases (KRISHNAMOORTHY *et al.*, 2018). Gas lift reduces the density of the fluid column, which reduces the hydrostatic pressure drop in the well and decreases the bottomhole pressure. However, if too much gas is injected, frictional pressure drop may increase to levels where an increase in gas injection rate may reduce the amount of produced oil. Therefore, the objective function is to find a desirable gas lift injection rate that maximizes oil production.

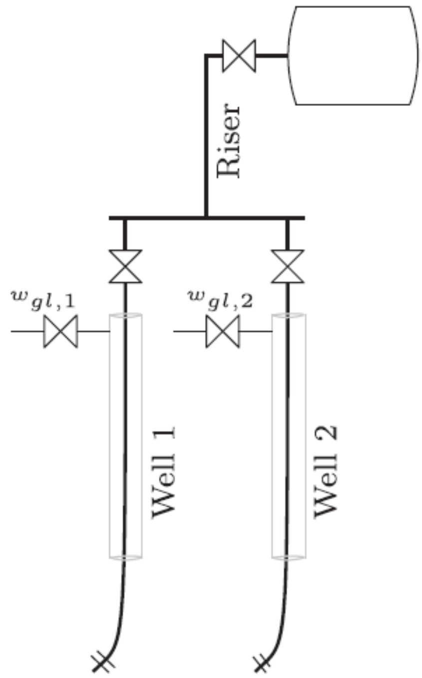


Figure 3.7: Two gas-lift wells scheme. From KRISHNAMOORTHY *et al.* (2018).

The system model used in this work was developed by KRISHNAMOORTHY *et al.* (2018) for two wells (oil production system with two gas-lift wells). The differential equations are given by Equations 3.29-3.33.

$$\dot{m}_{ga_i} = w_{gl_i} - w_{iv_i} \tag{3.29}$$

$$\dot{m}_{gt_i} = w_{iv_i} - w_{pg_i} + w_{rg_i} \tag{3.30}$$

$$\dot{m}_{ot_i} = w_{ro_i} - w_{po_i} \tag{3.31}$$

$$\dot{m}_{gr} = \sum_{i=1,2} w_{pg_i} - w_{tg} \tag{3.32}$$

$$\dot{m}_{or} = \sum_{i=1,2} w_{po_i} - w_{to} \tag{3.33}$$

where m_{ga_i} is the mass of gas in the annulus, w_{gl_i} is the mass rate of gas lift injection, w_{iv_i} is the gas flow rate from the annulus into the tubing, m_{gt_i} is the mass of gas in the well tubing, m_{ot_i} is the mass of oil in the well tubing, w_{pg_i} is the flow rate of produced gas, w_{rg_i} is the gas flow rate from the reservoir, w_{ro_i} is the oil flow rate from the reservoir and w_{po_i} is the produced oil flow rate, m_{gr} is the mass of gas in the riser, m_{or} is the mass of oil in the riser, w_{tg} is the total gas flow rate and w_{to} is the total oil flow rate; and i stands for each well, with $i = 1, 2$.

The algebraic equations are given by Equations 3.34-3.51.

$$\rho_{a_i} = \frac{M_w p_{a_i}}{T_{a_i} R} \tag{3.34}$$

$$\rho_{w_i} = \frac{m_{gt_i} + m_{ot_i} - \rho_o L_{bh_i} A_{bh_i}}{L_{w_i} A_{w_i}} \tag{3.35}$$

$$\rho_r = \frac{m_{gr} + m_{or}}{L_r A_r} \tag{3.36}$$

$$p_{a_i} = \left(\frac{T_{a_i} R}{M_w V_{a_i}} + \frac{g H_{a_i}}{V_{a_i}} \right) m_{ga_i} \tag{3.37}$$

$$p_{wh_i} = \frac{T_{w_i} R}{M_w} \left(\frac{m_{gt_i}}{L_{w_i} A_{w_i} + L_{bh_i} A_{bh_i} - \dfrac{m_{ot_i}}{\rho_o}} \right) - \frac{1}{2} \left(\frac{m_{gt_i} + m_{ot_i}}{L_{w_i} A_{w_i}} g H_{w_i} \right) \tag{3.38}$$

$$p_{wi_i} = p_{wh_i} + \frac{g}{L_{w_i} A_{w_i}} (m_{ot_i} + m_{gt_i} - \rho_o L_{bh_i} A_{bh_i}) H_{w_i} + \Delta p^t_{fric} \tag{3.39}$$

$$p_{bh_i} = p_{wi_i} + \rho_{w_i} g H_{bh_i} + \Delta p^{bh}_{fric} \tag{3.40}$$

$$p_{rh} = \frac{T_r R}{M_w} \left(\frac{m_{gr}}{L_r A_r} \right) \tag{3.41}$$

$$p_m = p_{rh} + \rho_r g H_r + \Delta p^r_{fric} \tag{3.42}$$

$$w_{iv_i} = C_{iv_i} \sqrt{max(0, \rho_{a_i}(p_{a_i} - p_{wi_i}))} \tag{3.43}$$

$$w_{pc_i} = C_{pc_i} \sqrt{max(0, \rho_{w_i}(p_{wh_i} - p_m))} \tag{3.44}$$

$$w_{pg_i} = \frac{m_{gt_i}}{m_{gt_i} + m_{ot_i}} w_{pc_i} \tag{3.45}$$

$$w_{po_i} = \frac{m_{ot_i}}{m_{gt_i} + m_{ot_i}} w_{pc_i} \tag{3.46}$$

$$w_{ro_i} = PI_i (p_{r_i} - p_{bh_i}) \tag{3.47}$$

$$w_{rg_i} = GOR_i \cdot w_{ro_i} \tag{3.48}$$

$$w_{rh} = C_{rh} \sqrt{\rho_r (p_{rh} - p_s)} \tag{3.49}$$

$$w_{tg} = \frac{m_{gr}}{m_{gr} + m_{or}} w_{rh} \tag{3.50}$$

$$w_{to} = \frac{m_{or}}{m_{gr} + m_{or}} w_{rh} \tag{3.51}$$

where ρ_{a_i} is the density of gas in the annulus, M_w is the molecular weight of the gas, R is the gas constant, T_{a_i} is the temperature in the annulus, ρ_{w_i} is the fluid mixture density in the tubing, ρ_o is the density of oil in the reservoir, L_{bh_i} and L_{w_i} are the lengths of each well below and above the injection point, respectively, A_{bh_i} and A_{w_i} are the cross-sectional areas of each well below and above the injection point, respectively. L_r and A_r are the length and cross-sectional area of the riser manifold.

p_{a_i}, L_{a_i}, A_{a_i} and V_{a_i} are the pressure, length, cross-sectional area and volume of each annulus, p_{wi_i} is the well injection point pressure, g is the acceleration of gravity constant, p_{wh_i} is the well-head pressure, H_{bh_i} and H_{w_i} are the vertical heights of each well tubing below and above the injection point, T_{w_i} is the temperature in each well tubing, p_{bh_i} is the bottom hole pressure, Δp^t_{fric} and Δp^{bh}_{fric} are the frictional pressure drop in the well tubing above and below the injection point, respectively. p_m is the manifold pressure, p_{rh} is the riser head pressure, L_r, A_r, H_r, T_r and Δp^t_{fric} are the length, cross-sectional area, vertical height, average temperature and frictional pressure drop in the riser. w_{iv_i} is the flow through the downhole gas lift injection valve, w_{pc_i} is the total flow through the production choke, C_{iv_i} and C_{pc_i} are the valve flow coefficients for the downhole injection valve and the production choke for each well, respectively. PI_i is the reservoir production index, p_{r_i} is the reservoir pressure and GOR_i is the gas-oil ratio. w_{rh} is the flow through the riser head choke, C_{rh} is the valve flow coefficient for the riser head valve and p_s is the separator pressure. The parameter values are in Tables 3.2 and 3.3.

Table 3.2: List of well parameter values (KRISHNAMOORTHY *et al.*, 2018):

Parameter	Well 1	Well 2	Units
L_w	1500	1500	m
H_w	1000	1000	m
D_w	0.121	0.121	m
L_bh	500	500	m
H_bh	500	500	m
D_bh	0.121	0.121	m
L_a	1500	1500	m
H_a	1000	1000	m
D_a	0.189	0.189	m
ρ_o	800	800	kg m^{-3}
C_iv	1x10^{-4}	1x10^{-4}	m^2
C_pc	2x10^{-3}	2x10^{-3}	m^2
p_r	150	155	bar
PI	0.7	0.7	kg s^{-1} bar^{-1}
T_a	28	28	℃
T_w	32	32	℃
GOR	0.1±0.05	0.12±0.02	kg/kg

Table 3.3: List of riser parameter values (KRISHNAMOORTHY *et al.*, 2018):

Parameter	Value	Units
L_r	500	m
H_r	500	m
D_r	0.121	m
C_rh	1x10^{-2}	m^2
p_s	20	bar
T_r	30	℃
M_w	20	g mol^{-1}
R	8.314	J mol^{-1} K^{-1}

The first principle model has a total of five differential equations and eighteen algebraic equations. Twelve variables were chosen to be outputs, $p_{wh_i}, p_{bh_i}, w_{pg_i}, w_{po_i}, p_{rh}, p_m, w_{to}, w_{tg}$, and two more to be the inputs, w_{gl_i}, $i = 1, 2$. The operating interval of the input variables is given by the equation below:

$$0 < w_{gl_i} < 8kgs^{-1} \tag{3.52}$$

Chapter 4

Results and Discussion

First of all, a study of types of disturbances was made in order to choose the type that is more adequate for nonlinear identification.

For each case study, before getting data from first principle models, a study of the process was made to know more about it, learning about its nonlinear behavior, calculate the time for the system to establish and noticing any delays.

After acquiring data, the black-box and gray-box identification were performed for each output at a time. The comparison of the two types of identification is made in the last section.

4.1 Pre-test

A series of pre-tests were made on the non-isothermal Van de Vusse reactor, testing types of input disturbances, and the results were extended to the second case study, the oil production system with two gas-lift wells. PRBS, multisine and RRSS were applied on each input variable of the Van de Vusse reactor. PRBS, in Figure 4.1 was generated with the order being set to 3 (for u_1) and 2 (for u_2).

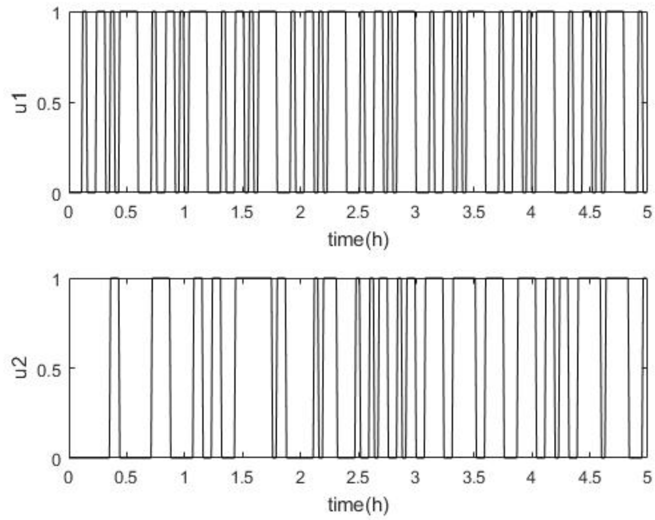


Figure 4.1: Pseudo-Random Binary Sequence.

In Figure 4.2, Multisine signal was generated by a linear combination of sines with random argument between 0 and $2\pi\omega t$, where ω was set to $20h^{-1}\times$ random value between 0 and 1, t was set to 501 points from 0 and 5 hours.

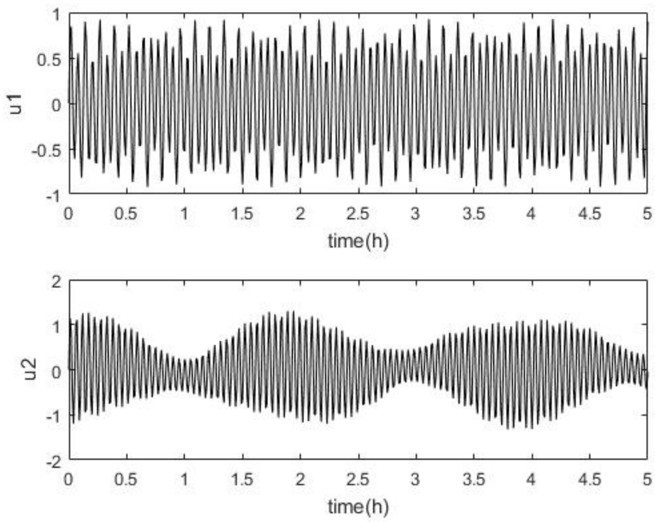


Figure 4.2: Multisine signal before sampling step.

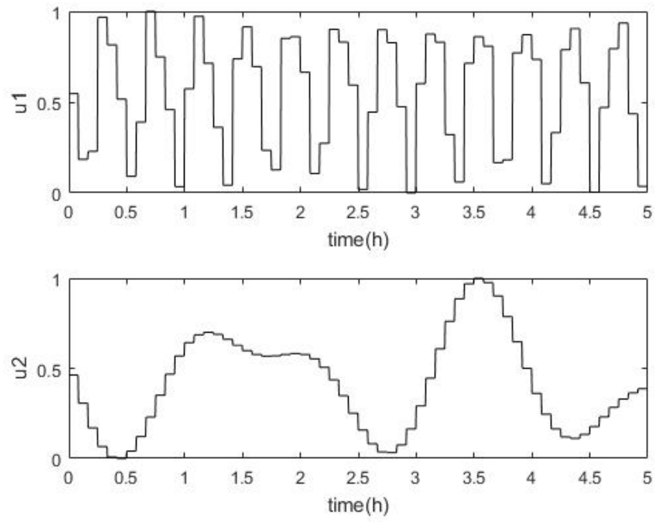


Figure 4.3: Multisine signal after sampling step.

In Figure 4.4, the RRSS signal was generated with 50 steps with duration of $0.1h$ each.

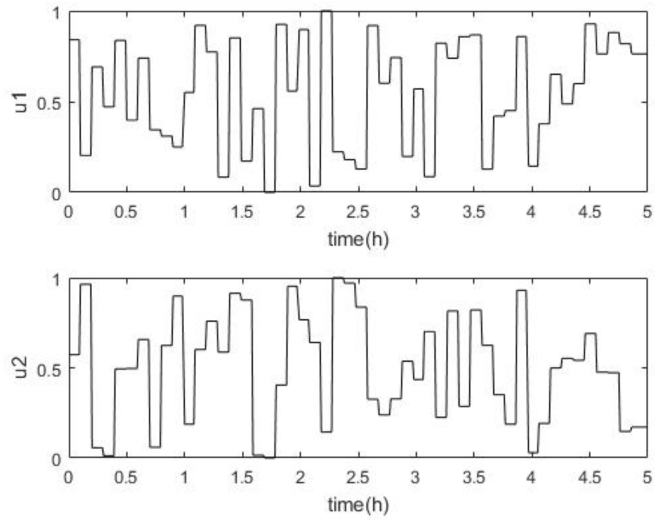


Figure 4.4: Random Range Step Sequence.

In Figure 4.1, it is clear that PRBS has some disadvantages when it comes to nonlinear identification. It does not cover a wide range of operating conditions, as it is inherently a local signal; at some point, the number of samples is not enough for the system to settle; or it has so many samples at some point that the system stays too long in steady state.

Figure 4.3 shows two possibilities when it comes to multisine signals, that is, it can turn into a signal with all the requirements to be a good input disturbance (u_1 case), or it can turn into a bad one (u_2 case), which does not cover abrupt changes on the input (one of the facts that may cause nonlinear behavior). In order to improve u_2 signal, the time sample period should be smaller, but the probability of getting a bad signal would still exist, as the input frequency is random.

On the other hand, Figure 4.4 shows a good input disturbance, as it covers all the operating range, it can be manipulated to stick to one value as long as the system needs to establish and contains abrupt change of values.

The process was identified using artificial neural networks (ANN). The identification procedure was performed in the software Statistica version 16 using multi-layer perceptrons (MLP) with regressors presented in Equation 4.1, as in BIJANZADEH *et al.* (2013) (higher maximum lags did not give much improvement in the identification). During training stage, 54 ANN were trained for the variable C_b, varying the number of neurons of the hidden layer (from 6, which is the number of regressors, to 15, there is not much improvement, in this case, for higher values), the activation function of the hidden nodes (logistic function or hyperbolic tangent) and the activation function of the output node (logistic function or hyperbolic tangent or identity). From those 54 ANN, the 5 best go to the validation stage. The validation method was the one-step-ahead cross-validation, which consists in the usage of data from the other types of signals for one-step-ahead validation.

$$y(k) = D[y(k-1), y(k-2), u_1(k-1), u_2(k-1), u_1(k-2), u_2(k-2)] \tag{4.1}$$

where D is the ANN model.

After that, the ANN performance was compared using the sum of quadratic fitting errors (SQE) and the determination coefficient between predicted and simulated data, using the other types of disturbance. In Table 4.1, the sum of squared error of the identification procedure is recorded; and in Table 4.2, each column shows the correlation coefficient between predicted data using the corresponding ANN and the reactor data, simulated for each type of disturbance.

Table 4.1: Sum of quadratic fitting error.

	J_{MS}	J_{RRSS}	J_{PRBS}
MS	0.1629	0.3712	8.7533
RRSS	0.5577	0.0509	4.0849
PRBS	3.2060	1.9091	0.5836

Table 4.2: R-squared values for identified ANN simulation from each type of disturbance.

	ANN-MS	ANN-RRSS	ANN-PRBS
MS	0.9990	0.9985	0.9723
RRSS	0.0255	0.9978	0.0301
PRBS	0.9942	0.9969	0.9987

Analyzing Table 4.1, where ANN-MS is the ANN identified using multisine signal to disturb the input variable, ANN-RRSS is the ANN identified using RRSS and ANN-PRBS is the ANN using PRBS. It can be observed, in the last column, high values of SQE, when validating ANN-PRBS. This corroborates the statement made by LEONTARITIS & BILLINGS (1987a) about the ineffectiveness of PRBS signals in the nonlinear identification. When comparing the correlation coefficient of ANN-MS using RRSS and ANN-RRSS using MS, in Table 4.2, it can be noted that ANN-MS does not have a good representation of the system when it is disturbed by a RRSS, *i.e.*, the ANN-RRSS adapts better to the other types of input disturbances.

This result was extended to the second case study. Therefore, the type of signal used in this work is a sequence of random range steps containing all possible operating conditions, with the same time period, being long enough to the system establish itself, but not too long as it makes the nonlinear identification more difficult (LEONTARITIS & BILLINGS, 1987a).

4.2 First Case Study

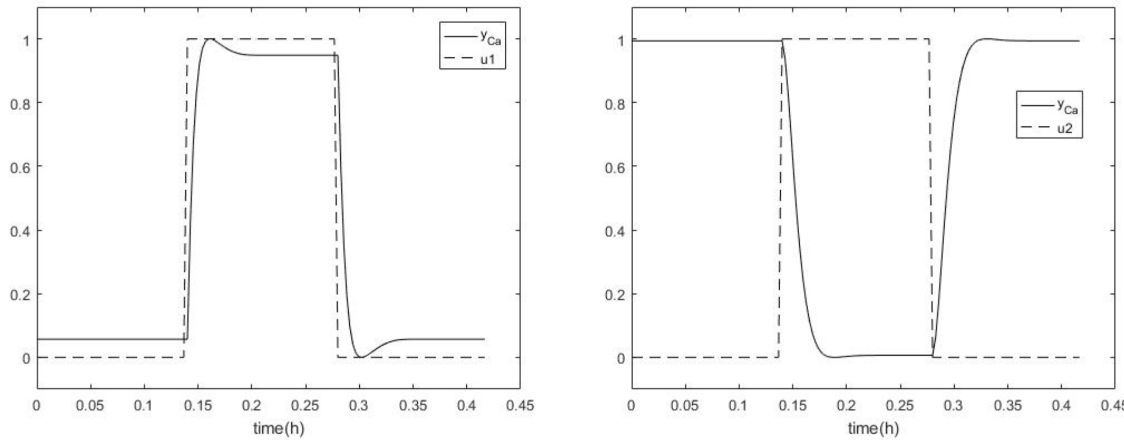
Two input variables (F/V and T_K) and three output variables (C_a, C_b and T) were selected to study the Van de Vusse reactor.

4.2.1 Gathering Information

In order to learn more about the process, the inputs were disturbed one at a time on an open loop simulation with sample time period of $0.0028h$ or $10s$, which was chosen to give a smooth output data. For the first test, after $0.14h$ of simulation, u_1 was given an unit pulse during $0.14h$, and for the second test, after $0.14h$ of simulation, u_2 was given an unit pulse with the same duration. All variables were normalized.

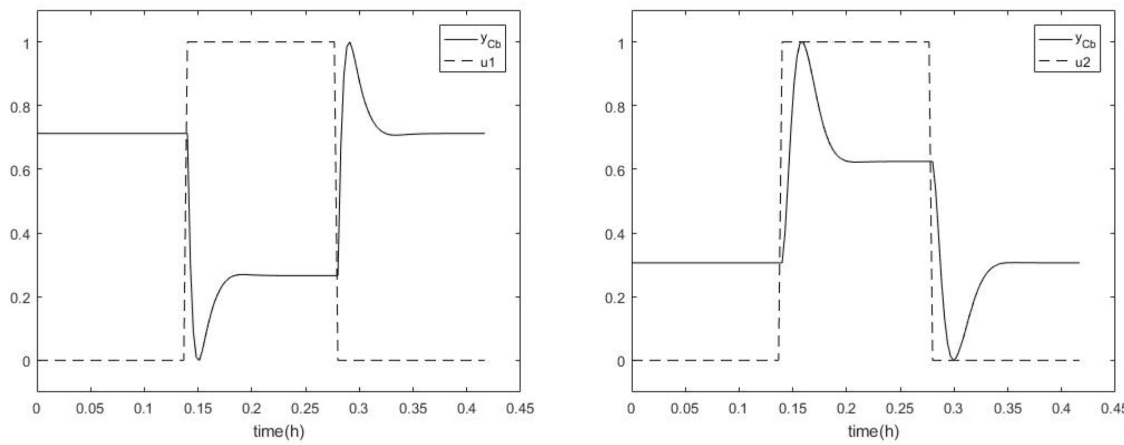
In Figures 4.5-4.7, it can be observed that all variables have no delays; C_a response has no delay and has negative gain in relation to u_2; C_b has negative gain in relation to u_1 and has a high overshoot, comparing to the other variables; and T

has inverse response in relation to u_1. The system takes at most $0.014h$ to reach the steady-state, so the PH is set to 25.



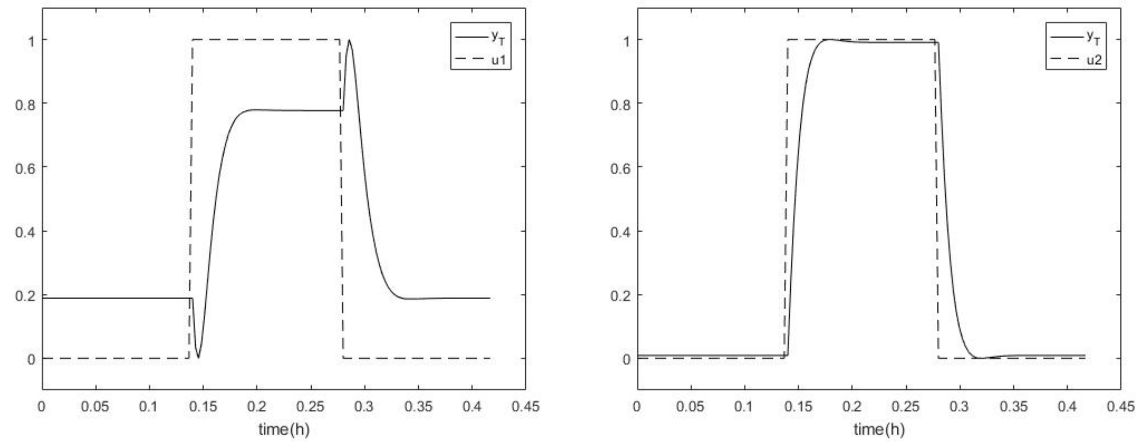
(a) (b)

Figure 4.5: Response of C_a due to disturbance on the inputs: (a) F/V; (b) T_K.



(a) (b)

Figure 4.6: Response of C_b due to disturbance on the inputs: (a) F/V; (b) T_K.



(a) (b)

Figure 4.7: Response of T due to disturbance on the inputs: (a) F/V; (b) T_K.

4.2.2 Data Acquisition

The system was simulated based on first principle models in order to represent the real behavior on a plant. It was executed on the programming environment MATLAB, where a series of 40 random steps with duration of $0.07h$ each were applied on the input variables (F/V and T_K) with operating intervals described in Equation 4.2 and sample time period of $10s$, as shown in Figures 4.8 and 4.9. It is important to emphasize that all variables were normalized and corrupted with white noise.

$$
\begin{aligned}
12h^{-1} \leqslant u_1 \leqslant 132h^{-1} \\
68℃ \leqslant u_2 \leqslant 188℃
\end{aligned}
\tag{4.2}
$$

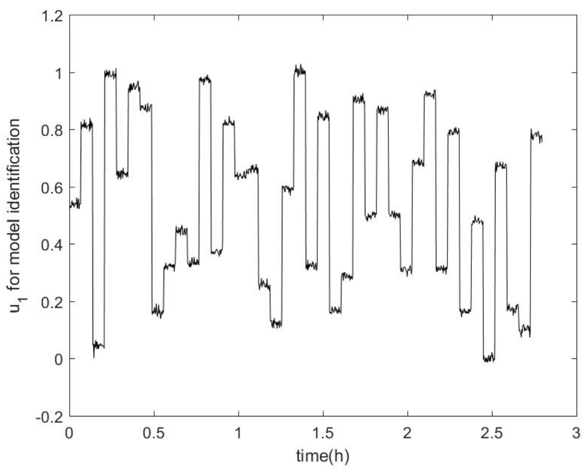


Figure 4.8: Simulated data - variable u_1 (F/V).

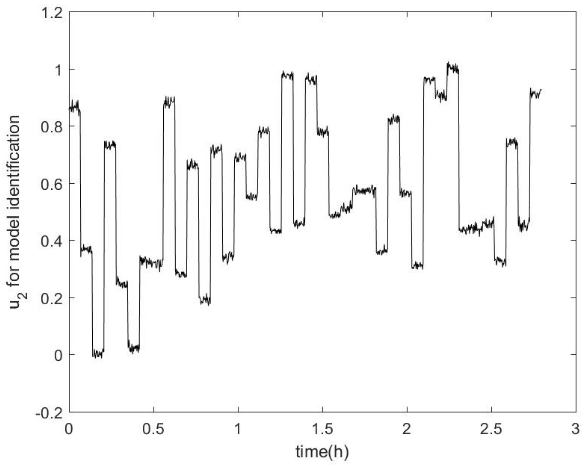


Figure 4.9: Simulated data - variable u_2 (T_K).

4.2.3 Black-box Identification

4.2.3.1 Parameter Estimation

The order parameters n_y, n_{ui} and n_e varied from 1 to 4 (higher values did not give much improvement to the model)and the delays were set to zero as a result from the previous section.

The optimal values of these parameters are listed in Table 4.3.

Table 4.3: Optimal values of order parameters of black-box identification.

Parameter	C_a	C_b	T
$n_{p_{NARX}}$	14	14	14
ℓ	2	2	2
n_y	2	4	4
n_{u_1}	4	3	4
n_{u_2}	2	4	2
n_e	3	4	3
$n_{p_{MA}}$	1	1	1

All identified models have second order nonlinearity degree and a total of 15 features.

The objective function values for the optimal solutions are in Table 4.4.

Table 4.4: Objective function for optimal parameters of black-box identification.

Variable	J_{OLS}
C_a	0.0324
C_b	0.0102
T	0.0106

The identified models are presented in Equations 4.3-4.5:

$$\hat{\boldsymbol{y}}_{C_a} = \boldsymbol{\Psi}^{*}_{C_a}\boldsymbol{\theta}_{C_a} \tag{4.3}$$

$$\boldsymbol{\Psi}^{*,T}_{C_a} = \begin{bmatrix} y_{C_a}(k-1) \\ y_{C_a}(k-2) \\ u_1(k-1) \\ u_1(k-2) \\ y_{C_a}(k-1)u_2(k-1) \\ y_{C_a}(k-1)^2 \\ u_2(k-1)u_2(k-2) \\ y_{C_a}(k-1)u_1(k-1) \\ y_{C_a}(k-2)u_1(k-2) \\ y_{C_a}(k-1)u_1(k-4) \\ u_1(k-4)^2 \\ u_2(k-1) \\ u_1(k-4) \\ y_{C_a}(k-1)u_1(k-3) \\ e(k-1) \end{bmatrix} \qquad \boldsymbol{\theta}_{C_a} = \begin{bmatrix} 1.5255 \\ -0.5178 \\ 0.3364 \\ -0.2199 \\ -0.0780 \\ 0.0107 \\ -0.0626 \\ -0.2279 \\ 0.1653 \\ -0.0523 \\ 0.0445 \\ 0.0274 \\ -0.0363 \\ -0.0272 \\ -0.4492 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{C_b} = \boldsymbol{\Psi}^*_{C_b}\boldsymbol{\theta}_{C_b} \tag{4.4}$$

$$\boldsymbol{\Psi}^{*,T}_{C_b} = \begin{bmatrix} y_{C_b}(k-1) \\ y_{C_b}(k-3)y_{C_b}(k-4) \\ u_1(k-3)u_2(k-1) \\ u_1(k-1)u_2(k-4) \\ u_1(k-2)u_2(k-4) \\ y_{C_b}(k-3)u_1(k-1) \\ u_2(k-4)^2 \\ y_{C_b}(k-3)u_1(k-2) \\ u_2(k-2) \\ y_{C_b}(k-4)u_1(k-3) \\ u_1(k-1)u_1(k-3) \\ u_1(k-3) \\ y_{C_b}(k-2)^2 \\ y_{C_b}(k-1)u_1(k-1) \\ e(k-3)u_2(k-1) \end{bmatrix} \quad \boldsymbol{\theta}_{C_b} = \begin{bmatrix} 1.0080 \\ -0.1718 \\ 0.0785 \\ 0.2339 \\ 0.0036 \\ -0.3318 \\ -0.1780 \\ 0.2120 \\ 0.0699 \\ 0.0524 \\ -0.1240 \\ 0.1346 \\ 0.1325 \\ -0.2126 \\ 0.2013 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{T} = \boldsymbol{\Psi}^*_{T}\boldsymbol{\theta}_{T} \tag{4.5}$$

$$\boldsymbol{\Psi}^{*,T}_{T} = \begin{bmatrix} y_T(k-1) \\ u_2(k-1) \\ y_T(k-4)u_2(k-2) \\ u_1(k-4) \\ y_T(k-1)u_1(k-4) \\ y_T(k-1)u_1(k-1) \\ u_1(k-1) \\ y_T(k-2) \\ u_2(k-2) \\ y_T(k-3)u_2(k-1) \\ u_1(k-1)u_2(k-1) \\ u_1(k-4)^2 \\ y_T(k-3)^2 \\ y_T(k-4)^2 \\ e(k-2)u_1(k-4) \end{bmatrix} \quad \boldsymbol{\theta}_{T} = \begin{bmatrix} 0.5091 \\ 0.1287 \\ -0.0925 \\ 0.0597 \\ -0.0195 \\ -0.2578 \\ 0.1574 \\ 0.2750 \\ 0.1327 \\ 0.0565 \\ -0.0304 \\ -0.0152 \\ 0.0704 \\ -0.0398 \\ -0.1834 \end{bmatrix}$$

where $\boldsymbol{e}$ is different for each model.

The final models were simulated and the one-step-ahead prediction was compared with the normalized data. These results are shown in Figures 4.10-4.12.

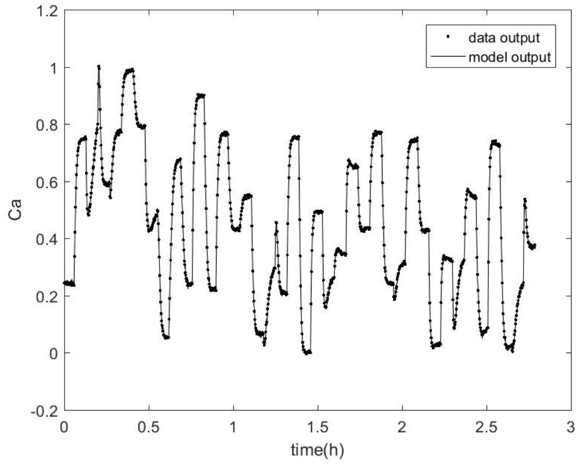


Figure 4.10: Simulation output of black-box identification for variable C_a.

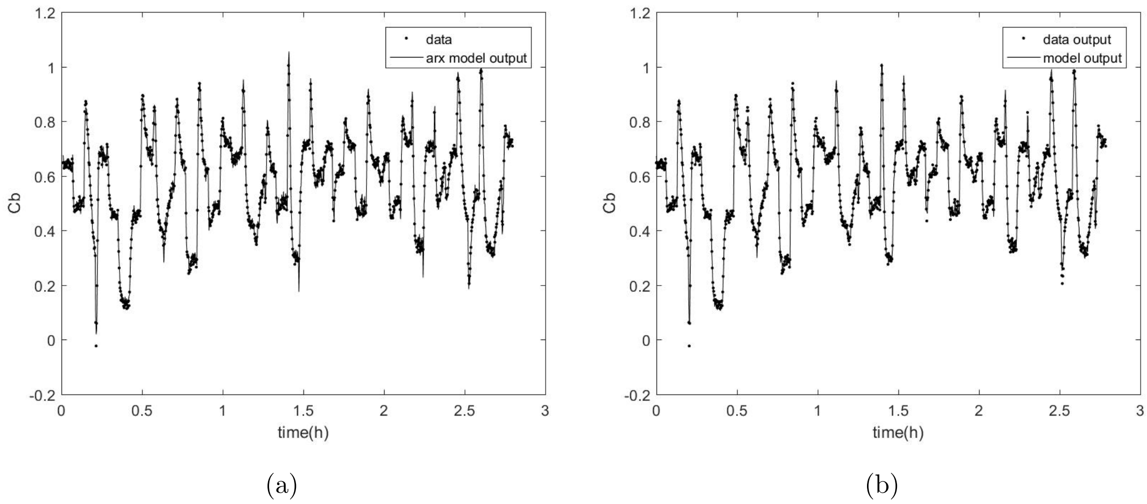


Figure 4.11: Simulation output of black-box identification for variable C_b: (a) ARX model; (b) NARMAX model.

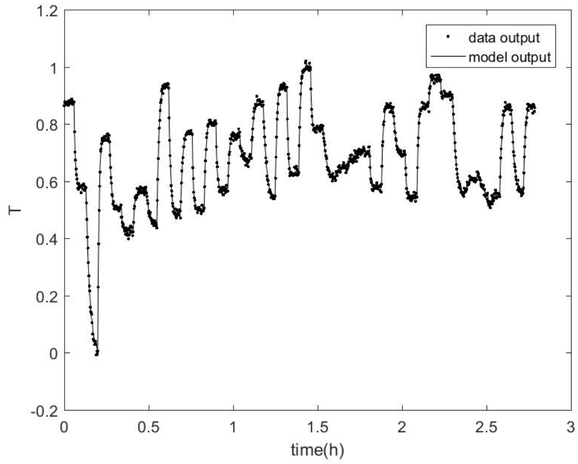


Figure 4.12: Simulation output of black-box identification for variable T.

Qualitatively, it can be said about Figures 4.10-4.12 that the identified models are quite good on describing the original data.

As ARX model is one of the most common type of model used on process control, the one-step-ahead prediction is compared. The ARX model gives more peaks than data presents for variable C_b. The differences between ARX and NARMAX models for the other variables cannot be seen graphically, so it is shown in the next section, comparing R-squared calculated with input data for model validation. In order carry out the model validation, the cross-validation is done.

4.2.3.2 Cross-validation

Figures 4.13 and 4.14 show the input data for model validation, that was generated with the same characteristics as the original set (same number of steps, same duration of each step of the data used in the identification). The same reference values were used for normalization to maintain the same scale, so it is possible that the new data set is not exactly between 0 and 1.

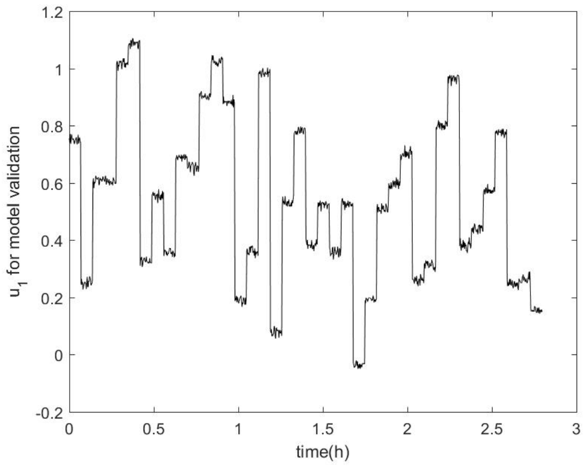


Figure 4.13: Input data u_1 for validation - First Case Study.

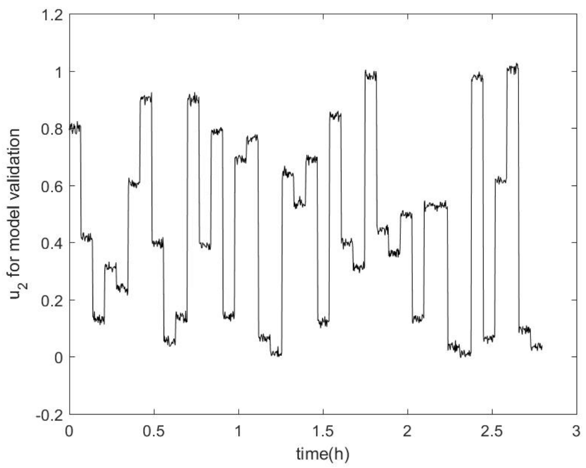


Figure 4.14: Input data u_2 for validation - First Case Study.

Figures 4.15, 4.17 and 4.19 show the k-step-ahead validation results ($k = 25$) and Table 4.5 shows the determination coefficients for the ARX model and for the NARMAX model.

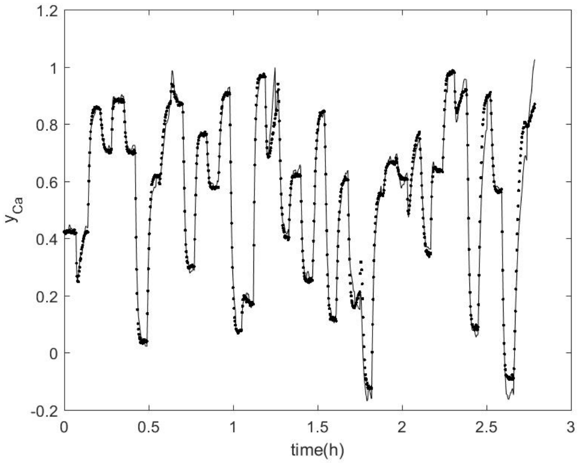


Figure 4.15: Cross-validation of C_a model using NARMAX from black-box identification.

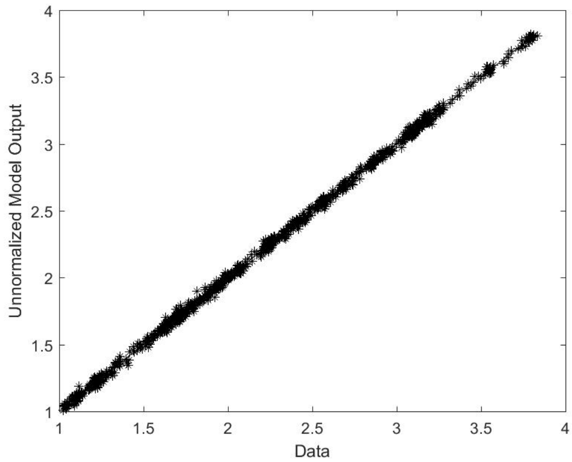


Figure 4.16: Comparison of both unnormalized predicted output using NARMAX from black-box identification, and data of variable C_a.

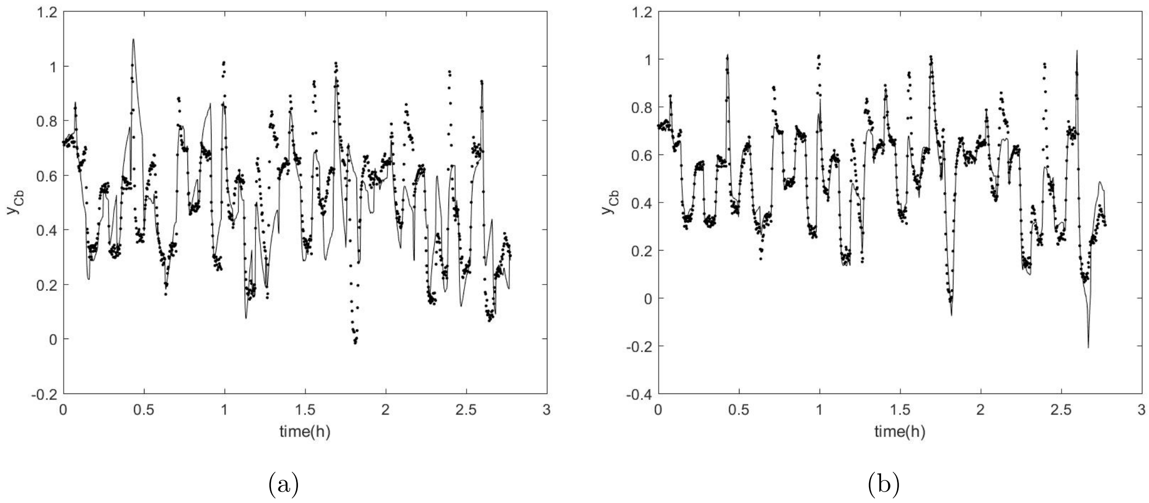


Figure 4.17: Cross-validation of C_b model from black-box identification using: (a) ARX model; (b) NARMAX model.

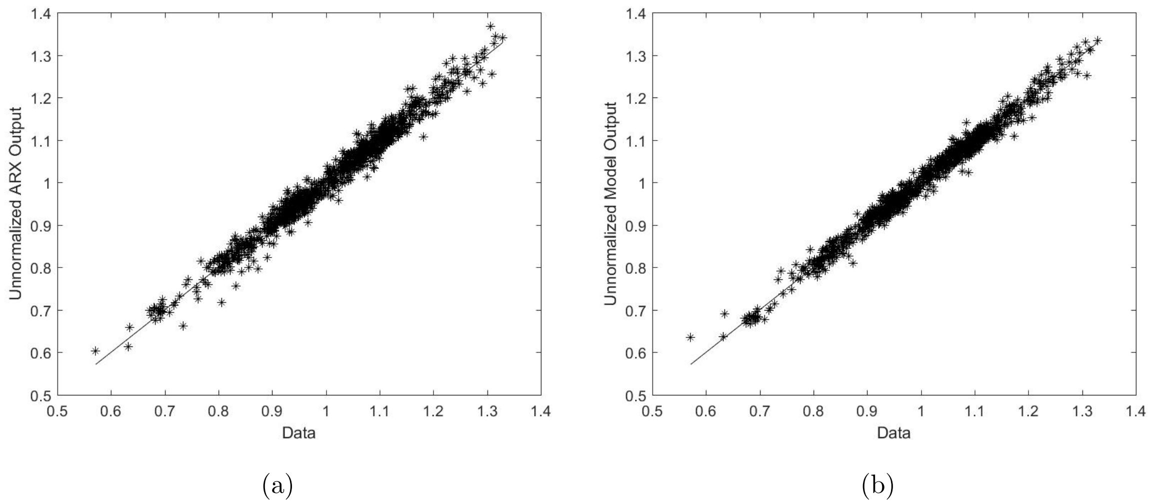


Figure 4.18: Comparison of both unnormalized predicted output from black-box identification and data of variable C_b using: (a) ARX model; (b) NARMAX model.

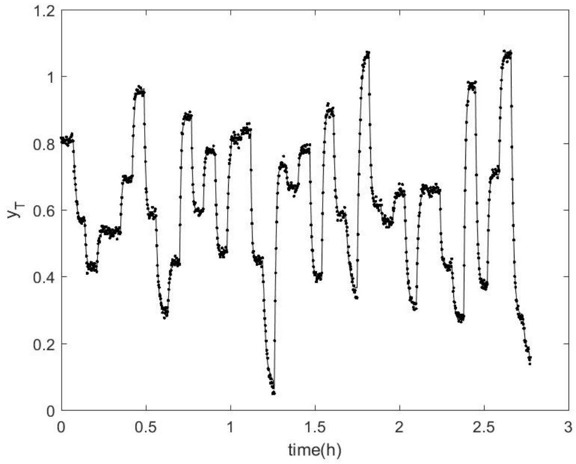


Figure 4.19: Cross-validation of T model using NARMAX from black-box identification.

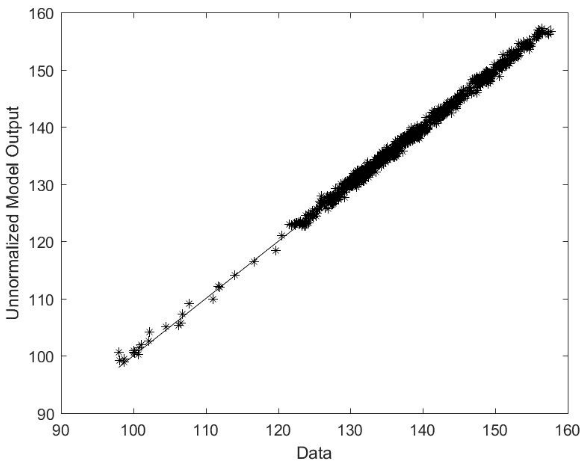


Figure 4.20: Comparison of both unnormalized predicted output using NARMAX from black-box identification and data of variable T.

Table 4.5: Determination coefficient of validation for black-box identification - First Case Study.

	R^2_{ARX}	R^2_{NARMAX}
C_a	0.8190	0.9789
C_b	0.4156	0.8590
T	0.9368	0.9971

The accuracy of prediction can be seen qualitatively in Figures 4.16, 4.18 and 4.20, and quantitatively in Table 4.5.

Qualitatively, C_a and T models seem to be more accurate on predicting the output, and C_b model is more disperse and ARX model gives a more disperse plot when comparing Figures 4.16a and 4.16b.

Comparing the values of R-squared in Table 4.5, besides reassuring that C_b model is less accurate on prediction (it has R-squared lower than 0.9), the ARX one is lower than the NARMAX one, which means that ARX model would not describe the nonlinear behavior of these variables for all the operating interval.

4.2.3.3 Dynamic Real-time Optimization

During DRTO run for the Van de Vusse reactor, F/V was varied in order to simulate changes on the flow rate and T_K was the decision variable with Equation 4.6 as economic objective function (with the optimization horizon of 50 and all variables restricted to positive values), as can be observed in Figures 4.21 and 4.22. F/V was given 3 steps: one at time = $50s$ from the $50h^{-1}$ to $100h^{-1}$, other at time = 500s from $100h^{-1}$ to $30h^{-1}$ and another at time = 1000s from $30h^{-1}$ to $72.128h^{-1}$, which is one of the steady-states of the process.

$$J_{DRTO} = \int_0^{50} (-p_{C_b} C_b + (p_{T_K} T_K)^2) dt \tag{4.6}$$

where the first term is f_1 from Equation 3.20: $p_{C_b} = 2.009$ is the price of product C_b and $p_{T_K} = 1.657 \times 10^{-4}$ is the utility cost (ALSTAD, 2005). The system is subjected to constraints:

$$u_i, y_j \geqslant 0, \ i = 1, 2 \text{ and } j = C_a, C_b, T \tag{4.7}$$

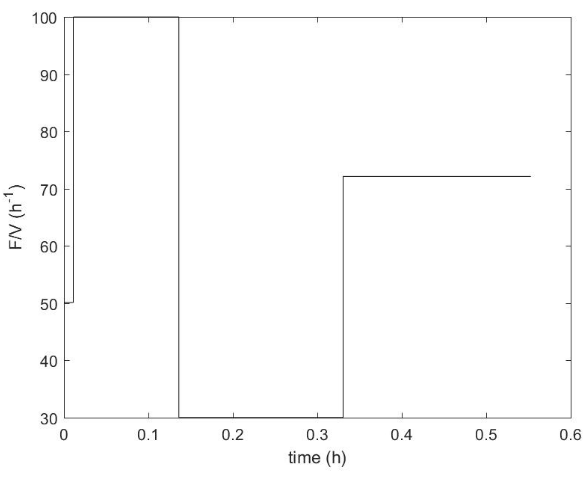


Figure 4.21: F/V variation.

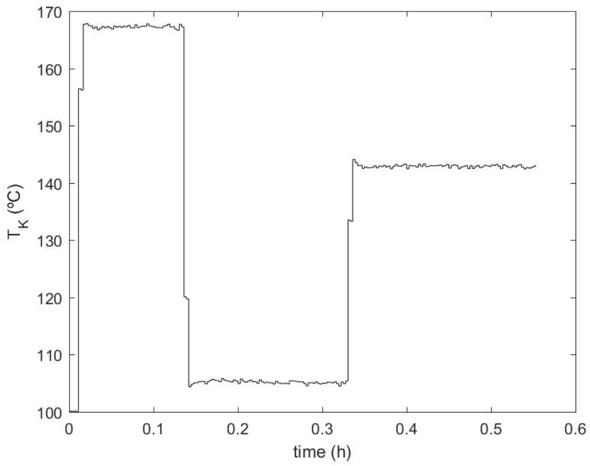


Figure 4.22: Control action on input T_K.

The closed loop responses of the black-box identified model to the DRTO actions on the temperature of the jacket (DRTO-NARMAX - DRTO based on NARMAX model) were compared with the one using first principle model (DRTO-ID - ideal DRTO), the results are shown in Figures 4.23-4.25. It can be noted that they have different solutions in all operating interval with 10% maximum difference, which is a good result because the minimum measurement error of the process is 10%.

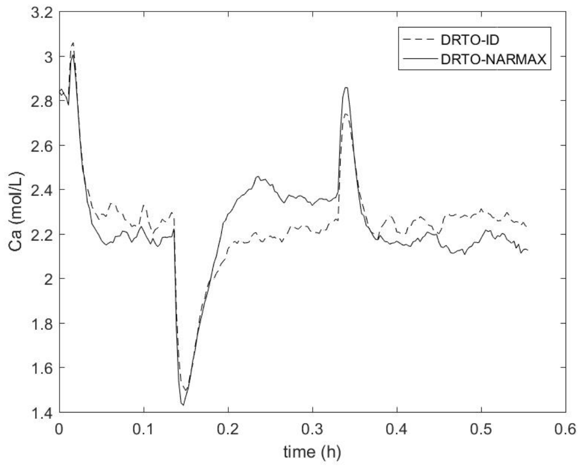


Figure 4.23: Comparison of DRTO performances to C_a.

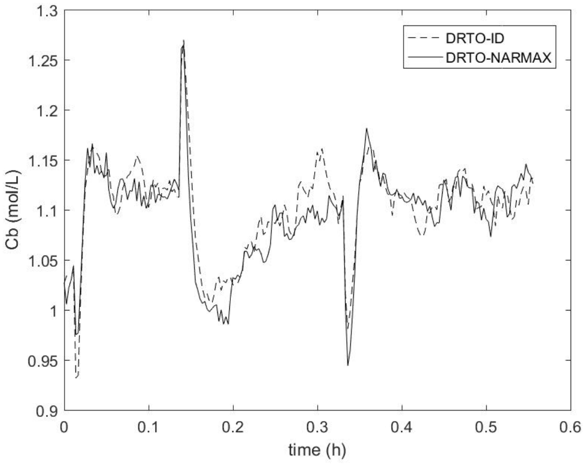


Figure 4.24: Comparison of DRTO performances for C_b.

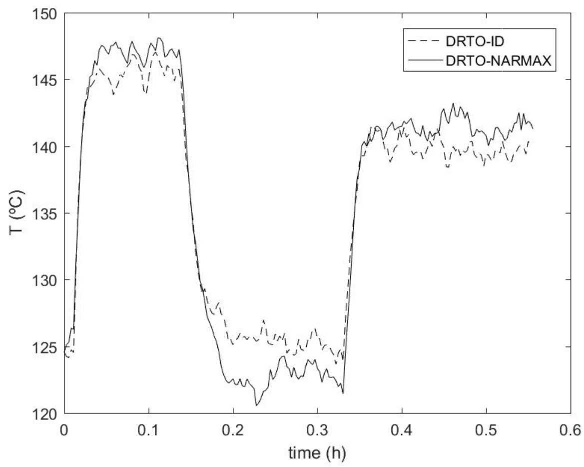


Figure 4.25: Comparison of DRTO performances for T.

In Figure 4.26, it can be seen that the objective function of the ideal scenario is much lower than the one calculated on the DRTO-NARMAX. It can be due to the fact that the model is far from perfection, *i.e.*, high uncertainty of the data and prediction error.

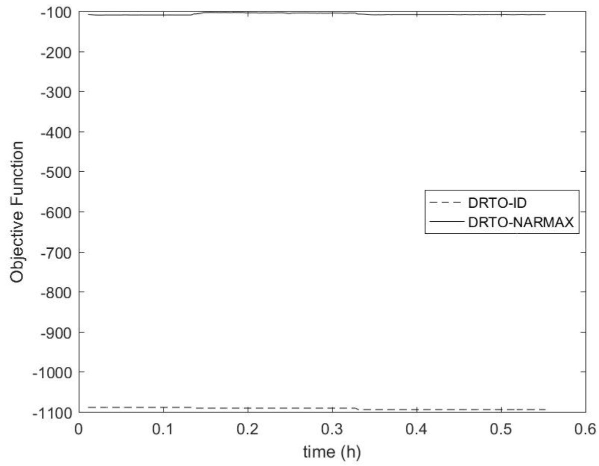


Figure 4.26: Comparison of objective function during DRTO.

4.2.4 Gray-box Identification

4.2.4.1 Parameter Estimation

The gray-box identification algorithm found no need for changing coordinates to identify C_a and T models. On the other hand, for C_b, it asked for suggestions, which are described in Appendix B, with each R-squared value. Some suggestions were made testing simple nonlinear combinations, such as $\sqrt{\boldsymbol{u_1}}$, $\boldsymbol{u_1}/\boldsymbol{u_2}$, $\boldsymbol{u_1^2}/\boldsymbol{u_2}$, *etc.*, and others based on energy or mass balance. The negative R-squared values are due to poor suggestion that makes the prediction error to be very big; and the NaN (Not-a-Number) means that at some point, there is addition of prediction errors with infinite magnitude (-Inf+Inf,).

The chosen suggestion (the one with maximum R-squared value of Table B.1) for C_b was changing $\boldsymbol{u_2}$ to $\boldsymbol{u_2^2}/\boldsymbol{u_1}$. The optimal values of the model orders are listed in Table 4.6.

Table 4.6: Optimal values of order parameters of the gray-box identification - First Case Study.

Parameter	C_b
$n_{p_{NARX}}$	14
ℓ	3
n_y	4
n_{u_1}	4
n_{u_2}	4
n_e	4
$n_{p_{MA}}$	1

The objective function value for the optimal solution is 0.0454, which is higher than the one calculated on the black-box identification, but it gives better prediction. This is due to the algorithm code that maximizes R-squared value instead of minimizing the objective function, as can be seen in the next section.

The estimated model of C_b is presented in Equation 4.8.

$$\hat{\boldsymbol{y}}_{C_b} = \boldsymbol{\Psi}^{*}_{C_b}\boldsymbol{\theta}_{C_b} \tag{4.8}$$

$$\boldsymbol{\Psi}^{*,T}_{C_b} = \begin{bmatrix} y_{C_b}(k-1) \\ u_1(k-3)^2u_2(k-1) \\ u_1(k-1)u_2(k-2)y_{C_b}(k-4) \\ u_1(k-2)u_2(k-2) \\ u_1(k-4)u_2(k-4) \\ y_{C_b}(k-3) \\ y_{C_b}(k-1)u_1(k-1)^2 \\ u_2(k-1)u_2(k-4)u_1(k-4) \\ u_1(k-3)u_2(k-1) \\ u_1(k-2)^2u_1(k-4) \\ u_1(k-1) \\ u_1(k-3)u_2(k-2)y_{C_b}(k-1) \\ u_1(k-1)u_1(k-2)y_{C_b}(k-1) \\ u_1(k-2)u_2(k-1)^2 \\ e(k-1)u_1(k-3)y_{C_b}(k-4) \end{bmatrix} \quad \boldsymbol{\theta}_{C_b} = \begin{bmatrix} 1.1595 \\ -0.2130 \\ -0.2899 \\ 0.5433 \\ 0.1185 \\ -0.1823 \\ -0.2495 \\ -1.3865 \\ 0.7723 \\ 0.0754 \\ -0.0483 \\ -0.3428 \\ 0.1102 \\ -0.2467 \\ -0.7995 \end{bmatrix}$$

The final model was simulated and the one-step-ahead simulation was compared with the normalized data, as shown in Figure 4.27. It can be seen that the model describes the normalized data very well.

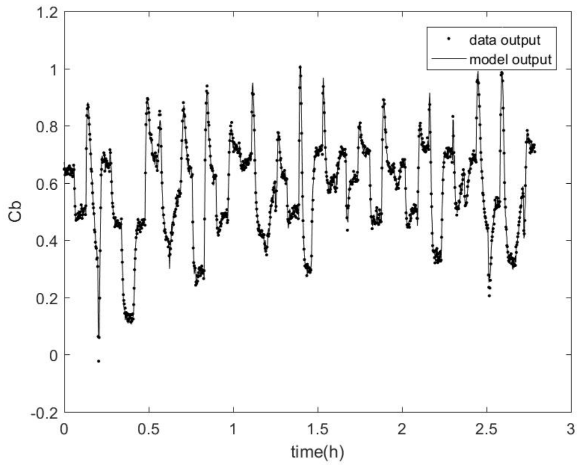
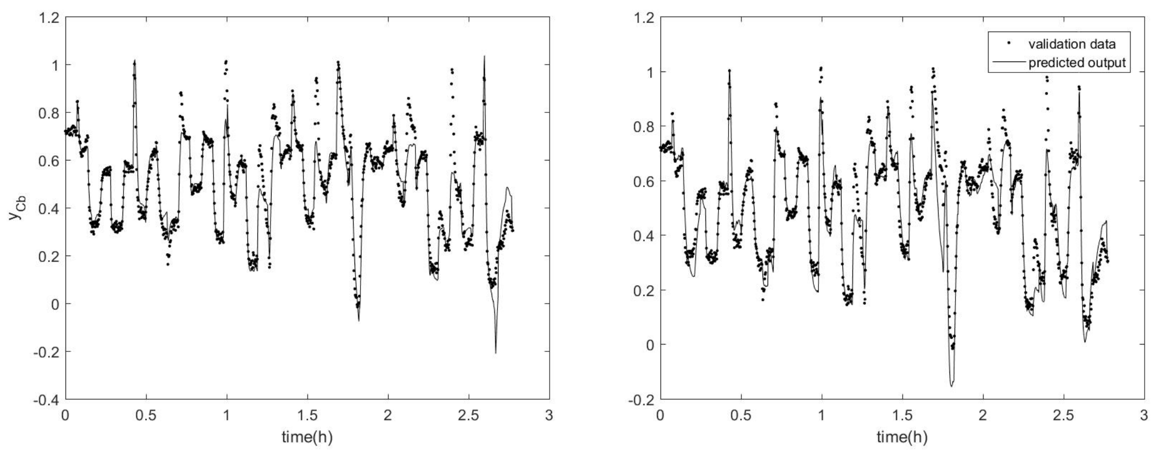


Figure 4.27: Simulation of gray-box identification for variable C_b.

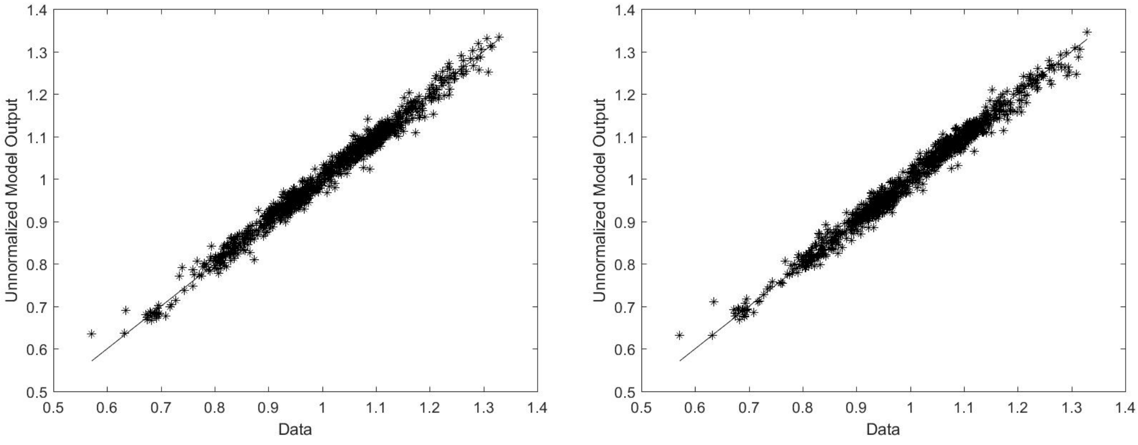
4.2.4.2 Cross-validation

Input data for validation of the gray-box identification model was the same as for the black-box one. This data was used before the change in coordinates, so $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ are used to calculate the new input variables, which are referred also as $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$, but they can be different for every identification procedure. In this case, as a result of the previous section, $\boldsymbol{u_2}$ values has been changed to $\boldsymbol{u_2}^2/\boldsymbol{u_1}$.



(a) (b)

Figure 4.28: Cross-validation of C_b model from (a) black-box identification; (b) gray-box identification.



(a) (b)

Figure 4.29: Comparison of both unnormalized predicted output and data of variable C_b using NARMAX (a) black-box identification; (b) gray-box identification.

The R-squared for the gray-box identification of C_b model is 0.8729, which is better than in black-box identification. In Figure 4.28b, it can be noticed, when comparing with Figure 4.28a, that the model can now provide a better representation of abrupt changes on the gain. And Figure 4.29b shows as much dispersion as in black-box validation, in Figure 4.29a.

4.2.4.3 Dynamic Real-time Optimization

The same procedure was done for the gray-box NARMAX model-based DRTO, F/V was varied and T_K was the decision variable, as can be observed in Figures 4.30 and 4.31.

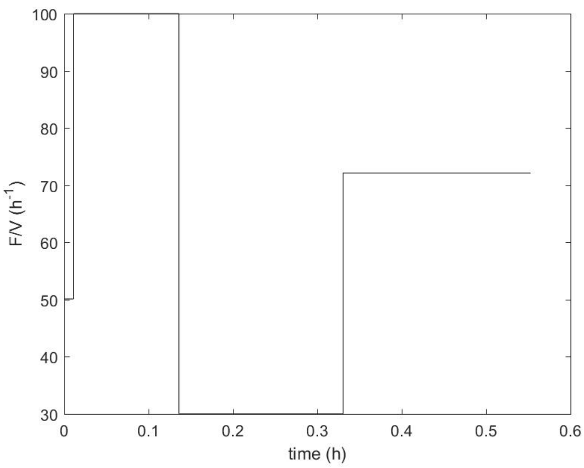


Figure 4.30: F/V variation.

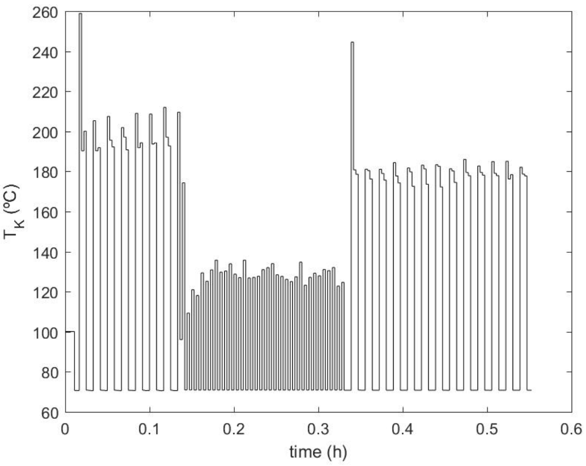


Figure 4.31: Control action on input T_K.

The closed loop responses of the identified model of the system to the DRTO actions on the temperature of the jacket were compared with the one using the first principle model, and the results are shown in Figures 4.32-4.34.

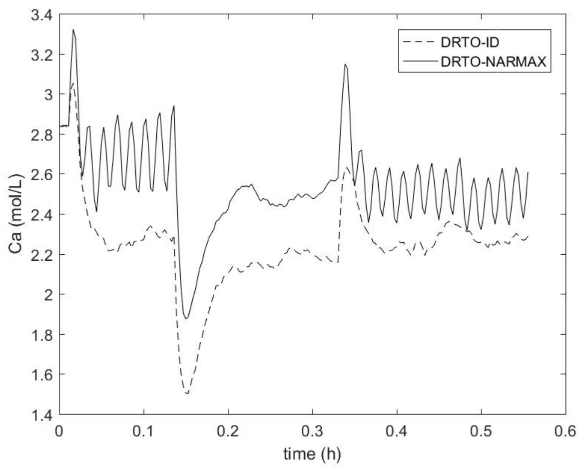


Figure 4.32: Comparison of DRTO performances for C_a.

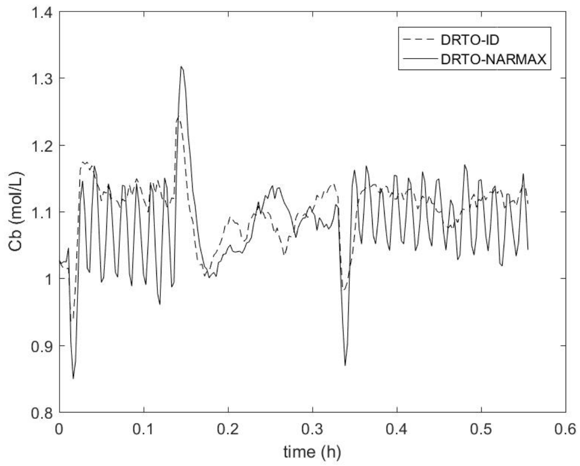


Figure 4.33: Comparison of DRTO performances for C_b.

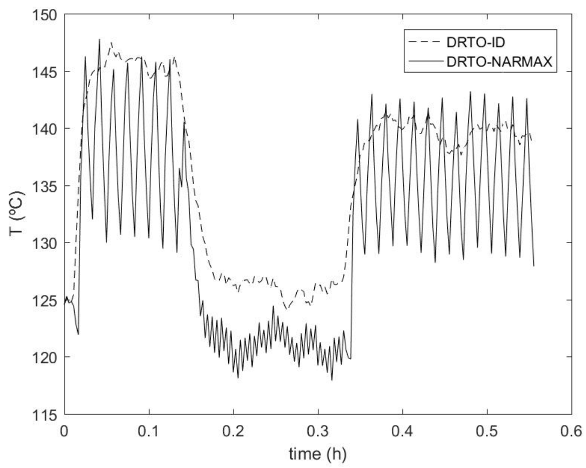


Figure 4.34: Comparison of DRTO performances for T.

The change on C_b model made the optimization unstable, which can be noted by the oscillations in Figures 4.32-4.34. It can be due to the fact that the optimization code sets the control horizon to be the same as the prediction horizon. In Figure 4.35, it can be seen that the objective function of the DRTO-NARMAX has much higher values than the ones in the ideal scenario, which was expected because of the uncertainties and the prediction error of the NARMAX models.

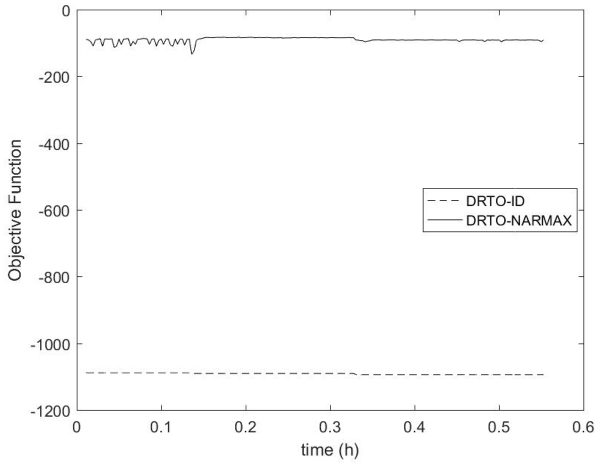


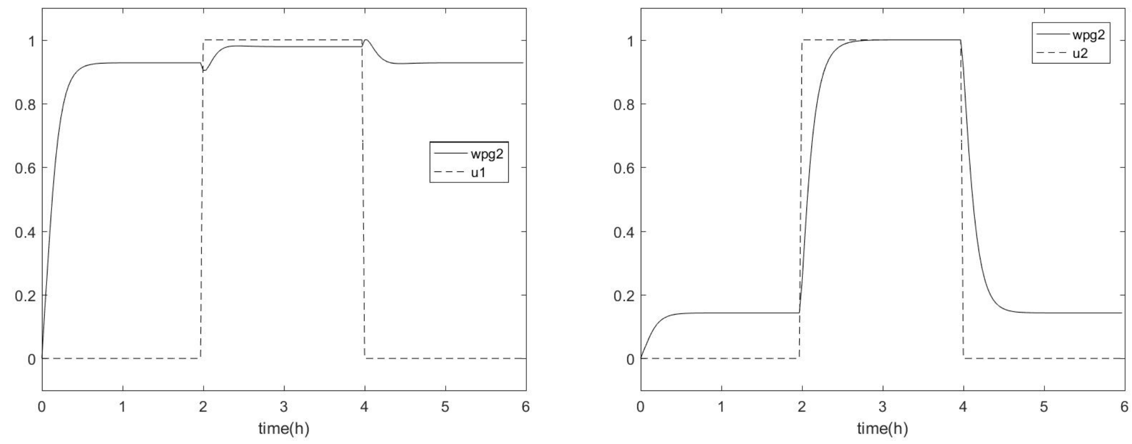
Figure 4.35: Comparison of objective function during DRTO.

4.3 Second Case Study

Oil production system with two gas-lift wells has twelve output variables, $p_{wh_i}, p_{bh_i}, w_{pg_i}, w_{po_i}, p_{rh}, p_m, w_{to}, w_{tg}$, and two input variables, w_{gl_i}, $i = 1, 2$.

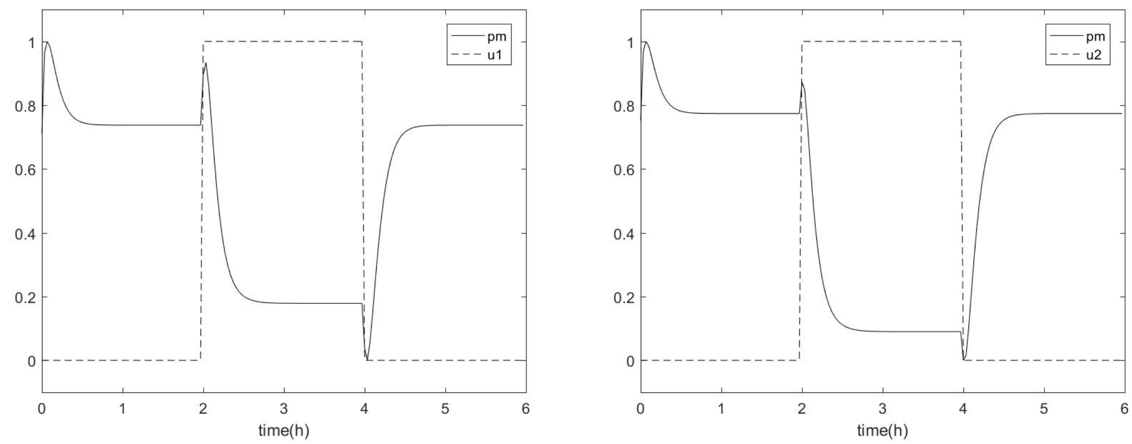
4.3.1 Gathering Information

The same tests were made for the second case study. On the first test, after $2h$, an unit pulse was applied on u_1 during $2h$ with a sample time period of $2min$. On the second test, the same pulse was applied on u_2. All variables are normalized. The slowest and the fastest variables are presented in Figures 4.36 and 4.37, respectively; the other results are in Appendix C. The system takes at maximum $20min$ to reach steady-state, so the PH is set to 50. Also, there are no delays regarding the input variables.



(a) (b)

Figure 4.36: Response of w_{pg_2} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



(a) (b)

Figure 4.37: Response of p_m due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.

4.3.2 Data Acquisition

The system was simulated based on first principle models on CasADi, where a series of 50 random steps with sample time period of $2min$ and duration of $0.028h$ were applied on the input variables (w_{gl_1} and w_{gl_2}) with operating intervals between 0 and 8 kg/s, as shown in Figures 4.38 and 4.39 after normalization and addition of white noise.

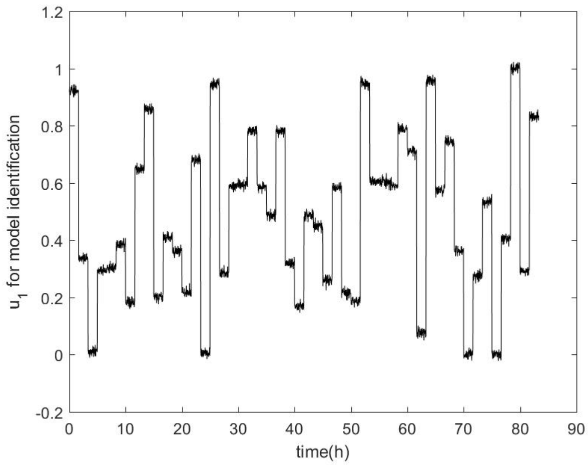


Figure 4.38: Simulation data of input variable u_1 for black-box identification - Second Case Study.

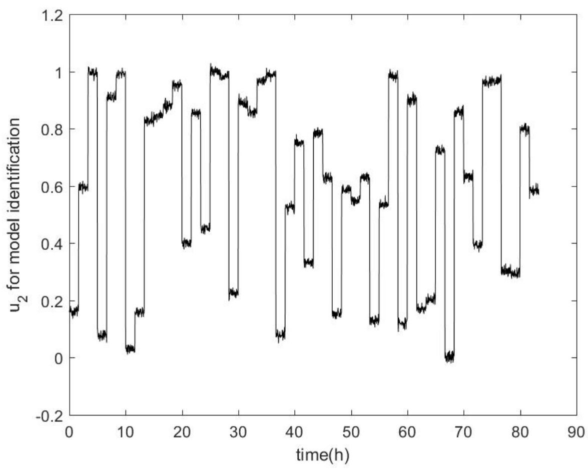


Figure 4.39: Simulation data of input variable u_2 for black-box identification - Second Case Study.

4.3.3 Black-box Identification

4.3.3.1 Parameter Estimation

The order parameters n_y, n_{ui} and n_e varied from 1 to 4 and the delays were set to zero as a result from the previous section. The optimal values are listed in Table 4.7.

Table 4.7: Optimal values of order parameters of black-box identification - Second Case Study.

Parameter	p_{wh_1}	p_{wh_2}	p_{bh_1}	p_{bh_2}	w_{pg_1}	w_{pg_2}	w_{po_1}	w_{po_2}	p_{rh}	p_m	w_{to}	w_{tg}
$n_{p_{NARX}}$	13	9	10	10	14	14	14	9	11	11	14	8
ℓ	2	2	2	2	1	2	2	1	1	2	1	1
n_y	2	1	2	2	3	4	4	4	3	3	4	2
n_{u_1}	1	1	4	3	3	1	4	4	3	4	2	3
n_{u_2}	2	2	3	4	2	4	2	3	4	4	4	4
n_e	1	3	1	1	3	1	2	2	2	2	1	0
$n_{p_{MA}}$	1	1	1	1	1	1	1	1	1	1	1	-

The model features are presented in Appendix D. The objective function values for the optimal solution of black-box identification are in Table 4.8.

Table 4.8: Objective function values for the optimal solution of black-box - Second Case Study.

Variable	J_{OLS}
p_{wh_1}	0.0617
p_{wh_2}	0.0965
p_{bh_1}	4.27×10^{-4}
p_{bh_2}	0.0331
w_{pg_1}	0.0607
w_{pg_2}	0.0111
w_{po_1}	0.0238
w_{po_2}	9.79×10^{-4}
p_{rh}	3.73×10^{-4}
p_m	0.0116
w_{to}	0.0596
w_{tg}	0.0697

4.3.3.2 Cross-validation

Figures 4.40 and 4.41 show the input data for validation, that was also generated with the same characteristics as the original set (same number of steps, same duration of each step of the data used on identification). It was used the same reference values for normalization.

Figures 4.42-4.53 show the k-step-ahead validation results ($k = 50$) and Table 4.9 shows the determination coefficients.

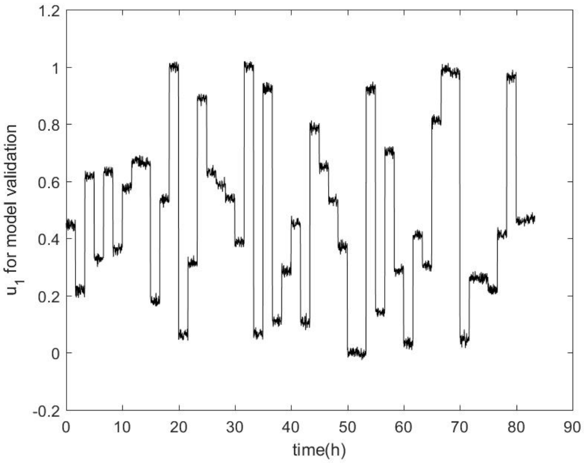


Figure 4.40: Input data u_1 for validation - Second Case Study.

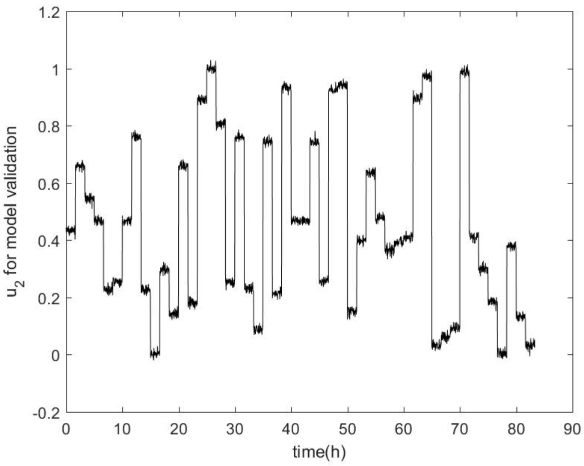
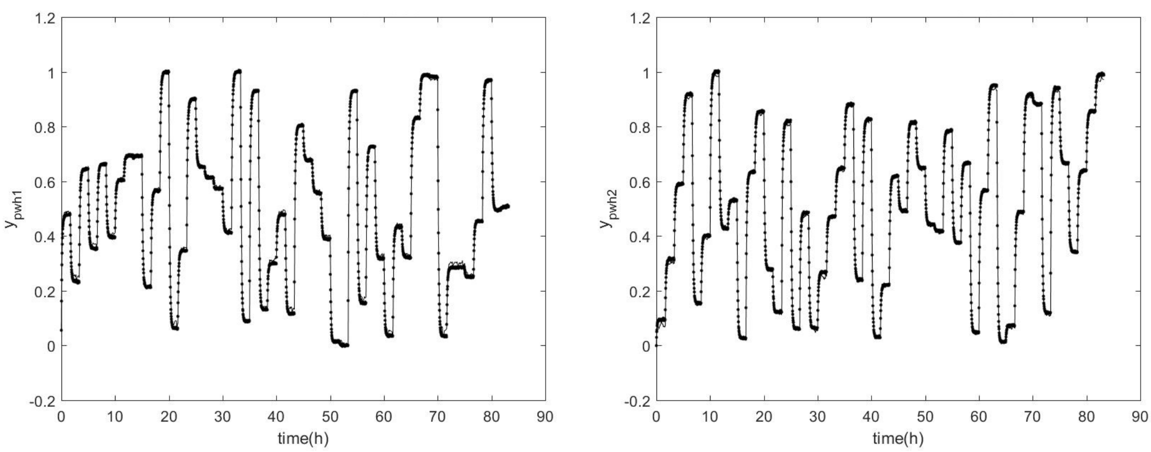
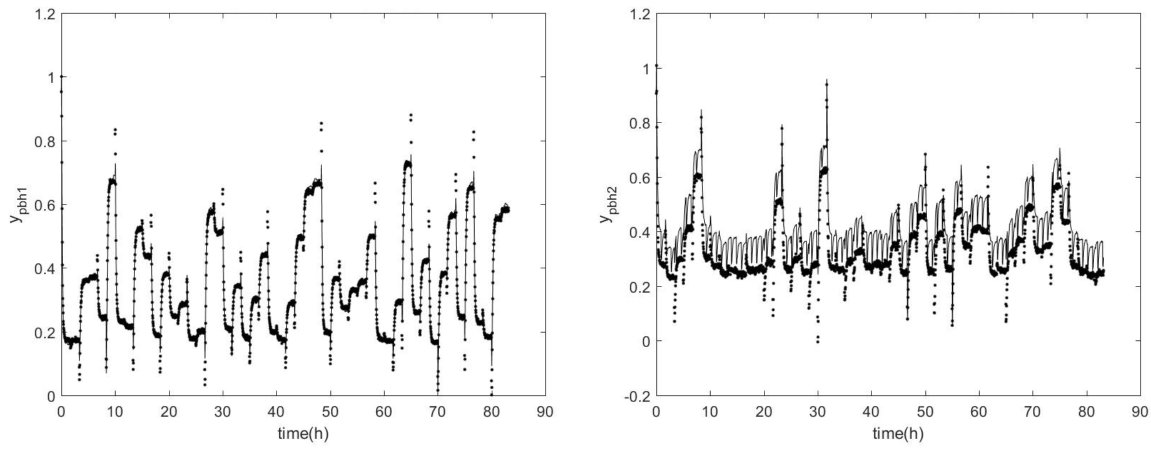


Figure 4.41: Input data u_2 for validation - Second Case Study.



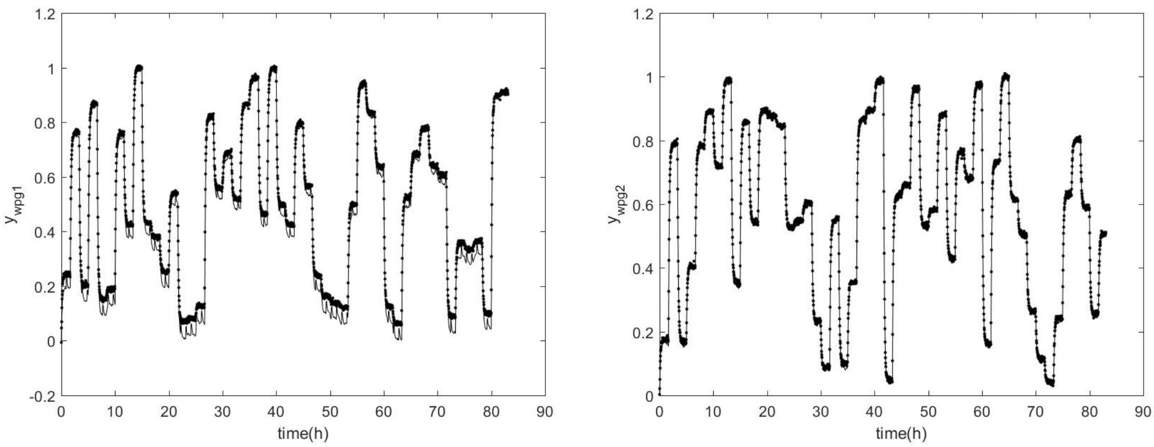
(a) (b)

Figure 4.42: Cross-validation of models for: (a) p_{wh_1}; (b) p_{wh_2}.



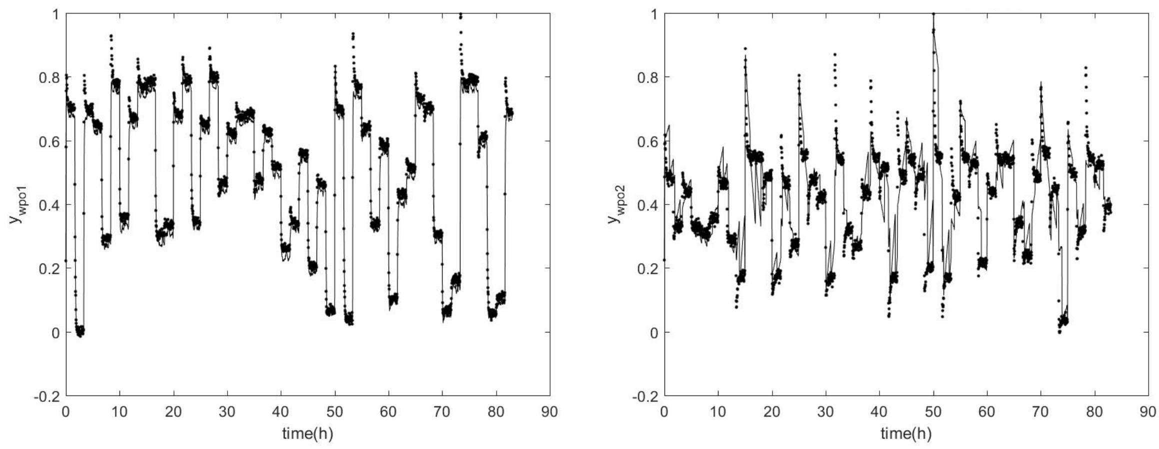
(a) (b)

Figure 4.43: Cross-validation of models for: (a) p_{bh_1}; (b) p_{bh_2}.



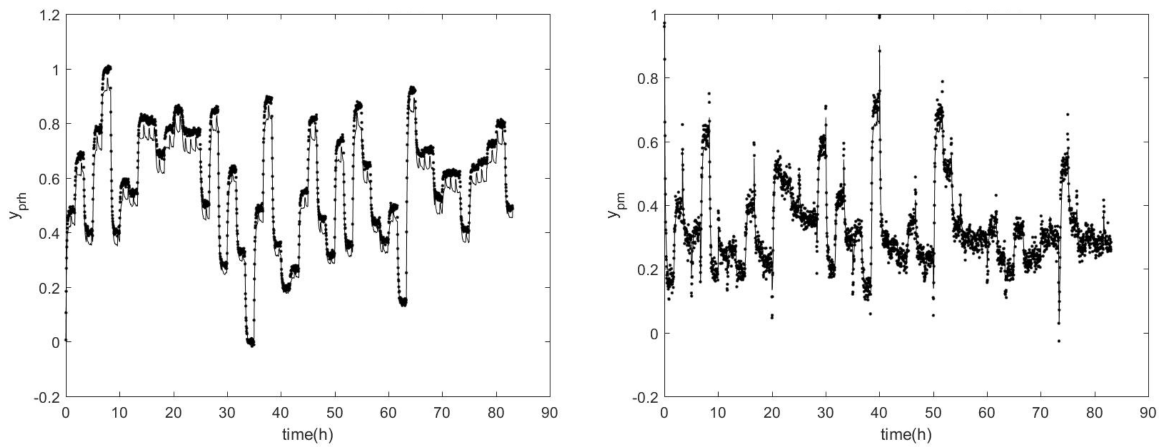
(a) (b)

Figure 4.44: Cross-validation of models for: (a) w_{pg_1}; (b) w_{pg_2}.



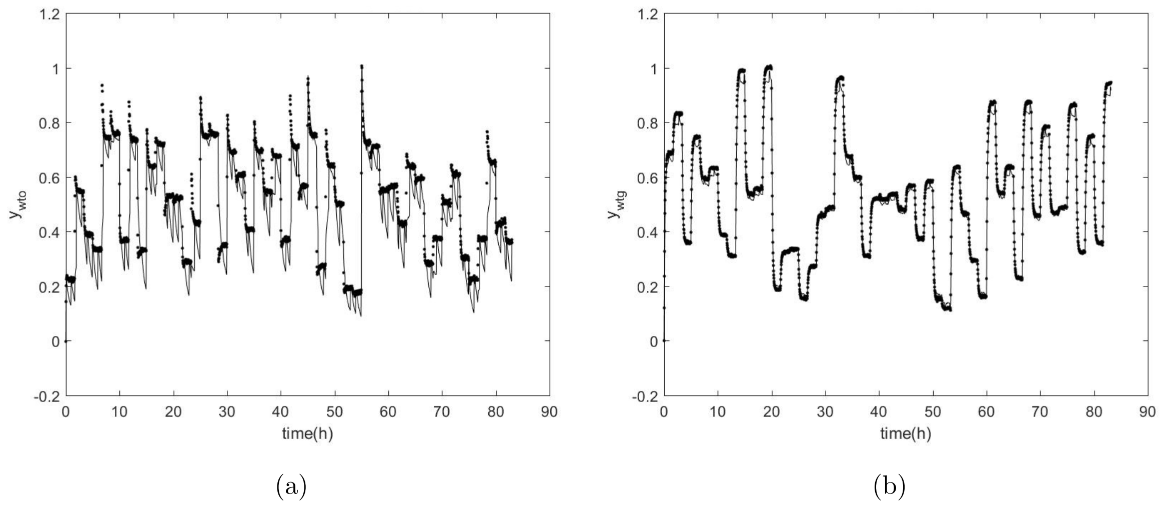
(a) (b)

Figure 4.45: Cross-validation of models for: (a) w_{po_1}; (b) w_{po_2}.



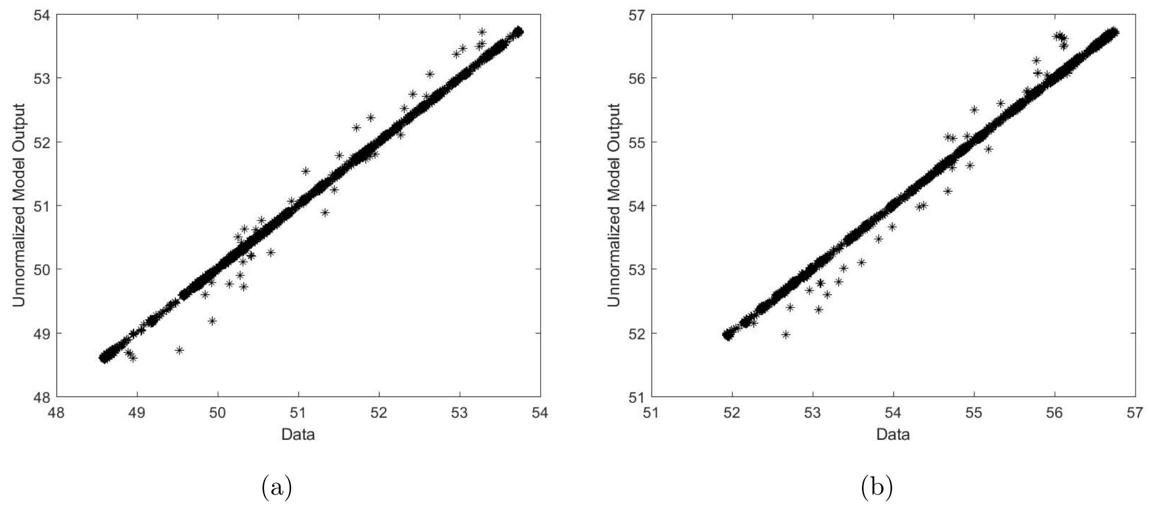
(a) (b)

Figure 4.46: Cross-validation of models for: (a) p_{rh}; (b) p_m.



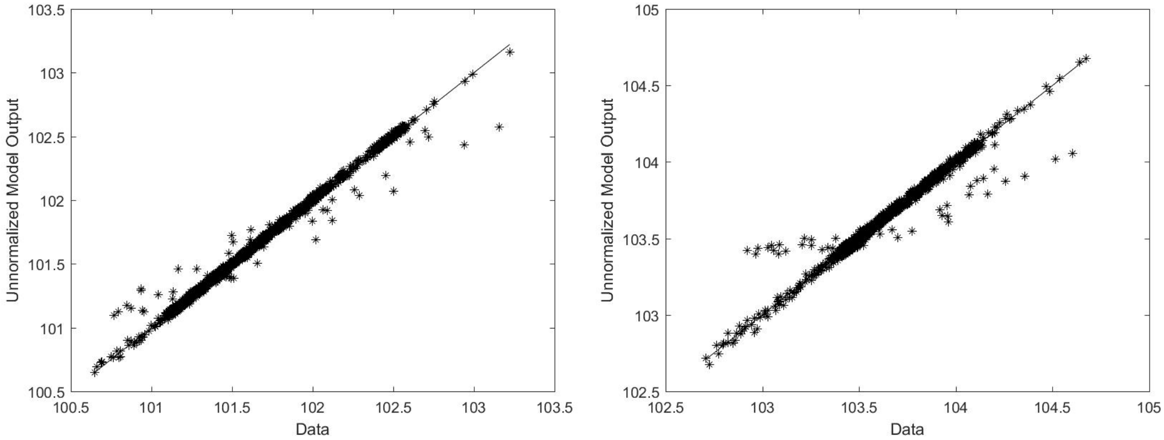
(a) (b)

Figure 4.47: Cross-validation of models for: (a) w_{to}; (b) w_{tg}.



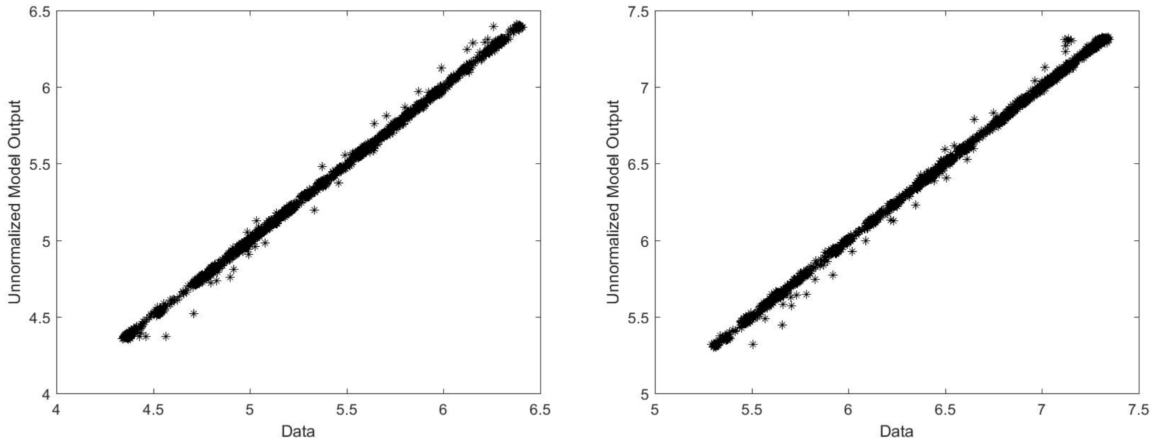
(a) (b)

Figure 4.48: Comparison of both unnormalized predicted output and data of variable: (a) p_{wh_1}; (b) p_{wh_2}.



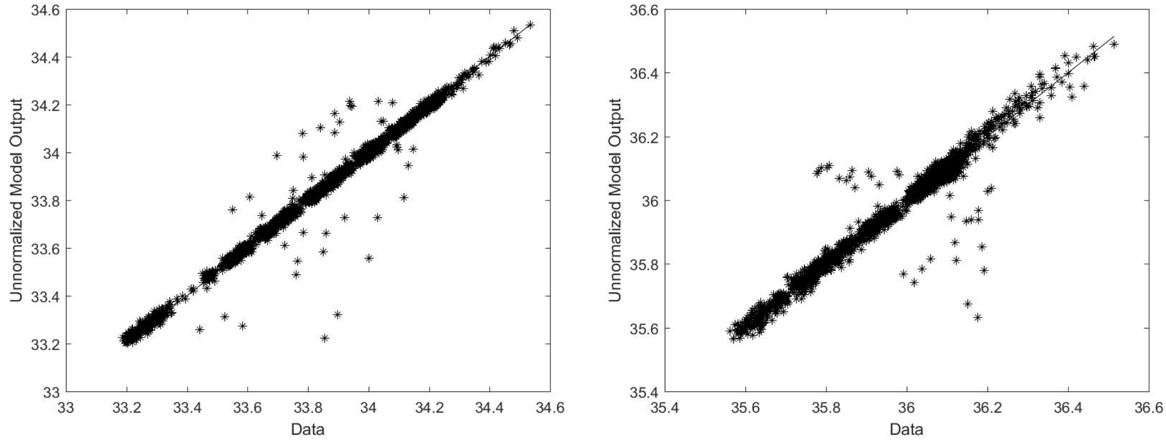
(a) (b)

Figure 4.49: Comparison of both unnormalized predicted output and data of variable: (a) p_{bh_1}; (b) p_{bh_2}.



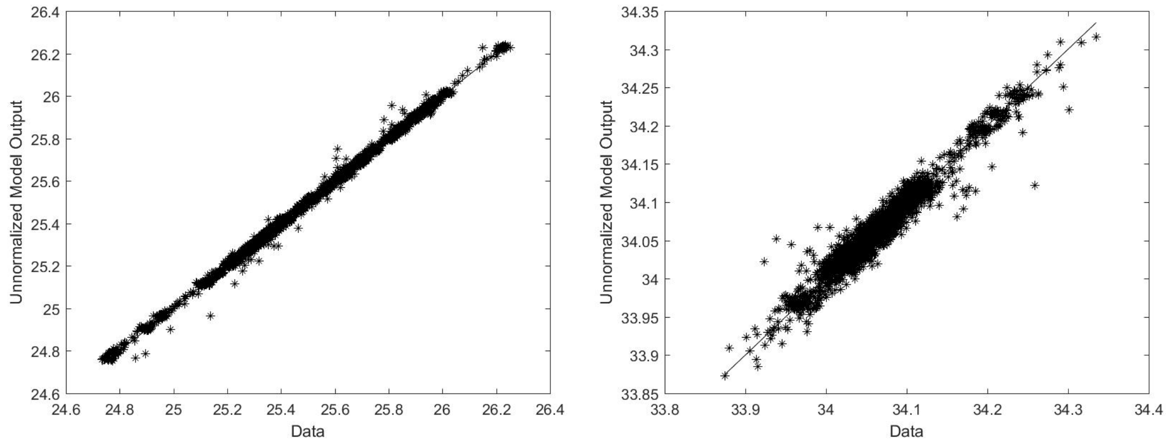
(a) (b)

Figure 4.50: Comparison of both unnormalized predicted output and data of variable: (a) w_{pg_1}; (b) w_{pg_2}.



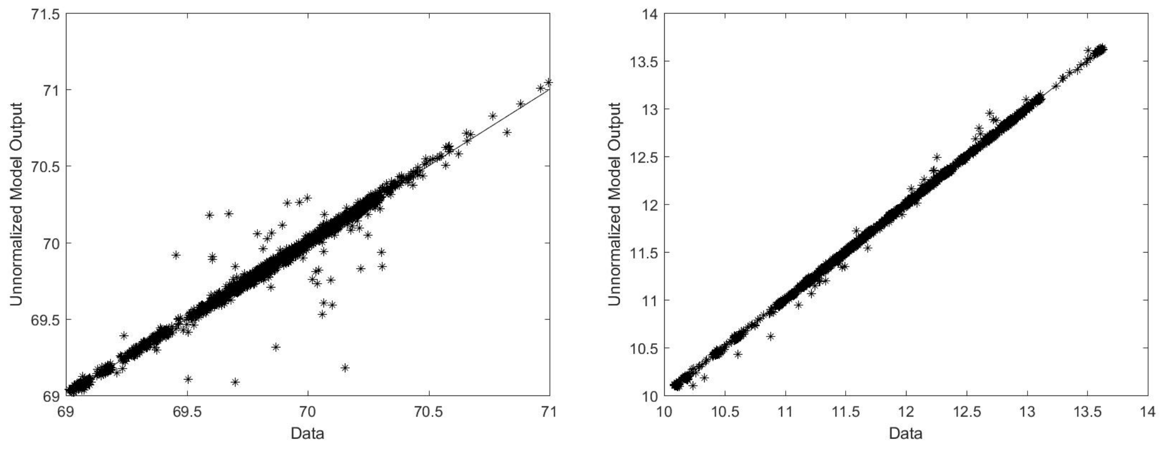
(a) (b)

Figure 4.51: Comparison of both unnormalized predicted output and data of variable: (a) w_{po_1}; (b) w_{po_2}.



(a) (b)

Figure 4.52: Comparison of both unnormalized predicted output and data of variable: (a) p_{rh}; (b) p_m.



(a)

(b)

Figure 4.53: Comparison of both unnormalized predicted output and data of variable: (a) w_{to}; (b) w_{tg}.

Table 4.9: Determination coefficient of validation for black-box identification - Second Case Study.

Variable	R^2_{NARMAX}
p_{wh_1}	0.9946
p_{wh_2}	0.9963
p_{bh_1}	0.9857
p_{bh_2}	0.5038
w_{pg_1}	0.9844
w_{pg_2}	0.9988
w_{po_1}	0.968
w_{po_2}	0.7331
p_{rh}	0.9332
p_m	0.9248
w_{to}	0.7978
w_{tg}	0.9887

From Figures 4.42-4.53, the variables that have the worst fit are, qualitatively, p_{bh_2}, w_{po_2}, p_m and w_{to}. It is corroborated by analyzing the worst R-squared values in Table 4.9. Beside the fact that p_m has one of the lowest R-squared values, it is higher than the minimum quality criterion, that is 0.9.

4.3.3.3 Dynamic Real-time Optimization

During DRTO run for the oil production system, all inputs were optimized with Equation 4.9 as economic function, and the constrains are given by Equation 4.10. The control horizon was set to 40, as can be noticed in Figures 4.54 and 4.55.

$$J_{DRTO} = \int_0^{40} (-(w_{to}^2) + 0.5 \times \sum_{i=1,2} w_{gl_i}^2)dt, \ i = 1, 2 \tag{4.9}$$

subject to

$$u_i, y_j \geqslant 0 \tag{4.10}$$

where $i = 1, 2$ and $j = p_{wh_i}, p_{bh_i}, w_{pg_i}, w_{po_i}, p_{rh}, p_m, w_{to}, w_{tg}$

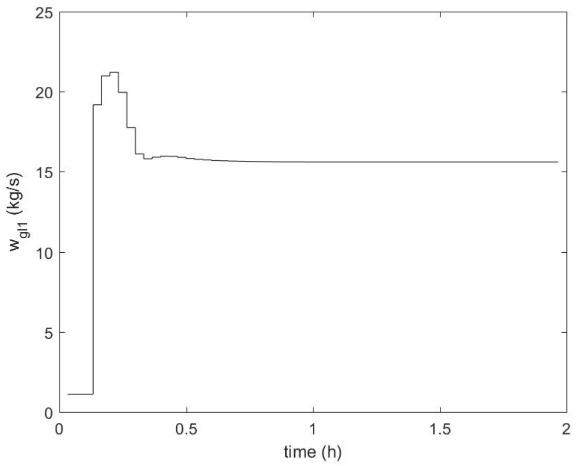


Figure 4.54: Control action on input w_{gl_1}.

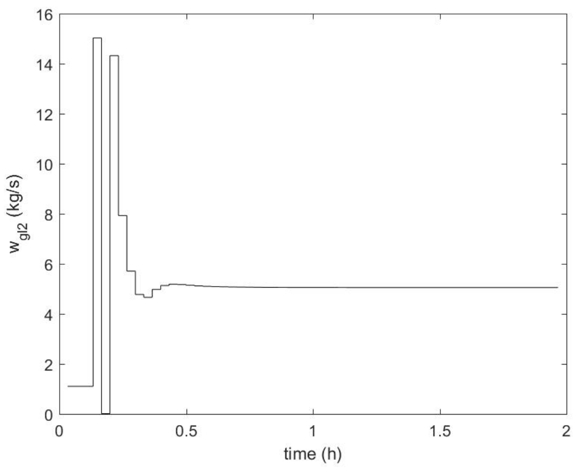


Figure 4.55: Control action on input w_{gl_2}.

The closed loop responses of the identified model of the system to the DRTO actions on the input variables (w_{gl_1} and w_{gl_2}) were compared with the one using first principle models, the results are shown in Figures 4.56-4.67.In Figures 4.58 and 4.65, the NARMAX model showed poor performance when trying to represent the nonlinearity of the respective variables. It can be due to the fact that one of the inputs changed so abruptly that made the trajectory to go to another stationary point.

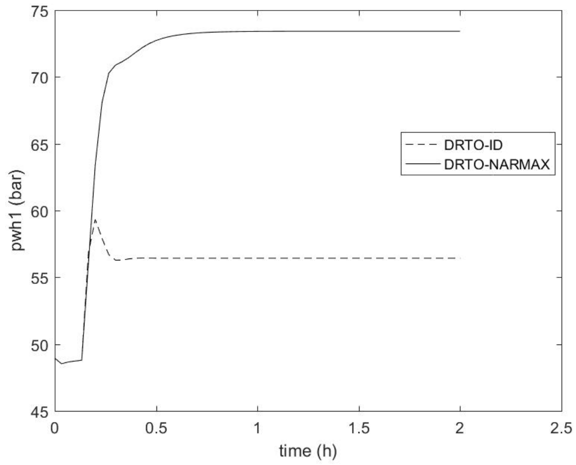


Figure 4.56: Comparison of DRTO performances for p_{wh_1}.

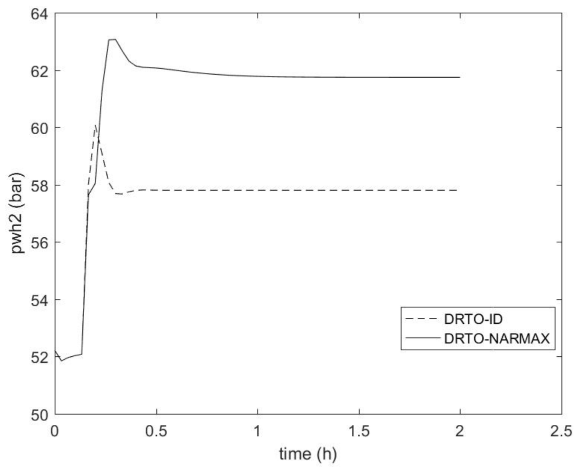


Figure 4.57: Comparison of DRTO performances for p_{wh_2}.

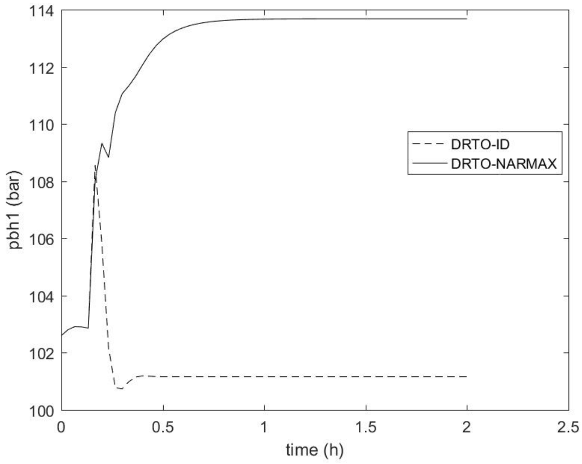


Figure 4.58: Comparison of DRTO performances for p_{bh_1}.

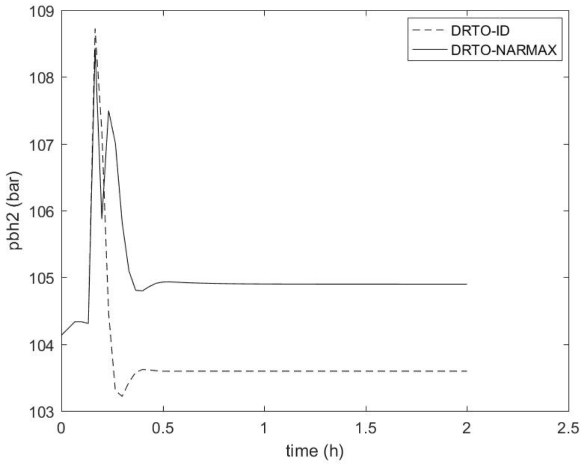


Figure 4.59: Comparison of DRTO performances for p_{bh_2}.

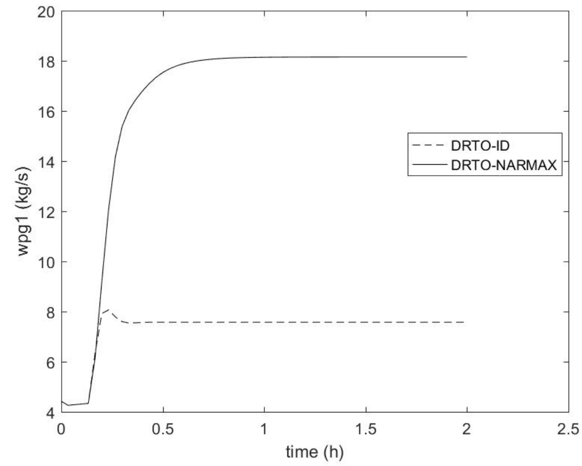


Figure 4.60: Comparison of DRTO performances for w_{pg_1}.

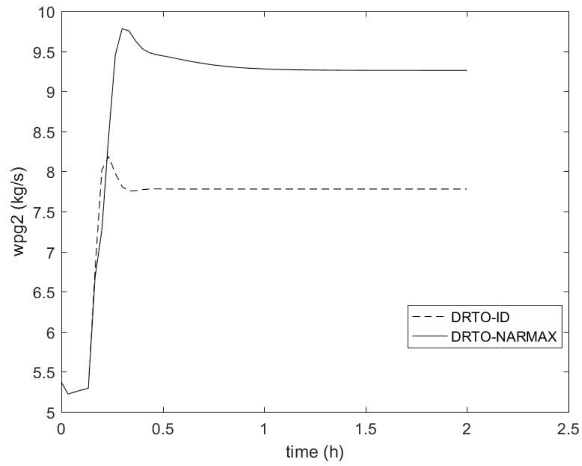


Figure 4.61: Comparison of DRTO performances for w_{pg_2}.

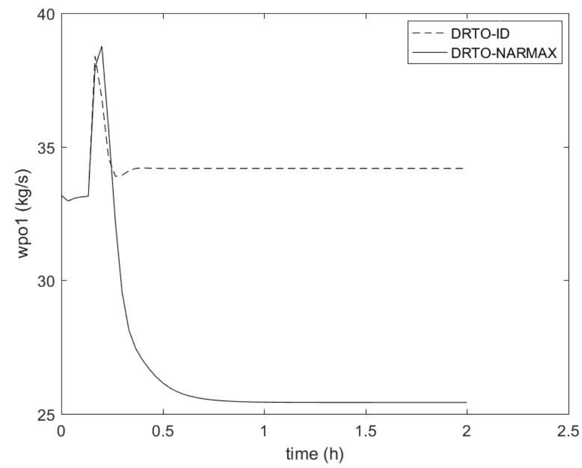


Figure 4.62: Comparison of DRTO performances for w_{po_1}.

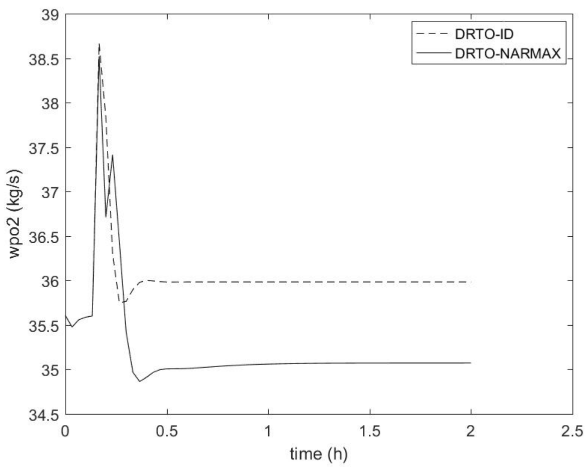


Figure 4.63: Comparison of DRTO performances for w_{po_2}.

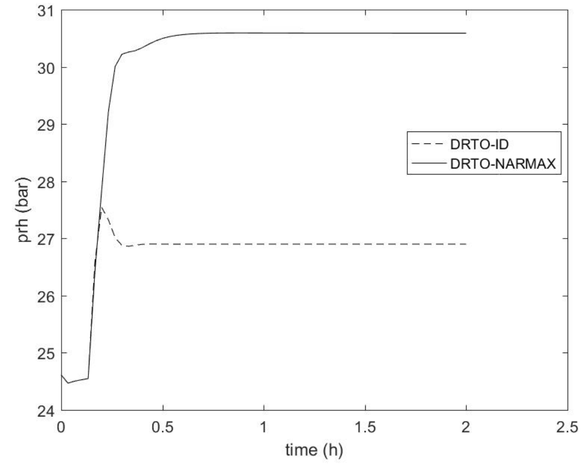


Figure 4.64: Comparison of DRTO performances for p_{rh}.

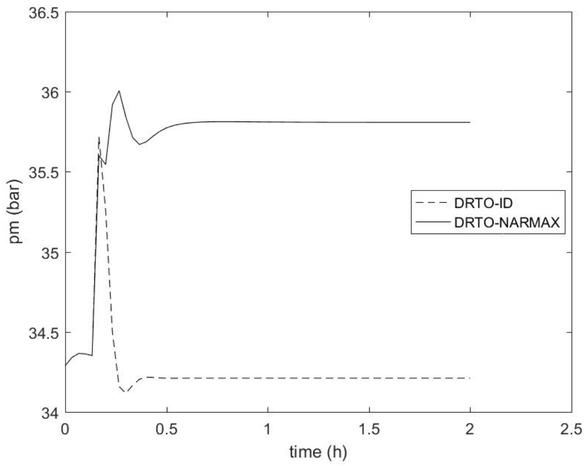


Figure 4.65: Comparison of DRTO performances for p_m.

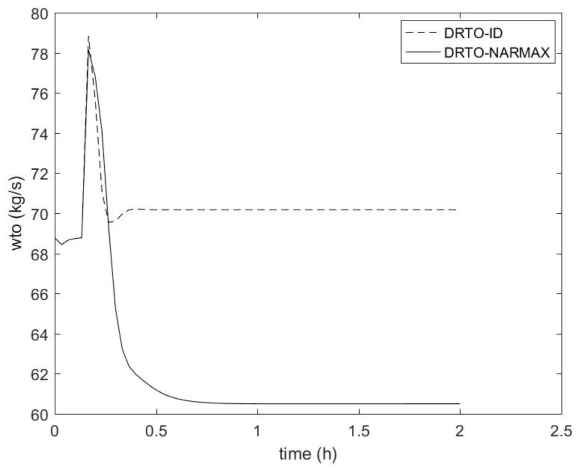


Figure 4.66: Comparison of DRTO performances for w_{to}.

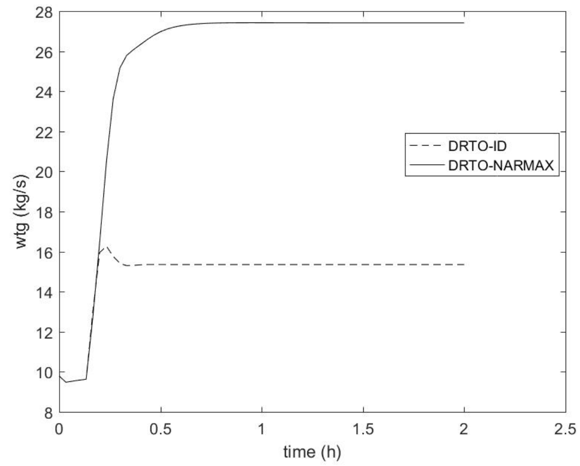


Figure 4.67: Comparison of DRTO performances for w_{tg}.

In Figure 4.68, it can be seen that the objective function of the DRTO-NARMAX has much higher values than the ones in the ideal scenario. It can be due to high uncertainties and the prediction error of the NARMAX models.

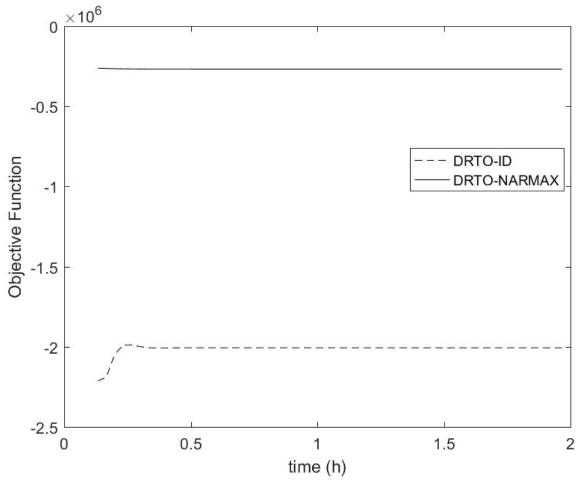


Figure 4.68: Comparison of objective function during DRTO.

4.3.4 Gray-box Identification

4.3.4.1 Parameter Estimation

As a result of the black-box identification, the minimum quality criterion made the gray-box identification algorithm to only find need for changing coordinates to identify p_{bh_2}, w_{po_2} and w_{to} models. The user suggestions are described in Appendix E, with each R-squared value and the chosen suggestions are in Table 4.10. The optimal values of the order parameters are listed in Table 4.11.

Table 4.10: Chosen modification on coordinates of gray-box identification - Second Case Study.

Output variable	First input	Second input
p_{bh_2}	u_1/u_2	u_2/u_1
w_{po_2}	u_2/u_1^2	$\sqrt{u_1}$
w_{to}	u_1^2/u_2	$\sqrt{u_2}$

Table 4.11: Optimal values of order parameters of gray-box identification - Second Case Study.

Parameter	p_{bh_2}	w_{po_2}	w_{to}
$n_{p_{NARX}}$	10	9	11
ℓ	1	1	1
n_y	3	4	4
n_{u_1}	4	2	4
n_{u_2}	4	4	4
n_e	0	0	0
$n_{p_{MA}}$	-	-	-

Table 4.12: Objective function values for the optimal solution of gray-box identification - Second Case Study.

Variable	J_{OLS}
p_{bh_2}	0.7964
w_{po_2}	2.8517
w_{to}	1.4141

The objective function values in Table 4.12 are much higher than the ones resulted from black-box identification. This is due to the fact that the algorithm chose a NARX model instead of a NARMAX one, so the noise was not modeled. The identified models are presented in Appendix F. They were simulated and the one-step-ahead prediction outputs were compared with the normalized data. These results are shown in Figures 4.69-4.71. It can be seen that the models describe the normalized data very well, despite not having the MA part in the model.

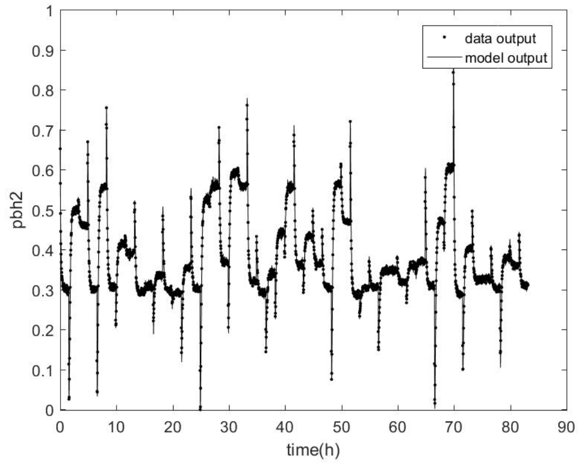


Figure 4.69: Simulation output of gray-box identification for variable p_{bh_2}.

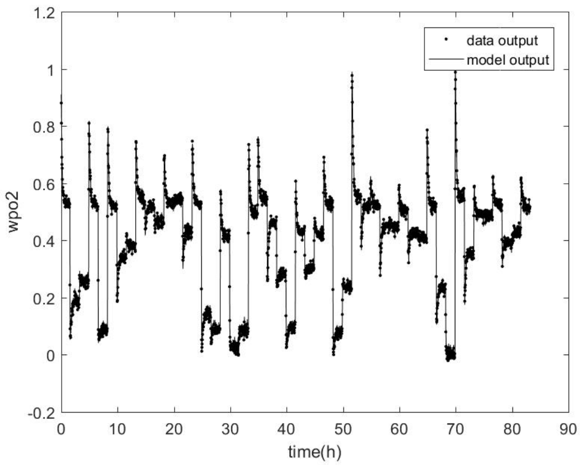


Figure 4.70: Simulation output of gray-box identification for variable w_{po_2}.

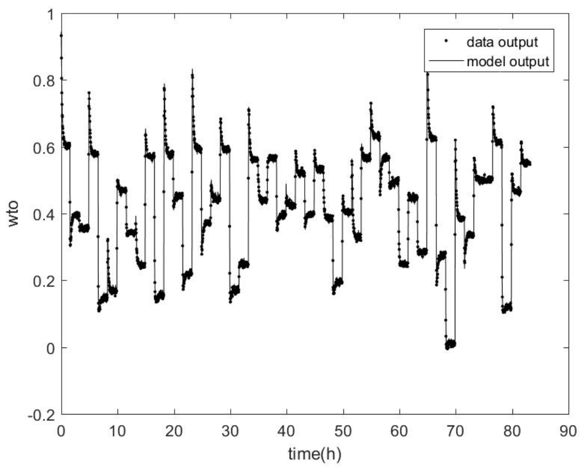


Figure 4.71: Simulation output of gray-box identification for variable w_{to}.

4.3.4.2 Cross-validation

Figures 4.72 and 4.73 show the input data for validation. It used the same reference values as the original data to be normalized, so it is possible that it is not exactly between 0 and 1. The the k-step-ahead validation results ($k = 50$) are presented in Figures 4.74-4.76 and the determination coefficients are shown in Table 4.13.

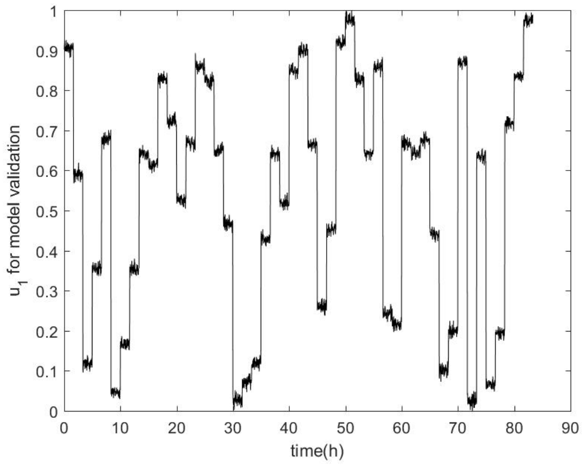


Figure 4.72: Input data u_1 for validation - Second Case Study.

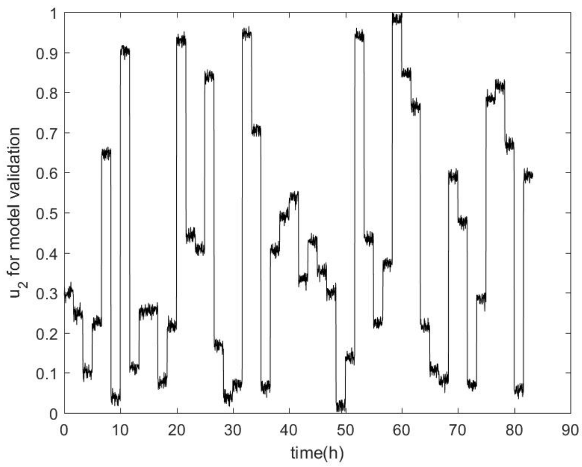
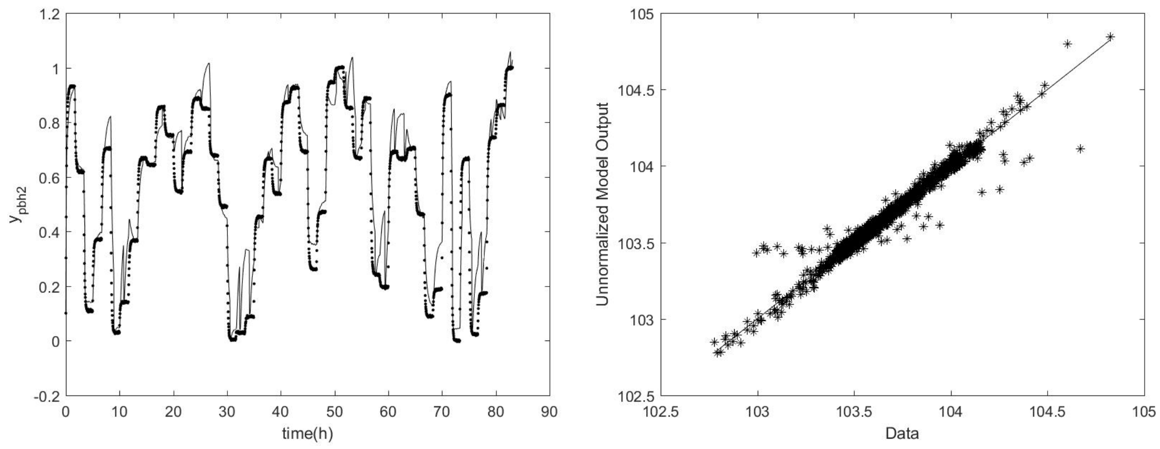
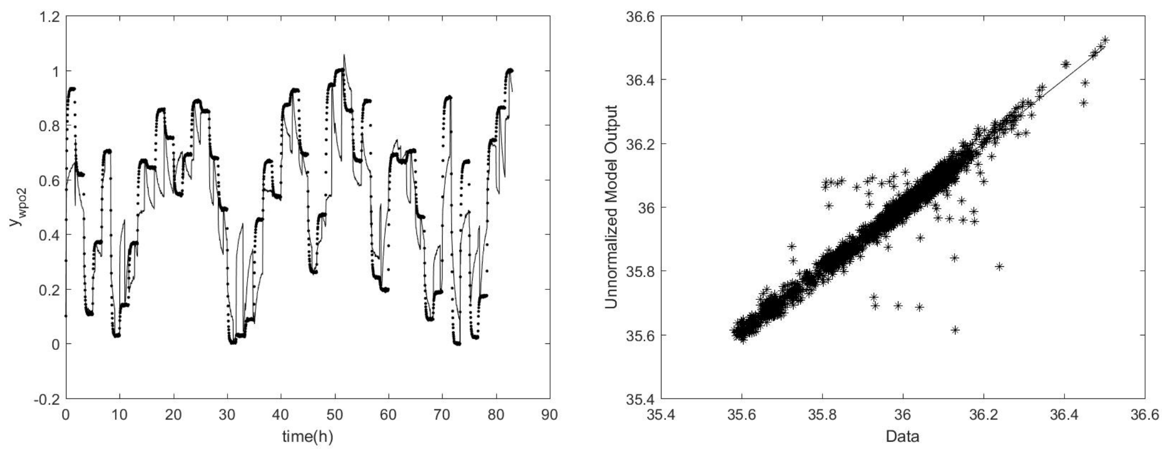


Figure 4.73: Input data u_2 for validation - Second Case Study.



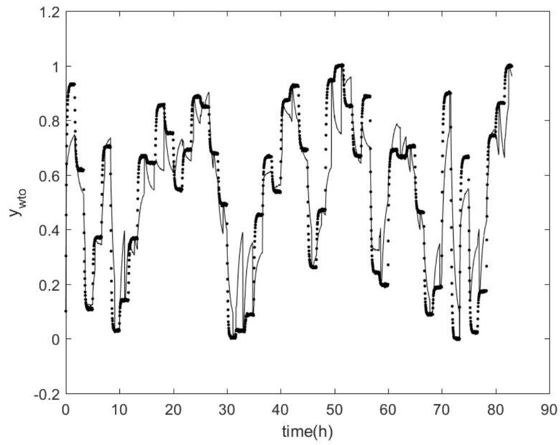
(a) (b)

Figure 4.74: (a) Cross-validation of p_{bh_2} model; (b) Comparison of unnormalized predicted output and unnormalized data.

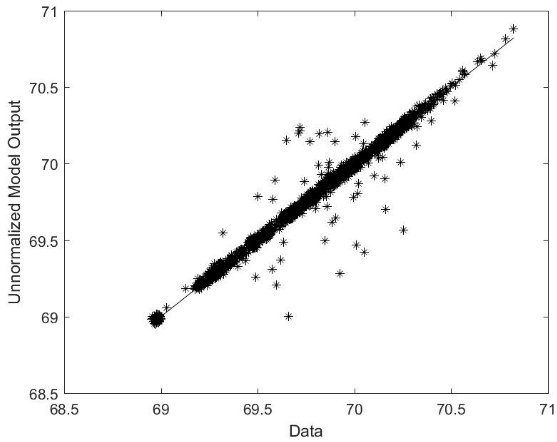


(a) (b)

Figure 4.75: (a) Cross-validation of w_{po_2} model; (b) Comparison of unnormalized predicted output and unnormalized data.



(a)



(b)

Figure 4.76: (a) Cross-validation of w_{to} model; (b) Comparison of unnormalized predicted output and unnormalized data.

Table 4.13: Determination coefficient of validation for gray-box identification - Second Case Study.

Variable	R^2_{NARX}
p_{bh_2}	0.8867
w_{po_2}	0.7442
w_{to}	0.8151

In Figures 4.74-4.76, it can be seen that model outputs for variables w_{po_2} and w_{to} are still quite disperse and do not give a good prediction, although the R-squared values, in Table 4.13 are higher, when compared with black-box identified model, in Table 4.9.

4.3.4.3 Dynamic Real-time Optimization

During DRTO run, all inputs were optimized, as can be observed in Figures 4.77 and 4.78.

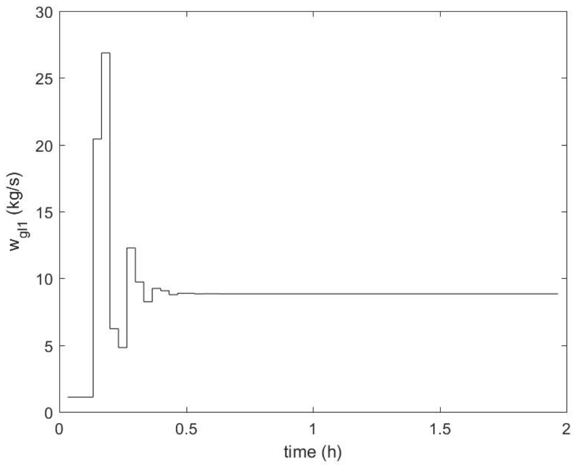


Figure 4.77: Control action on input w_{gl_1}.

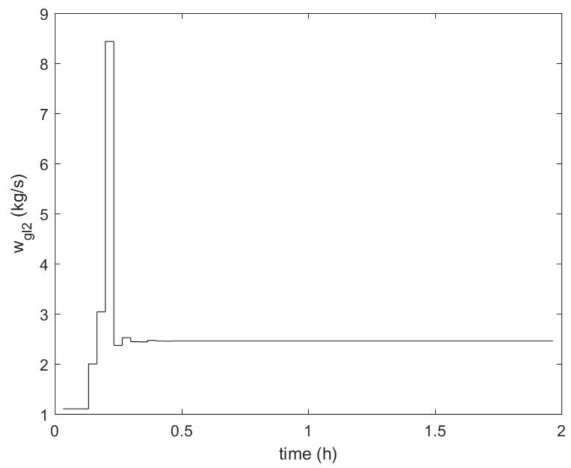


Figure 4.78: Control action on input w_{gl_2}.

The closed loop responses of the identified model of the system to the DRTO actions on the input variables (w_{gl_1} and w_{gl_2}) were compared with the one using first principle models, the results are shown in Figures 4.79-4.90. The performance improved because the DRTO-NARMAX solution based on the gray-box NARMAX models went closer to the ideal solution when comparing to the one based on the NARMAX model from the black-box identification.

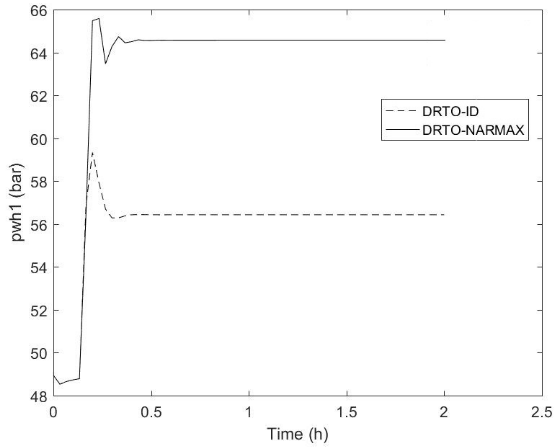


Figure 4.79: Comparison of DRTO performances for p_{wh_1}.

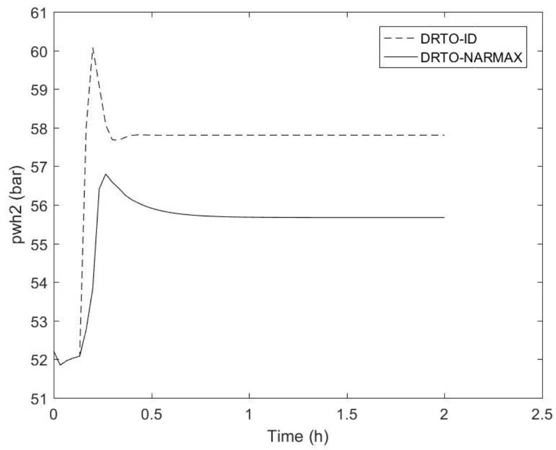


Figure 4.80: Comparison of DRTO performances for p_{wh_2}.

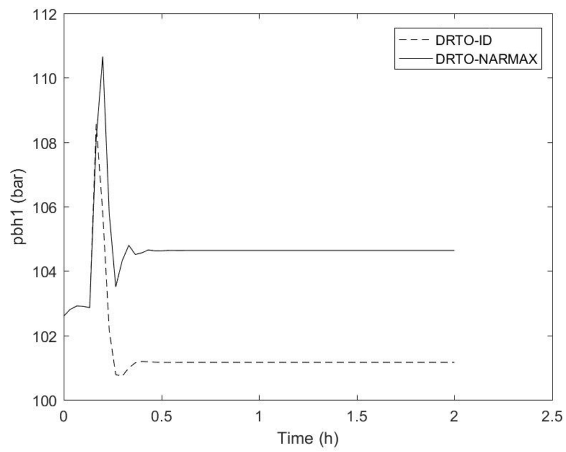


Figure 4.81: Comparison of DRTO performances for p_{bh_1}.

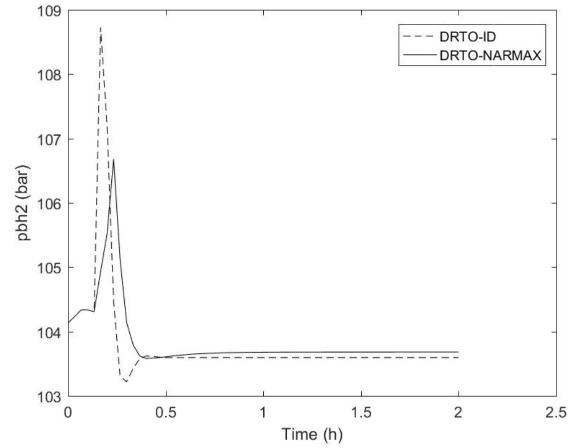


Figure 4.82: Comparison of DRTO performances for p_{bh_2}.

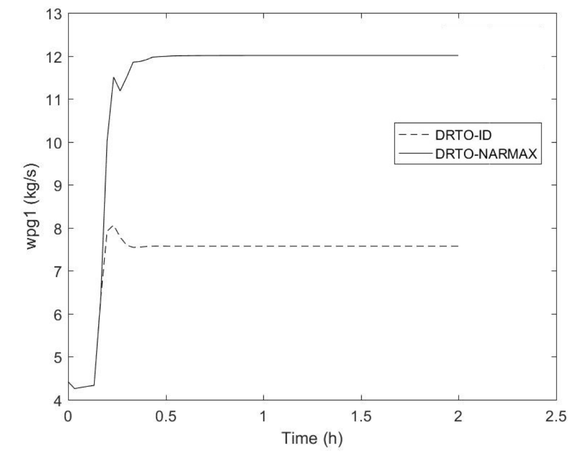


Figure 4.83: Comparison of DRTO performances for w_{pg_1}.

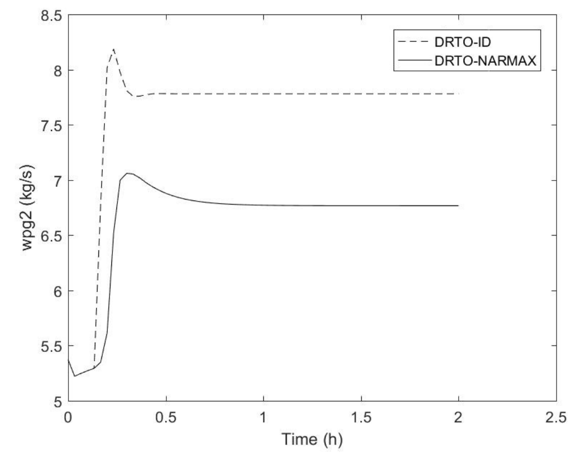


Figure 4.84: Comparison of DRTO performances for w_{pg2}.

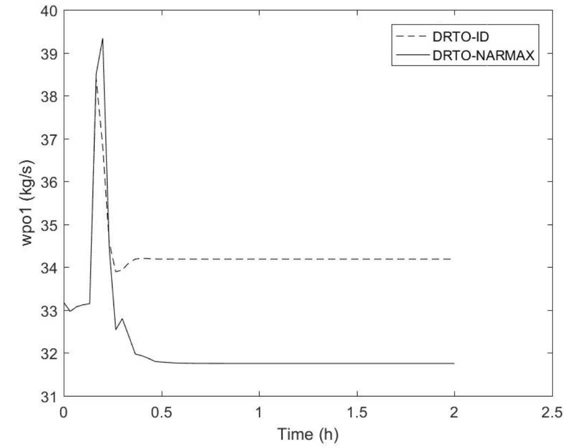


Figure 4.85: Comparison of DRTO performances for w_{po_1}.

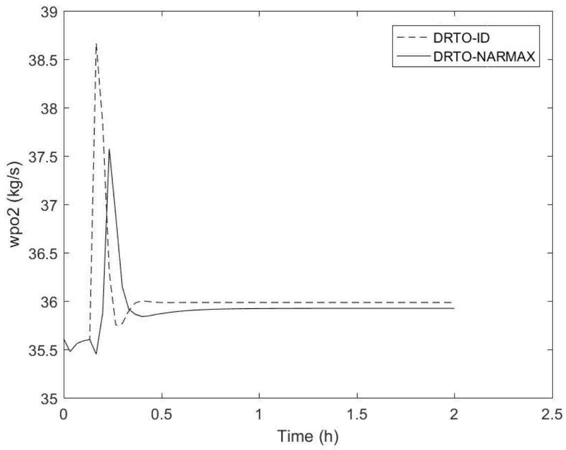


Figure 4.86: Comparison of DRTO performances for w_{po_2}.

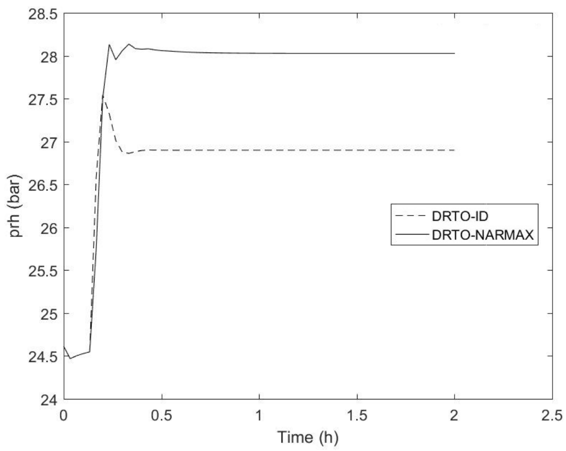


Figure 4.87: Comparison of DRTO performances for p_{rh}.

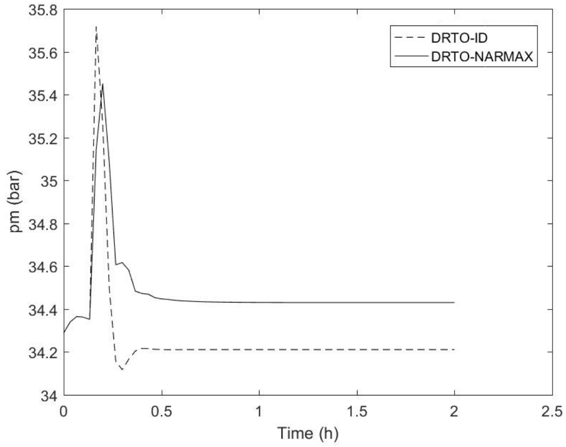


Figure 4.88: Comparison of DRTO performances for p_m.

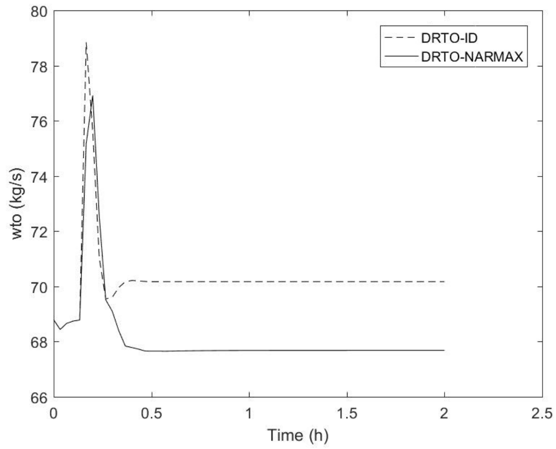


Figure 4.89: Comparison of DRTO performances for w_{to}.

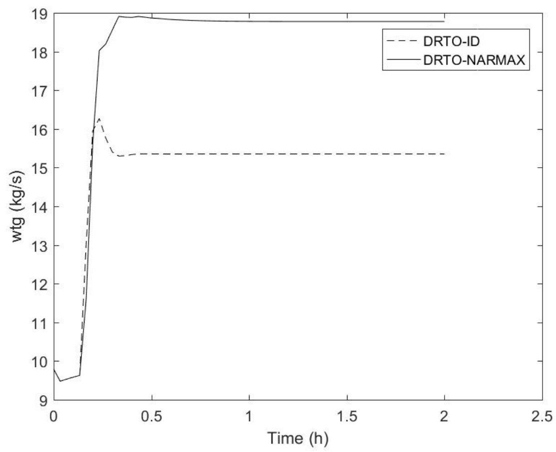


Figure 4.90: Comparison of DRTO performances for w_{tg}.

Comparing the performance of gray-box NARMAX model-based DRTO with the black-box NARMAX model-based DRTO, it can be observed that the difference between the solutions for the variables p_{bh_2}, w_{po_2} and w_{to} on DRTO-ID and DRTO-NARMAX were reduced a lot.

In Figure 4.91, it can be seen that the objective function of the DRTO-NARMAX has much higher values than the ones in the ideal scenario. This is due to imperfection of the model, generating high prediction error.

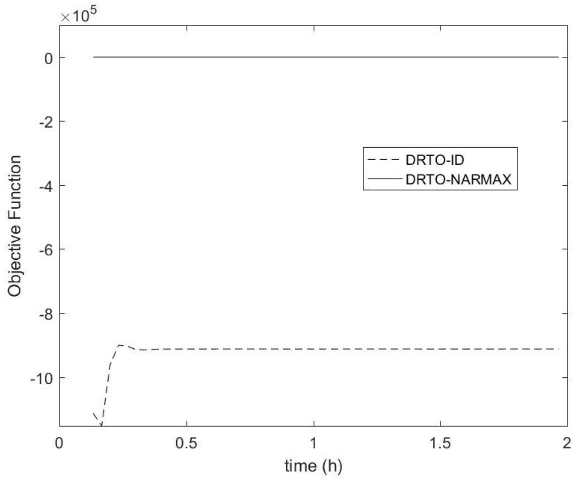


Figure 4.91: Comparison of objective function during DRTO.

Chapter 5

Conclusions and Suggestions

A study of the Van de Vusse reactor and the oil production system with two gas-lift wells was made by disturbing each input variable at a time, the required duration of each step was found and the prediction horizon was calculated.

The identification using NARMAX structure was performed for both cases and the user interaction gave improvement to the model. On the other hand, it makes the identification a little hard to improve when there is lack of knowledge, that is when the study of the process is more important and should be made attentive and exhaustively.

Despite the fact that the search for the optimal is local and exhaustive, the gray-box identification algorithm had a great performance, because it is based on analytical solution. The usage of prior knowledge by changing the coordinates avoids gray-box usual complexity, because it does not change the fact that the model is still linear on the parameters. For both the case studies, the gray-box algorithm showed low complexity when leading with a combination of the parameters that should be given by the user, and it did not take a lot of computational effort to find the solution, although it took longer than the black-box one. The results of the gray-box identification were better than the black-box one for both case studies. Therefore, gray-box identification gives more accurate, sometimes smaller and less complex model than the black-box one.

The application on optimization and control made it clearer that the gray-box identification improves the modeling of the system. The gray-box identified model provided a better performance than the black-box one, even though for Van de Vusse CSTR, the optimization went unstable. This can be due to the fact that the control horizon of the dynamic real-time optimization (DRTO) is set to be the same as the prediction horizon and it can lead the system to oscillate. Moreover, it demanded acquisition of knowledge regarding to the optimization tool CasADi, even though there is still lack of it when the results are oscillating.

Regarding the algorithm itself, the orthogonal least square method avoids the ill-conditioning problem, but it needs a lot of optimization layers in order to find optimal orders and nonlinearity degree for the model. One suggestion for continuing this work is to compile a multi-objective optimization algorithm for the batch estimation. Other suggestions are a hybrid estimation (using moving horizon estimation) that could be implemented with the dynamic real-time optimization, and improve the DRTO algorithm to avoid instability.

Bibliography

ABDULLAH, S. M., YASSIN, A. I. M., TAHIR, N. M., 2015, "Particle swarm optimization and least squares estimaton of NARMAX", *ARPN Journal of Engineering and Applied Sciences*, v. 10, n. 22, pp. 17139–17145. ISSN: 18196608.

AGUIRRE, L. A., 2000, *Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais.* Belo Horizonte, UFMG. ISBN: 85-7041-220-7.

AGUIRRE, L. A., BILLINGS, S. A., 1995, "Improved structure selection for nonlinear models based on term clustering", *International Journal of Control*, v. 62, n. 3, pp. 569–587. ISSN: 13665820. doi: 10.1080/00207179508921557.

AGUIRRE, L. A., RODRIGUES, G. G., JACOME, C. R. F., 1998, "Identificação de sistemas nao-lineares utilizando modelos NARMAX polinomiais - uma revisão e novos resultados", *Controle e Automação*, v. 9, n. 2, pp. 90–106. ISSN: 01031759.

AKAIKE, H., 1974, "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, v. 19, n. 6, pp. 716–723. ISSN: 15582523. doi: 10.1109/TAC.1974.1100705.

ALSTAD, V., 2005, *Studies on Selection of Controlled Variables.* Tese de Doutorado, NTNU.

ANDERSSON, J., 2013, *A General-Purpose Software Framework for Dynamic Optimization.* Tese de Doutorado, Arenberg Doctoral School, KU Leuven, Heverlee.

BARBOSA, B. H., AGUIRRE, L. A., MARTINEZ, C. B., et al., 2011, "Black and gray-box identification of a hydraulic pumping system", *IEEE Transactions on Control Systems Technology*, v. 19, n. 2, pp. 398–406. ISSN: 10636536. doi: 10.1109/TCST.2010.2042600.

BIJANZADEH, M., KAHANI, D., KOHAN, E. D., et al., 2013, "NARMAX-OLS Representation of a semi-active dynamic leg joint model for a paraplegic subject using functional electrical stimulation". In: *World Congress on Engineering*, v. 2, pp. 1303–1308, London, UK. ISBN: 9789881925282.

BILLINGS, S. A., 2013, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. John Wiley & Sons. ISBN: 9781118535554.

BILLINGS, S. A., CHEN, S., 1989, "Identification of non-linear rational systems using a prediction-error estimation algorithm", *International Journal of Systems Science*, v. 20, n. 3, pp. 467–494. ISSN: 14645319. doi: 10.1080/00207728908910143.

BILLINGS, S. A., CHEN, S., 1998, "The determination of multivariable non-linear models for dynamic systems using neural networks", *Neural Network Systems Techniques and Applications*, pp. 231–278.

BILLINGS, S. A., FADZIL, M. B., 1985, "The Practical Identification of Systems with Nonlinearities", *Proceedings of the 7th IFAC Symposium ldententification and System Parameter Estimation*, v. 18, n. 5, pp. 155–160. ISSN: 14746670. doi: 10.1016/S1474-6670(17)60551-2.

BILLINGS, S. A., LEONTARITIS, I. J., 1981, "Identification of nonlinear systems using parameter estimation techniques". In: *Proceedings of the IEE Conference Control and Its Applications*, n. 194, pp. 183–187, Warwick.

BILLINGS, S. A., LEONTARITIS, I. J., 1982, "Parameter estimation techniques for nonlinear systems". In: *Proceedings of the 6th IFAC Symposium on Identification and System Parameter Estimation*, v. 15, p. 427, Washinton, D.C. doi: 10.1016/S1474-6670(17)63039-8.

BILLINGS, S. A., VOON, W. S. F., 1983, "Structure detection and model validity tests in the identification of nonlinear systems". In: *Proceedings of the IEE Control Theory and Applications, Pt. D*, v. 130, pp. 193–199. ISBN: 0143-7054. doi: 10.1049/ip-d.1983.0034.

BILLINGS, S. A., VOON, W. S. F., 1984, "Least squares parameter estimation algorithms for non-linear systems", *International Journal of Systems Science*, v. 15, n. 6, pp. 601–615. ISSN: 14645319. doi: 10.1080/00207728408547198.

BILLINGS, S. A., VOON, W. S. F., 1986, “A prediction-error and stepwise-regression estimation algorithm for non-linear systems”, *International Journal of Control*, v. 44, n. 3, pp. 803–822. ISSN: 0020-7179. doi: 10.1080/00207178608933633.

BILLINGS, S. A., KORENBERG, M. J., CHEN, S., 1988, “Identification of non-linear output-affine systems using an orthogonal least-squares algorithm”, *International Journal of Systems Science*, v. 19, n. 8, pp. 1559–1568. ISSN: 14645319. doi: 10.1080/00207728808964057.

CHEN, S., BILLINGS, S. A., 1989, “Representations of non-linear systems: The narmax model”, *International Journal of Control*, v. 49, n. 3, pp. 1013–1032. ISSN: 13665820. doi: 10.1080/00207178908559683.

CHEN, S., BILLINGS, S. A., LUO, W., 1989, “Orthogonal least squares methods and their application to non-linear system identification”, *International Journal of Control*, v. 50, n. 5, pp. 1873–1896. ISSN: 0020-7179. doi: 10.1080/00207178908953472.

CORRÊA, M. V., AGUIRRE, L. A., 2004, “Identificação não-linear caixa-cinza: uma revisão e novos resultados”, *Controle & Automação*, v. 15, n. 2, pp. 109–126. ISSN: 0103-1759. doi: 10.1590/S0103-17592004000200001.

DRAPER, N. R., SMITH, H., 1998, *Applied Regression Analysis*. 3rd ed. New York, John Wiley & Sons. ISBN: 9780471170822.

JÁCOME, C. R. F., 1996, *Uso de Conhecimento Prévio na Identificação de Modelos Polinomiais NARMAX*. Tese de Doutorado, Universidade Federal de Minas Gerais, Belo Horizonte.

JAMALUDIN, M. Z., SWARTZ, C. L. E., 2016, “Closed-loop Formulation for Nonlinear Dynamic Real-time Optimization”, *IFAC-PapersOnLine*, v. 49, n. 7, pp. 406–411. ISSN: 24058963. doi: 10.1016/j.ifacol.2016.07.376.

JOHANSEN, T. A., 1996, “Identification of non-linear systems using empirical data and prior knowledge—an optimization approach”, *Automatica*, v. 32, n. 3, pp. 337–356. ISSN: 0005-1098. doi: 10.1016/0005-1098(95)00146-8.

JØRGENSEN, J. B., 2004, *Moving Horizon Estimation and Control*. Tese de Doutorado, Technical University of Denmark.

KARPLUS, W. J., 1977, “The spectrum of mathematical modeling and systems simulation”, *ACM SIGSIM Simulation Digest*, v. 9, n. 1, pp. 32–38. ISSN: 01636103. doi: 10.1145/1102505.1102522.

KORENBERG, M., BILLINGS, S. A., LIU, Y. P., et al., 1988, "Orthogonal parameter estimation algorithm for non-linear stochastic systems", *International Journal of Control*, v. 48, n. 1, pp. 193–210. ISSN: 13665820. doi: 10.1080/00207178808906169.

KRISHNAMOORTHY, D., FOSS, B., SKOGESTAD, S., 2018, "Steady-state real-time optimization using transient measurements", *Computers and Chemical Engineering*, v. 115, pp. 34–45. ISSN: 00981354. doi: 10.1016/j.compchemeng.2018.03.021.

KUKREJA, S. L., GALIANA, H. L., KEARNEY, R. E., 2004, "A bootstrap method for structure detection of NARMAX models", *International Journal of Control*, v. 77, n. 2, pp. 132–143. ISSN: 00207179. doi: 10.1080/00207170310001646264.

LEONTARITIS, I. J., BILLINGS, S. A., 1985, "Input-output parametric models for non-linear systems Part I: Deterministic non-linear systems", *International Journal of Control*, v. 41, n. 2, pp. 303–328. ISSN: 13665820. doi: 10.1080/0020718508961129.

LEONTARITIS, I. J., BILLINGS, S. A., 1987a, "Experimental design and identifiability for non-linear systems", *International Journal of Systems Science*, v. 18, n. 1, pp. 189–202. ISSN: 14645319. doi: 10.1080/00207728708963958.

LEONTARITIS, I. J., BILLINGS, S. A., 1987b, "Model selection and validation methods for non-linear systems", *International Journal of Control*, v. 45, n. 1, pp. 311–341. ISSN: 13665820. doi: 10.1080/00207178708933730.

LJUNG, L., 1999, *System Identification: Theory for the user.* 2ª ed. New Jersey, Prentice Hall. ISBN: 0136566952.

MARIUS, O., NICOLAE, P., 2015, "Identification method based on NARMAX polynomials". In: *19th International Conference on System Theory, Control and Computing*, n. 2, pp. 907–911, Cheile Gradistei, Romania, oct. ISBN: 9781479984817. doi: 10.1109/ICSTCC.2015.7321410.

SCHMITZ, M. J., GREEN, R. A., 2012, "Multisine excitation design to increase the efficiency of system identification analysis through undersampling and DFT optimization", *Measurement*, v. 45, n. 6, pp. 1576–1586. ISSN: 0263-2241. doi: 10.1016/j.measurement.2012.02.019.

TEIXEIRA, B. O., AGUIRRE, L. A., 2011, "Using uncertain prior knowledge to improve identified nonlinear dynamic models", *Journal of Process Control*, v. 21, n. 1, pp. 82–91. ISSN: 09591524. doi: 10.1016/j.jprocont.2010.10.008.

THOMSON, M., SCHOOLING, S. P., SOUFIAN, M., 1996, "The practical application of a nonlinear identification methodology", *Control Engineering Practice*, v. 4, n. 3, pp. 295–306. ISSN: 09670661. doi: 10.1016/0967-0661(96)00006-8.

TRIERWEILER, J. O., 1997, *A Systematic Approach to Control Structure Design*. Tese de Doutorado, Universidade de Dortmund, Dortmund.

WÜRTH, L., HANNEMANN, R., MARQUARDT, W., 2011, "A two-layer architecture for economically optimal process control and operation", *Journal of Process Control*, v. 21, n. 3, pp. 311–321. ISSN: 09591524. doi: 10.1016/j.jprocont.2010.12.008.

ZHU, Q. M., BILLINGS, S. A., 1991, *Recursive Parameter Estimation for Nonlinear Rational Models*. Relatório técnico, University of Sheffield, Sheffield.

Appendix A - Householder Transformation with QR Decomposition

The decomposition QR of $\boldsymbol{\Psi}$ (defined in Section 3.3) results in a matrix $\boldsymbol{Q}$ that satisfies the following equations:

$$\boldsymbol{Q\Psi} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} \tag{A.1}$$

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I} \tag{A.2}$$

Seeing that, the extended matrix $\tilde{\boldsymbol{\Psi}}$ is defined, as in Equation 3.9.

A matrix of Householder transformation is defined by the Equations A.3 to A.7:

$$\boldsymbol{H}^{(i)} = \boldsymbol{I} - \boldsymbol{v}^{(i)}\beta^{(i)}(\boldsymbol{v}^{(i)})^T, i = 1, ..., n_\theta \tag{A.3}$$

$$\boldsymbol{v}^{(i)} = \begin{bmatrix} v_1^{(i)} \\ v_2^{(i)} \\ \vdots \\ v_N^{(i)} \end{bmatrix} \tag{A.4}$$

$$v_j^{(i)} = \begin{cases} 0, & j < i \\ \tilde{\psi}_{ii}^{(i-1)} + sign(\tilde{\psi}_{ii}^{(i-1)})\sigma^{(i)}, & j = i \\ \tilde{\psi}_{ji}^{(i-1)}, & j > i \end{cases} \tag{A.5}$$

where $\tilde{\psi}_{ji}^{(i-1)}$ is the term number (ji) of the matrix $\tilde{\boldsymbol{\Psi}}^{(i-1)}$ and $sign(\boldsymbol{X})$ is a function that, for each element of matrix $\boldsymbol{X}$, returns 1, if the element is greater than zero; it returns zero, if the element is equal to zero; and -1, if the element is less than zero.

$$\beta^{(i-1)} = \frac{1}{\sigma^{(i)}(\sigma^{(i)} + |\tilde{\psi}_{ii}^{(i-1)}|)} \tag{A.6}$$

$$\sigma^{(i)} = \sqrt{\sum_{j=i}^{N} (\tilde{\psi}_{ji}^{(i-1)})^2} \tag{A.7}$$

For each transformation, $\tilde{\Psi}^{(i)}$ is calculated, according to Equations A.8 and A.9:

$$\tilde{\boldsymbol{\Psi}}^{(i)} = \boldsymbol{H}^{(i)} \tilde{\boldsymbol{\Psi}}^{(i-1)} \tag{A.8}$$

$$\begin{aligned} \tilde{\boldsymbol{\Psi}}^{(i)} &= [\boldsymbol{I} - \boldsymbol{v}^{(i)} \beta^{(i)} (\boldsymbol{v}^{(i)})^T] \tilde{\boldsymbol{\Psi}}^{(i-1)} \\ &= \tilde{\boldsymbol{\Psi}}^{(i-1)} - \boldsymbol{v}^{(i)} \beta^{(i)} (\boldsymbol{v}^{(i)})^T \tilde{\boldsymbol{\Psi}}^{(i-1)}, \ i = 1, ..., n_\theta \end{aligned} \tag{A.9}$$

After n_θ transformations, it generates $\tilde{\boldsymbol{\Psi}}^{(n_\theta)}$, according to Equation A.10:

$$\tilde{\boldsymbol{\Psi}}^{(n_\theta)} = \begin{bmatrix} \boldsymbol{V}_{n_\theta} & \boldsymbol{y_1^*} \\ \boldsymbol{0} & \boldsymbol{y_2^*} \end{bmatrix} \tag{A.10}$$

Appendix B - User Suggestions - First Case Study

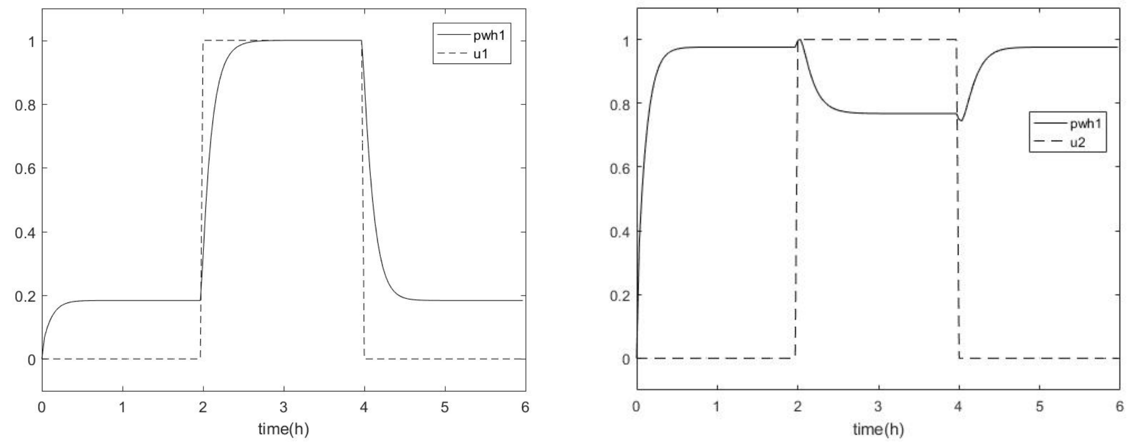
Table B.1: Change on coordinates to identify the model of C_b - First Case Study.

First input	Second input	$R^2_{C_b}$
u_1	u_1/u_2	0.5658
u_1	u_2/u_1	0.1929
u_1	u_1^2/u_2	0.4661
u_1	u_1/u_2^2	0.4790
u_1	u_2^2/u_1	0.8729
u_1	u_2/u_1^2	0.1522
u_1	e^{u_1}	-0.4614
u_1	e^{u_2}	-0.6242
u_1	$\sqrt{u_1}$	-0.0585
u_1	$\sqrt{u_2}$	0.8199
u_1	$\sqrt{u_1}/u_2$	0.6727
u_1	$\sqrt{u_2}/u_1$	0.4089
u_2	u_1/u_2	NaN
u_2	u_2/u_1	NaN
u_2	u_1^2/u_2	-0.7735
u_2	u_1/u_2^2	0.5435
u_2	u_2^2/u_1	0.0816
u_2	u_2/u_1^2	-0.9064
u_2	e^{u_1}	0.4837
u_2	e^{u_2}	0.4879
u_2	$\sqrt{u_1}$	-0.3753
u_2	$\sqrt{u_2}$	0.5426
u_2	$\sqrt{u_1}/u_2$	0.5885
u_2	$\sqrt{u_2}/u_1$	-0.0010

u_1/u_2	u_2/u_1	0.6181
u_1/u_2	u_1^2/u_2	-1.0126
u_1/u_2	u_1/u_2^2	0.5222
u_1/u_2	u_2^2/u_1	0.8141
u_1/u_2	u_2/u_1^2	0.4382
u_1/u_2	e^{u_1}	0.2883
u_1/u_2	e^{u_2}	0.2840
u_1/u_2	$\sqrt{u_1}$	0.3106
u_1/u_2	$\sqrt{u_2}$	0.7630
u_1/u_2	$\sqrt{u_1}/u_2$	0.7226
u_1/u_2	$\sqrt{u_2}/u_1$	0.1834
u_2/u_1	u_1^2/u_2	NaN
u_2/u_1	u_1/u_2^2	NaN
u_2/u_1	u_2^2/u_1	-0.0011
u_2/u_1	u_2/u_1^2	NaN
u_2/u_1	e^{u_1}	NaN
u_2/u_1	e^{u_2}	0.2776
u_2/u_1	$\sqrt{u_1}$	-0.0867
u_2/u_1	$\sqrt{u_2}$	0.8005
u_2/u_1	$\sqrt{u_1}/u_2$	-0.0425
u_2/u_1	$\sqrt{u_2}/u_1$	-0.0011
u_1^2/u_2	u_1/u_2^2	0.3688
u_1^2/u_2	u_2^2/u_1	-0.0550
u_1^2/u_2	u_2/u_1^2	-1.1686
u_1^2/u_2	e^{u_1}	0.1595
u_1^2/u_2	e^{u_2}	0.1473
u_1^2/u_2	$\sqrt{u_1}$	-0.6069
u_1^2/u_2	$\sqrt{u_2}$	0.8026
u_1^2/u_2	$\sqrt{u_1}/u_2$	0.5184
u_1^2/u_2	$\sqrt{u_2}/u_1$	0.1855
u_1/u_2^2	u_2^2/u_1	0.6489
u_1/u_2^2	u_2/u_1^2	0.6030
u_1/u_2^2	e^{u_1}	0.6866
u_1/u_2^2	e^{u_2}	0.6820
u_1/u_2^2	$\sqrt{u_1}$	0.5589
u_1/u_2^2	$\sqrt{u_2}$	0.7648
u_1/u_2^2	$\sqrt{u_1}/u_2$	0.0047
u_1/u_2^2	$\sqrt{u_2}/u_1$	0.1609
u_2^2/u_1	u_2/u_1^2	-0.0024

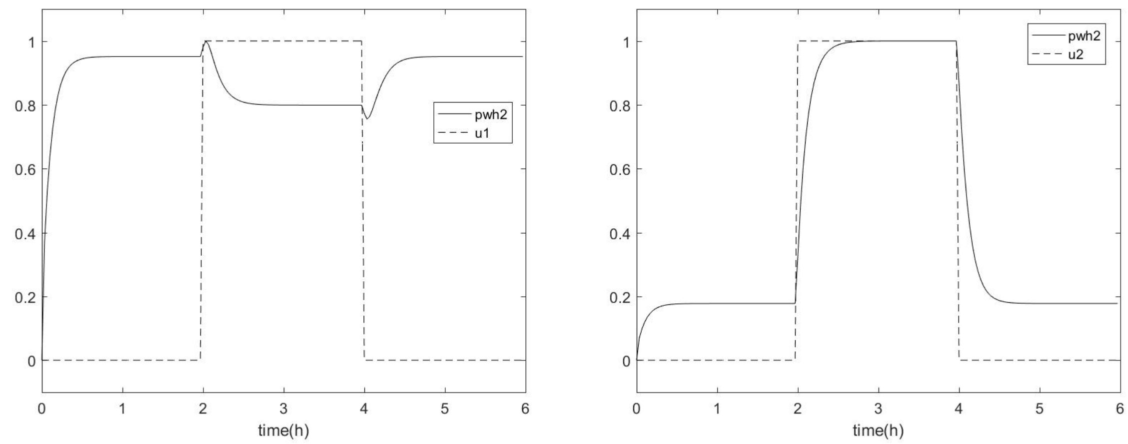
u_2^2/u_1	e^{u_1}	NaN
u_2^2/u_1	e^{u_2}	0.6295
u_2^2/u_1	$\sqrt{u_1}$	NaN
u_2^2/u_1	$\sqrt{u_2}$	0.8039
u_2^2/u_1	$\sqrt{u_1}/u_2$	-0.0010
u_2^2/u_1	$\sqrt{u_2}/u_1$	-0.0010
u_2/u_1^2	e^{u_1}	NaN
u_2/u_1^2	e^{u_2}	-0.3775
u_2/u_1^2	$\sqrt{u_1}$	NaN
u_2/u_1^2	$\sqrt{u_2}$	0.7142
u_2/u_1^2	$\sqrt{u_1}/u_2$	NaN
u_2/u_1^2	$\sqrt{u_2}/u_1$	-0.0010
e^{u_1}	e^{u_2}	-0.1730
e^{u_1}	$\sqrt{u_1}$	-0.0010
e^{u_1}	$\sqrt{u_2}$	0.4995
e^{u_1}	$\sqrt{u_1}/u_2$	-0.4394
e^{u_1}	$\sqrt{u_2}/u_1$	-0.5312
e^{u_2}	$\sqrt{u_1}$	-0.1317
e^{u_2}	$\sqrt{u_2}$	0.4994
e^{u_2}	$\sqrt{u_1}/u_2$	-0.9909
e^{u_2}	$\sqrt{u_2}/u_1$	-0.8326
$\sqrt{u_1}$	$\sqrt{u_2}$	0.8397
$\sqrt{u_1}$	$\sqrt{u_1}/u_2$	0.8587
$\sqrt{u_1}$	$\sqrt{u_2}/u_1$	0.3155
$\sqrt{u_2}$	$\sqrt{u_1}/u_2$	0.5133
$\sqrt{u_2}$	$\sqrt{u_2}/u_1$	NaN
$\sqrt{u_1}/u_2$	$\sqrt{u_2}/u_1$	0.3249

Appendix C - Gathering Information - Second Case Study



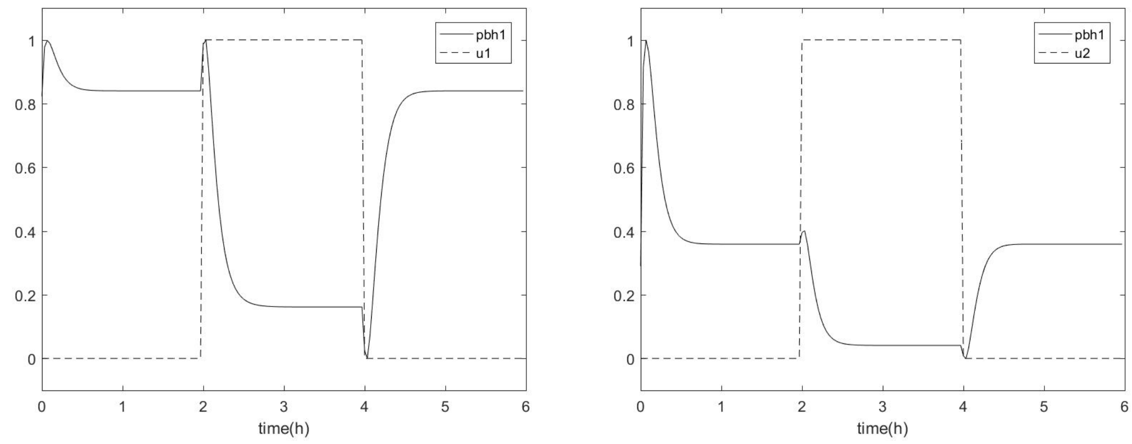
(a) (b)

Figure C.1: Response of p_{wh_1} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



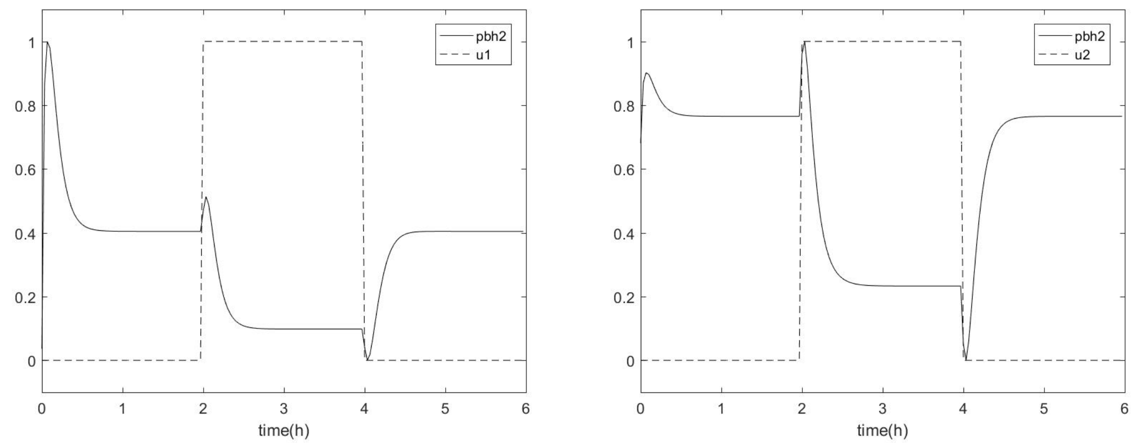
(a) (b)

Figure C.2: Response of p_{wh_2} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



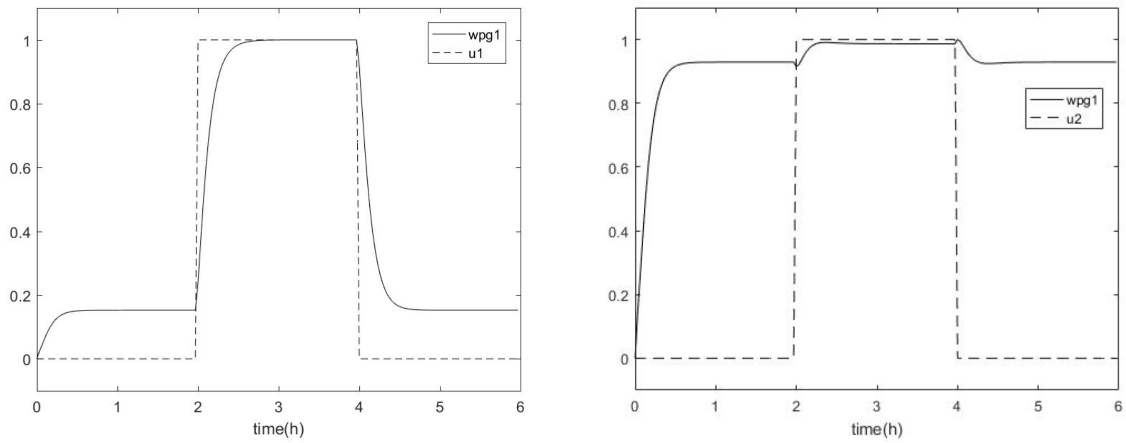
(a) (b)

Figure C.3: Response of p_{bh_1} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



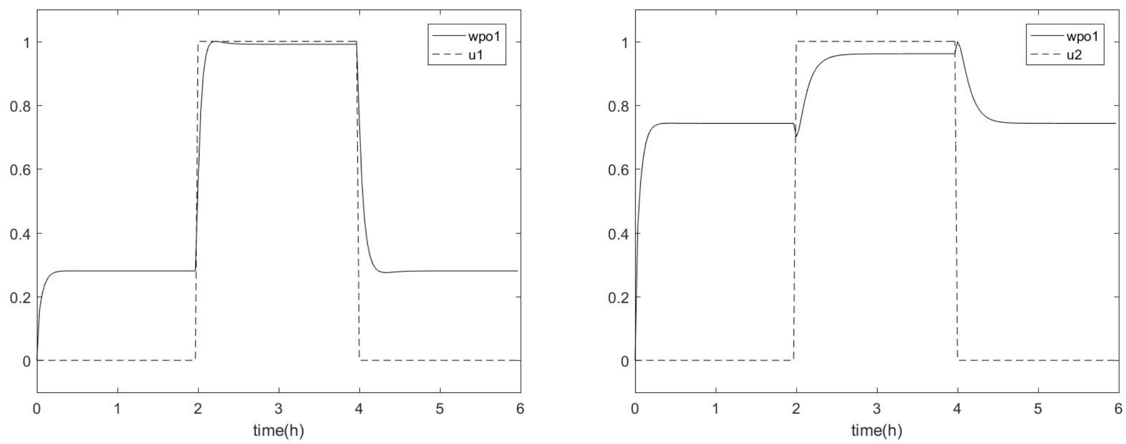
(a) (b)

Figure C.4: Response of p_{bh_2} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



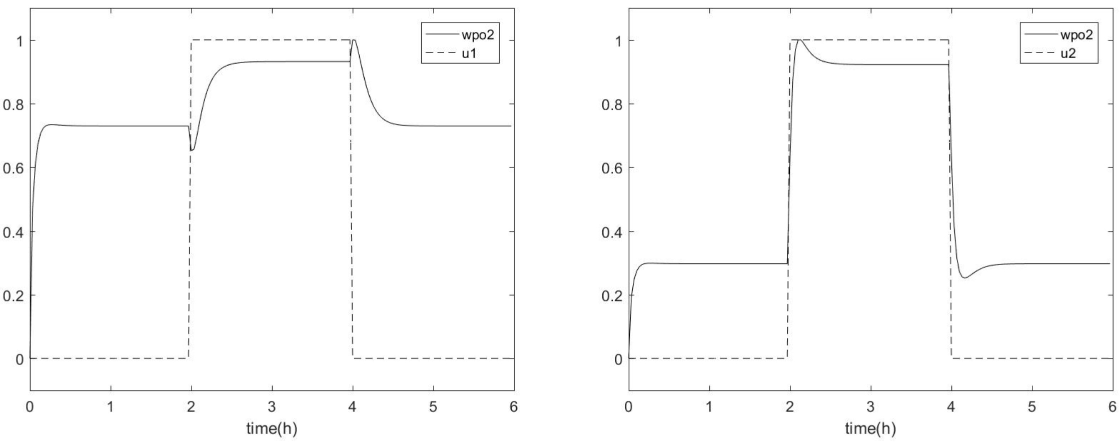
(a) (b)

Figure C.5: Response of w_{pg_1} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



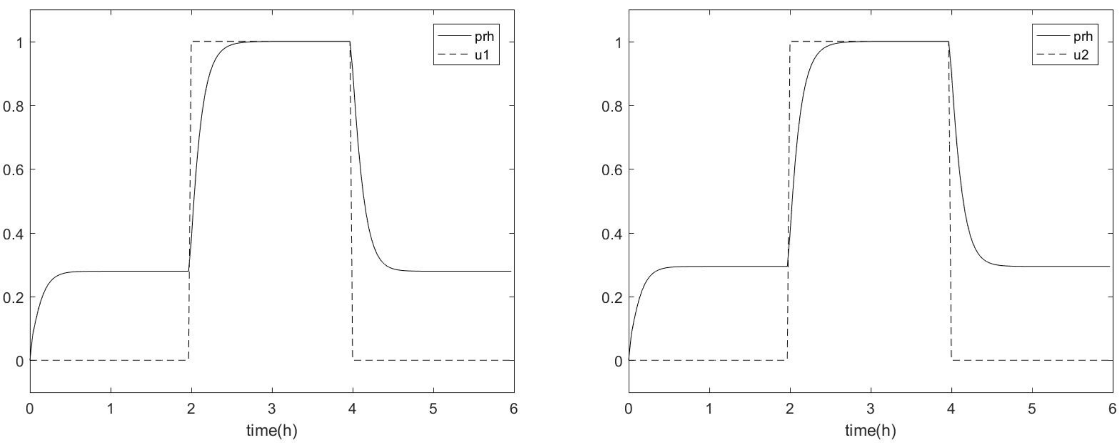
(a) (b)

Figure C.6: Response of w_{po_1} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



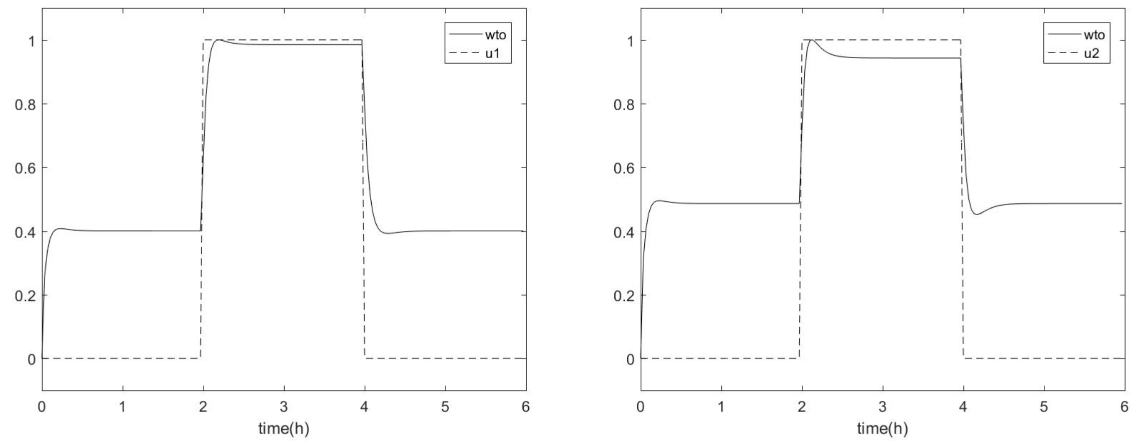
(a) (b)

Figure C.7: Response of w_{po_2} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



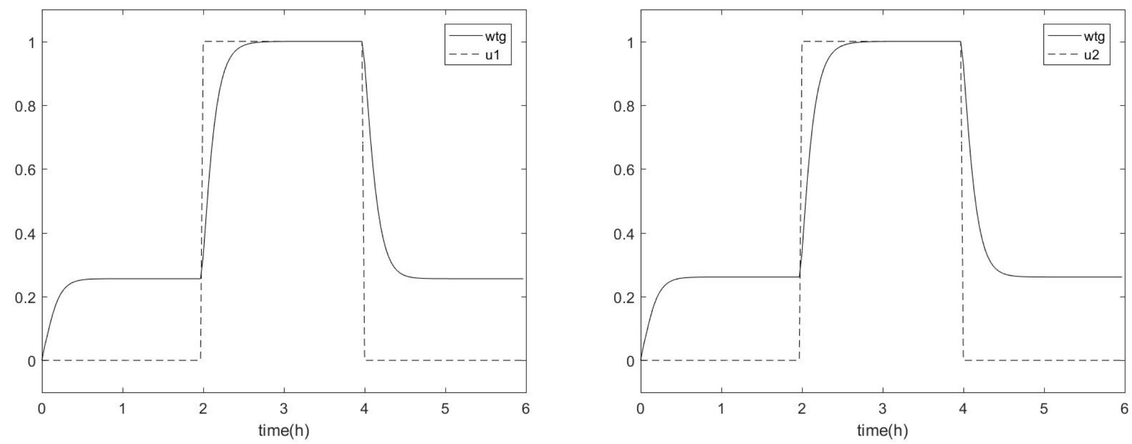
(a) (b)

Figure C.8: Response of p_{rh} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



(a) (b)

Figure C.9: Response of w_{to} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.



(a) (b)

Figure C.10: Response of w_{tg} due to disturbance on the inputs: (a) w_{gl_1}; (b) w_{gl_2}.

Appendix D - Features of the black-box NARMAX models - Second Case Study

$$\hat{\boldsymbol{y}}_{p_{wh_1}} = \boldsymbol{\Psi}^*_{p_{wh_1}} \boldsymbol{\theta}_{p_{wh_1}} \tag{D.1}$$

$$\boldsymbol{\Psi}^{*T}_{p_{wh_1}} = \begin{bmatrix} y_{p_{wh_1}}(k-1) \\ y_{p_{wh_1}}(k-2) \\ u_1(k-1) \\ y_{p_{wh_1}}(k-1)u_1(k-1) \\ y_{p_{wh_1}}(k-2)u_2(k-2) \\ u_1(k-1)u_2(k-1) \\ y_{p_{wh_1}}(k-2)^2 \\ y_{p_{wh_1}}(k-1)^2 \\ y_{p_{wh_1}}(k-1)y_{p_{wh_1}}(k-2) \\ u_2(k-2) \\ u_2(k-2)^2 \\ y_{p_{wh_1}}(k-1)u_2(k-2) \\ u_1(k-1)u_2(k-2) \\ e(k-1)y_{p_{wh_1}}(k-2) \end{bmatrix} \quad \boldsymbol{\theta}_{p_{wh_1}} = \begin{bmatrix} 1.464149 \\ -0.54346 \\ 0.103848 \\ 0.06166 \\ 0.286443 \\ 0.011124 \\ 0.064889 \\ -0.607 \\ 0.453616 \\ 0.006359 \\ -0.00534 \\ -0.36463 \\ 0.066143 \\ -0.05045 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{p_{wh_2}} = \boldsymbol{\Psi}^*_{p_{wh_2}} \boldsymbol{\theta}_{p_{wh_2}} \tag{D.2}$$

$$\boldsymbol{\Psi}^{*T}_{p_{wh_2}} = \begin{bmatrix} y_{p_{wh_2}}(k-1) \\ u_2(k-1) \\ y_{p_{wh_2}}(k-1)u_2(k-1) \\ u_1(k-1) \\ y_{p_{wh_2}}(k-1)^2 \\ u_1(k-1)^2 \\ u_2(k-1)^2 \\ u_2(k-1)u_2(k-2) \\ e(k-2)^2 \end{bmatrix} \quad \boldsymbol{\theta}_{p_{wh_2}} = \begin{bmatrix} 0.7844 \\ 0.2381 \\ -0.0616 \\ 0.0070 \\ 0.0391 \\ -0.0056 \\ -0.0131 \\ 0.0079 \\ 0.1055 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{p_{bh_1}} = \boldsymbol{\Psi}^*_{p_{bh_1}} \boldsymbol{\theta}_{p_{bh_1}} \tag{D.3}$$

$$\boldsymbol{\Psi}^{*T}_{p_{bh_1}} = \begin{bmatrix} y_{p_{bh_1}}(k-1) \\ y_{p_{bh_1}}(k-2) \\ y_{p_{bh_1}}(k-1)u_1(k-1) \\ y_{p_{bh_1}}(k-1)y_{p_{bh_1}}(k-2) \\ u_1(k-4) \\ u_1(k-2) \\ y_{p_{bh_1}}(k-1)u_1(k-4) \\ y_{p_{bh_1}}(k-2)^2 \\ y_{p_{bh_1}}(k-1)u_2(k-3) \\ y_{p_{bh_1}}(k-1)u_1(k-3) \\ u_1(k-1) \\ y_{p_{bh_1}}(k-2)u_1(k-3) \\ u_1(k-4)^2 \\ u_1(k-1)u_1(k-2) \\ e(k-1) \end{bmatrix} \quad \boldsymbol{\theta}_{p_{bh_1}} = \begin{bmatrix} 1.7873 \\ -0.5652 \\ -0.0465 \\ -1.4597 \\ 0.1040 \\ -0.1227 \\ -0.1779 \\ 1.1525 \\ -0.0200 \\ -0.7419 \\ 0.0865 \\ 0.5680 \\ -0.0735 \\ 0.0428 \\ -0.02186 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{p_{bh_2}} = \boldsymbol{\Psi}^*_{p_{bh_2}} \boldsymbol{\theta}_{p_{bh_2}} \tag{D.4}$$

$$\boldsymbol{\Psi}^{*T}_{p_{bh_2}} = \begin{bmatrix} y_{p_{bh_2}}(k-1) \\ y_{p_{bh_2}}(k-2) \\ y_{p_{bh_2}}(k-1)u_2(k-1) \\ y_{p_{bh_2}}(k-1)^2 \\ y_{p_{bh_2}}(k-1)u_2(k-2) \\ y_{p_{bh_2}}(k-2)u_2(k-1) \\ u_2(k-4)^2 \\ y_{p_{bh_2}}(k-1)u_2(k-4) \\ y_{p_{bh_2}}(k-2)u_2(k-3) \\ y_{p_{bh_2}}(k-2)^2 \\ u_1(k-1)u_2(k-4) \\ y_{p_{bh_2}}(k-1)u_1(k-3) \\ y_{p_{bh_2}}(k-1)u_1(k-1) \\ y_{p_{bh_2}}(k-2)u_2(k-4) \\ e(k-1)u_2(k-2) \end{bmatrix} \quad \boldsymbol{\theta}_{p_{bh_2}} = \begin{bmatrix} 1.8734 \\ -0.6051 \\ -0.0395 \\ -0.8908 \\ -0.2835 \\ 0.4207 \\ 0.0399 \\ -0.3133 \\ -0.1858 \\ 0.5294 \\ 0.0177 \\ -0.0498 \\ 0.0276 \\ 0.1388 \\ 0.026872 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{w_{pg_1}} = \boldsymbol{\Psi}^{*}_{w_{pg_1}} \boldsymbol{\theta}_{w_{pg_1}} \tag{D.5}$$

$$\boldsymbol{\Psi}^{*T}_{w_{pg_1}} = \begin{bmatrix} y_{w_{pg_1}}(k-1) \\ u_1(k-1) \\ u_1(k-2) \\ y_{w_{pg_1}}(k-2) \\ u_1(k-3) \\ u_2(k-2) \\ y_{w_{pg_1}}(k-3) \\ e(k-3) \end{bmatrix} \quad \boldsymbol{\theta}_{w_{pg_1}} = \begin{bmatrix} 0.3661 \\ 0.2060 \\ 0.0626 \\ 0.2269 \\ 0.0378 \\ 0.0023 \\ 0.1013 \\ -0.0235 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{w_{pg_2}} = \boldsymbol{\Psi}^{*}_{w_{pg_2}} \boldsymbol{\theta}_{w_{pg_2}} \tag{D.6}$$

$$\boldsymbol{\Psi}^{*T}_{w_{pg_2}} = \begin{bmatrix} y_{w_{pg_2}}(k-1) \\ u_2(k-1) \\ y_{w_{pg_2}}(k-1)y_{w_{pg_2}}(k-4) \\ y_{w_{pg_2}}(k-2) \\ u_2(k-2) \\ u_1(k-1) \\ y_{w_{pg_2}}(k-4)u_1(k-1) \\ y_{w_{pg_2}}(k-3)^2 \\ u_2(k-1)^2 \\ u_1(k-1)^2 \\ u_2(k-3)^2 \\ y_{w_{pg_2}}(k-1)y_{w_{pg_2}}(k-2) \\ y_{w_{pg_2}}(k-1)u_2(k-4) \\ y_{w_{pg_2}}(k-2)u_2(k-2) \\ e(k-1)u_2(k-1) \end{bmatrix} \quad \boldsymbol{\theta}_{w_{pg_2}} = \begin{bmatrix} 0.3599 \\ 0.2418 \\ 0.1648 \\ 0.3706 \\ 0.0540 \\ 0.0187 \\ -0.0141 \\ 0.0554 \\ -0.0301 \\ -0.0086 \\ 0.0353 \\ -0.3261 \\ 0.0371 \\ 0.0360 \\ -0.03476 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{w_{po_1}} = \boldsymbol{\Psi}^*_{w_{po_1}} \boldsymbol{\theta}_{w_{po_1}} \tag{D.7}$$

$$\boldsymbol{\Psi}^{*T}_{w_{po_1}} = \begin{bmatrix} y_{w_{po_1}}(k-1) \\ y_{w_{po_1}}(k-2) \\ y_{w_{po_1}}(k-1)^2 \\ u_1(k-1) \\ u_1(k-1)u_1(k-4) \\ u_1(k-2) \\ u_2(k-2) \\ y_{w_{po_1}}(k-2)^2 \\ u_1(k-3)u_2(k-1) \\ y_{w_{po_1}}(k-1)u_1(k-3) \\ y_{w_{po_1}}(k-4) \\ u_1(k-4)^2 \\ y_{w_{po_1}}(k-3)u_1(k-2) \\ u_2(k-1)^2 \\ e(k-2)u_1(k-3) \end{bmatrix} \quad \boldsymbol{\theta}_{w_{po_1}} = \begin{bmatrix} 0.843586 \\ -0.00153 \\ -0.19414 \\ 0.510861 \\ -0.16781 \\ -0.16492 \\ 0.02431 \\ 0.218338 \\ -0.01662 \\ -0.18764 \\ -0.09074 \\ 0.046826 \\ 0.120682 \\ -0.00657 \\ -0.04545 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{w_{po_2}} = \boldsymbol{\Psi}^*_{w_{po_2}} \boldsymbol{\theta}_{w_{po_2}} \tag{D.8}$$

$$\boldsymbol{\Psi}^{*T}_{w_{po_2}} = \begin{bmatrix} y_{w_{po_2}}(k-1) \\ y_{w_{po_2}}(k-2) \\ u_1(k-2) \\ y_{w_{po_2}}(k-4) \\ u_2(k-2) \\ u_2(k-1) \\ u_2(k-3) \\ u_1(k-1) \\ y_{w_{po_2}}(k-3) \\ u_1(k-4) \\ e(k-2) \end{bmatrix} \quad \boldsymbol{\theta}_{w_{po_2}} = \begin{bmatrix} 0.9467 \\ 0.0736 \\ 0.0252 \\ 0.0469 \\ -0.1777 \\ 0.2316 \\ -0.0620 \\ -0.0197 \\ -0.0707 \\ 0.0058 \\ 0.0041 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{p_{rh}} = \boldsymbol{\Psi}^{*}_{p_{rh}} \boldsymbol{\theta}_{p_{rh}} \tag{D.9}$$

$$\boldsymbol{\Psi}^{*T}_{p_{rh}} = \begin{bmatrix} y_{p_{rh}}(k-1) \\ y_{p_{rh}}(k-3) \\ u_1(k-1) \\ u_2(k-1) \\ y_{p_{rh}}(k-2) \\ u_2(k-2) \\ u_1(k-2) \\ u_1(k-3) \\ u_2(k-3) \\ e(k-2) \end{bmatrix} \quad \boldsymbol{\theta}_{p_{rh}} = \begin{bmatrix} 0.2574 \\ 0.1290 \\ 0.1364 \\ 0.1307 \\ 0.2713 \\ 0.0320 \\ 0.0304 \\ 0.0154 \\ 0.0112 \\ -0.0035 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{p_m} = \boldsymbol{\Psi}^{*}_{p_m} \boldsymbol{\theta}_{p_m} \tag{D.10}$$

$$\boldsymbol{\Psi}^{*T}_{p_m} = \begin{bmatrix} y_{p_m}(k-1) \\ y_{p_m}(k-2) \\ y_{p_m}(k-3)u_1(k-2) \\ u_1(k-1) \\ y_{p_m}(k-1)u_2(k-2) \\ y_{p_m}(k-3)u_2(k-1) \\ u_1(k-3) \\ y_{p_m}(k-1)^2 \\ u_1(k-1)u_2(k-1) \\ y_{p_m}(k-1)u_2(k-3) \\ y_{p_m}(k-1)u_1(k-4) \\ y_{p_m}(k-3)u_1(k-3) \\ y_{p_m}(k-1)u_2(k-4) \\ u_1(k-4)^2 \\ e(k-2)y_{p_m}(k-3) \end{bmatrix} \quad \boldsymbol{\theta}_{p_m} = \begin{bmatrix} 1.2171 \\ 0.2443 \\ -0.1815 \\ 0.1121 \\ -0.1652 \\ 0.1619 \\ -0.1657 \\ -0.6616 \\ 0.1249 \\ -0.2359 \\ -0.3037 \\ 0.2930 \\ -0.1550 \\ 0.0339 \\ -0.27136 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{wto} = \boldsymbol{\Psi}^{*}_{wto}\boldsymbol{\theta}_{wto} \tag{D.11}$$

$$\boldsymbol{\Psi}^{*T}_{wto} = \begin{bmatrix} y_{w_{to}}(k-1) \\ y_{w_{to}}(k-2) \\ y_{w_{to}}(k-3) \\ u_1(k-1) \\ u_1(k-2) \\ u_2(k-1) \\ u_2(k-2) \\ y_{w_{to}}(k-4) \\ u_2(k-3) \\ e(k-1) \end{bmatrix} \quad \boldsymbol{\theta}_{wto} = \begin{bmatrix} 1.2545 \\ -0.2245 \\ -0.1290 \\ 0.1298 \\ -0.1114 \\ 0.0740 \\ -0.0855 \\ 0.0655 \\ 0.0182 \\ -0.0019 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{wtg} = \boldsymbol{\Psi}^{*}_{wtg}\boldsymbol{\theta}_{wtg} \tag{D.12}$$

$$\boldsymbol{\Psi}^{*T}_{wtg} = \begin{bmatrix} y_{w_{tg}}(k-1) \\ y_{w_{tg}}(k-2) \\ u_1(k-1) \\ u_2(k-1) \\ u_1(k-2) \\ u_2(k-2) \\ u_2(k-3) \\ u_1(k-3) \end{bmatrix} \quad \boldsymbol{\theta}_{wtg} = \begin{bmatrix} 0.5027 \\ 0.2428 \\ 0.0972 \\ 0.1003 \\ 0.0238 \\ 0.0191 \\ 0.0114 \\ 0.0116 \end{bmatrix}$$

Appendix E - User Suggestions - Second Case Study

Table E.1: Change on coordinates to identify the models - Second Case Study.

First input	Second input	$R^2_{p_{bh_2}}$	$R^2_{w_{po_2}}$	$R^2_{w_{to}}$
u_1	u_1/u_2	-0.6769	0.2601	-0.0406
u_1	u_2/u_1	-1.1680	0.4182	0.6663
u_1	u_1^2/u_2	-0.4230	0.3538	-0.0890
u_1	u_1/u_2^2	-0.4561	0.4068	0.1867
u_1	u_2^2/u_1	-0.8604	0.5572	0.7188
u_1	u_2/u_1^2	0.0111	0.6082	0.7658
u_1	$\sqrt{u_1}$	-28.9291	-19.7444	-15.3672
u_1	$\sqrt{u_2}$	-1.1052	0.2540	0.5652
u_1	$\sqrt{u_1}/u_2$	-0.1662	0.3498	0.2837
u_1	$\sqrt{u_2}/u_1$	-0.7201	-874.1604	0.6888
u_2	u_1/u_2	-0.2654	0.5775	0.5587
u_2	u_2/u_1	-0.6767	0.5725	-0.2282
u_2	u_1^2/u_2	-0.6916	0.6851	0.4574
u_2	u_1/u_2^2	-0.8165	0.7002	0.5760
u_2	u_2^2/u_1	-0.6767	0.6105	-0.7905
u_2	u_2/u_1^2	-0.6767	0.5760	-0.1071
u_2	$\sqrt{u_1}$	-0.7013	0.5813	0.0155
u_2	$\sqrt{u_2}$	-22.7267	-18.1608	-21.7015
u_2	$\sqrt{u_1}/u_2$	-0.8112	0.5768	0.5992
u_2	$\sqrt{u_2}/u_1$	-0.6767	0.5590	0.3318
u_1/u_2	u_2/u_1	0.8867	0.5843	0.7016
u_1/u_2	u_1^2/u_2	-4.1315	-2.1390	-1.4719
u_1/u_2	u_1/u_2^2	-2.3793	0.5967	0.1436
u_1/u_2	u_2^2/u_1	0.8130	0.6425	0.7380

u_1/u_2	u_2/u_1^2	0.7424	0.4860	0.6097
u_1/u_2	$\sqrt{u_1}$	0.4621	0.1572	-0.8914
u_1/u_2	$\sqrt{u_2}$	-0.0762	0.3226	0.7808
u_1/u_2	$\sqrt{u_1}/u_2$	-5.5474	-0.8549	-2.5010
u_1/u_2	$\sqrt{u_2}/u_1$	0.8358	0.4761	0.6229
u_2/u_1	u_1^2/u_2	0.4732	0.7037	0.5633
u_2/u_1	u_1/u_2^2	0.1410	0.5646	0.5271
u_2/u_1	u_2^2/u_1	-5.8955	-0.6152	-2.0907
u_2/u_1	u_2/u_1^2	-6.1042	0.5175	-1.9163
u_2/u_1	$\sqrt{u_1}$	-1.1304	0.5795	0.2443
u_2/u_1	$\sqrt{u_2}$	-1.6755	0.3506	-1.9162
u_2/u_1	$\sqrt{u_1}/u_2$	0.2033	-1.0062	0.5486
u_2/u_1	$\sqrt{u_2}/u_1$	-11.8883	-6.0243	-3.4989
u_1^2/u_2	u_1/u_2^2	-0.9713	0.1706	-0.5357
u_1^2/u_2	u_2^2/u_1	0.4972	0.1941	0.7552
u_1^2/u_2	u_2/u_1^2	0.7286	0.6294	0.6316
u_1^2/u_2	$\sqrt{u_1}$	0.2633	0.3756	0.4050
u_1^2/u_2	$\sqrt{u_2}$	-0.9282	0.4064	0.8151
u_1^2/u_2	$\sqrt{u_1}/u_2$	-2.0311	-0.0224	-0.2853
u_1^2/u_2	$\sqrt{u_2}/u_1$	0.7866	0.6586	0.6887
u_1/u_2^2	u_2^2/u_1	0.6666	0.4732	0.6467
u_1/u_2^2	u_2/u_1^2	0.5758	0.4097	0.5257
u_1/u_2^2	$\sqrt{u_1}$	0.7000	0.3560	-0.2723
u_1/u_2^2	$\sqrt{u_2}$	-0.1480	0.4173	0.7634
u_1/u_2^2	$\sqrt{u_1}/u_2$	-6.1006	-3.1060	-2.6313
u_1/u_2^2	$\sqrt{u_2}/u_1$	0.3652	-2.6345	0.5794
u_2^2/u_1	u_2/u_1^2	-2.8397	0.5723	0.2160
u_2^2/u_1	$\sqrt{u_1}$	-0.8767	0.5347	0.2645
u_2^2/u_1	$\sqrt{u_2}$	-1.5390	-0.3817	-1.7555
u_2^2/u_1	$\sqrt{u_1}/u_2$	0.1923	0.5188	0.5724
u_2^2/u_1	$\sqrt{u_2}/u_1$	-3.6250	0.5980	0.3742
u_2/u_1^2	$\sqrt{u_1}$	-0.1947	0.7442	0.0697
u_2/u_1^2	$\sqrt{u_2}$	-1.5980	0.3392	-1.5139
u_2/u_1^2	$\sqrt{u_1}/u_2$	0.0146	0.4372	0.5373
u_2/u_1^2	$\sqrt{u_2}/u_1$	-3.8088	-1.4521	-3.3082
$\sqrt{u_1}$	$\sqrt{u_2}$	-1.1733	0.2241	0.5165
$\sqrt{u_1}$	$\sqrt{u_1}/u_2$	-0.6151	0.3687	0.3083
$\sqrt{u_1}$	$\sqrt{u_2}/u_1$	-1.5104	0.4003	0.6615
$\sqrt{u_2}$	$\sqrt{u_1}/u_2$	-0.7675	0.6919	0.5940

$\sqrt{u_2}$	$\sqrt{u_2}/u_1$	-0.7219	0.6469	0.2953
$\sqrt{u_1}/u_2$	$\sqrt{u_2}/u_1$	0.6644	0.3817	0.5547

Appendix F - Features of the gray-box NARMAX models - Second Case Study

$$\hat{\boldsymbol{y}}_{p_{bh_2}} = \boldsymbol{\Psi}^{*}_{p_{bh_2}} \boldsymbol{\theta}_{p_{bh_2}} \tag{F.1}$$

$$\boldsymbol{\Psi}^{*T}_{p_{bh_2}} = \begin{bmatrix} y_{p_{bh_2}}(k-1) \\ y_{p_{bh_2}}(k-2) \\ u_1(k-1) \\ u_2(k-4) \\ u_2(k-2) \\ u_1(k-4) \\ u_1(k-2) \\ u_2(k-3) \\ y_{p_{bh_2}}(k-3) \\ u_1(k-3) \end{bmatrix} \quad \boldsymbol{\theta}_{p_{bh_2}} = \begin{bmatrix} 1.4030 \\ -0.4210 \\ 0.0325 \\ 0.0063 \\ -0.0053 \\ -0.0422 \\ 0.0458 \\ 0.0104 \\ -0.0194 \\ -0.0014 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{w_{po_2}} = \boldsymbol{\Psi}^{*}_{w_{po_2}} \boldsymbol{\theta}_{w_{po_2}} \tag{F.2}$$

$$\boldsymbol{\Psi}^{*T}_{w_{po_2}} = \begin{bmatrix} y_{w_{po_2}}(k-1) \\ y_{w_{po_2}}(k-2) \\ y_{w_{po_2}}(k-4) \\ u_1(k-1) \\ u_1(k-2) \\ y_{w_{po_2}}(k-3) \\ u_2(k-1) \\ u_2(k-2) \\ u_1(k-3) \\ u_2(k-3) \\ u_1(k-4) \end{bmatrix} \quad \boldsymbol{\theta}_{w_{po_2}} = \begin{bmatrix} 1.0324 \\ -0.0024 \\ 0.0397 \\ 0.7475 \\ -0.5893 \\ -0.0877 \\ -0.5960 \\ 0.4609 \\ -0.1477 \\ 0.1194 \\ 0.0129 \end{bmatrix}$$

$$\hat{\boldsymbol{y}}_{w_{to}} = \boldsymbol{\Psi}^{*}_{w_{to}}\boldsymbol{\theta}_{w_{to}} \tag{F.3}$$

$$\boldsymbol{\Psi}^{*T}_{w_{to}} = \begin{bmatrix} y_{w_{to}}(k-1) \\ y_{w_{to}}(k-2) \\ y_{w_{to}}(k-4) \\ u_1(k-1) \\ u_2(k-4) \\ u_1(k-2) \\ u_2(k-2) \\ u_2(k-1) \\ y_{w_{to}}(k-3) \\ u_2(k-3) \\ u_1(k-3) \end{bmatrix} \qquad \boldsymbol{\theta}_{w_{to}} = \begin{bmatrix} 1.3654 \\ -0.3129 \\ 0.0745 \\ 0.1405 \\ 0.0179 \\ -0.1483 \\ -0.1304 \\ 0.0930 \\ -0.1581 \\ 0.0323 \\ 0.0296 \end{bmatrix}$$