JOSÉ OTÁVIO ASSUMPÇÃO MATIAS

Real-time Optimization with Persistent
Parameter Adaptation using Online
Parameter Estimation

São Paulo

2018

JOSÉ OTÁVIO ASSUMPÇÃO MATIAS

**Real-time Optimization with Persistent Parameter
Adaptation using Online Parameter Estimation**

Thesis submitted to the Escola Politécnica
of the University of São Paulo in partial
fulfillment of the requirements for the degree
of Doctor of Science.

Area of concentration: Chemical Engi-
neering

Supervisor: Prof. Dr. Galo Antônio Carrillo
Le Roux

São Paulo
2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Matias, José Otávio Assumpção

Real-time Optimization with Persistent Parameter Adaptation using Online Parameter Estimation / J. O. A. Matias -- versão corr. -- São Paulo, 2018.

133 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Química.

1.Otimização em tempo real 2.Estimação em tempo real 3.Detecção de estado estacionário 4.Otimização global da planta I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Química II.t.

Acknowledgements

Há muitas pessoas que merecem, de certa maneira, algum crédito nessa tese. Seja me ajudando durante as matérias da pós-graduação, com a discussão de ideias, diretamente com os códigos, ou apenas me acompanhando nas inúmeras pausas e cafézinhos durante os anos de doutorado. Por isso meu primeiro agradecimento vai a todos os meus colegas que me ajudaram tanto (e, principalmente, me atrapalharam tanto) na conclusão dessa tese. Sua inspiração e encorajamento foram essenciais. Não vou citá-los nominalmente, pois com certeza cometeria alguma injustiça, esquecendo o nome de alguém.

Sem dúvida, gostaria de agradecer ao professor Galo Le Roux pela supervisão do trabalho. Sem as suas contribuições e sugestões certamente essa tese não estaria pronta. Também queria agradecer ao professor Johannes Jäschke pela supervisão no meu período de doutorado sanduíche na Noruega.

Como não poderia faltar, gostaria de agradecer às instâncias superiores que me ajudaram a realizar o doutorado. Às que me proporcionaram o suporte financeiro (ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Petrobrás e ao INTPART/SUBPRO) e às transcendentais que me fizeram suportar esses anos todos. Obrigado a Deus, a todos os santos, a São Marcos, aos anjos e arcanjos, aos Orixás, aos Uerês, a Alá, a Jeová, a Judá, a todos os espíritos de luz, às Fadas, ao deus da tribo da Polinésia e quem mais estiver por aí, porque não é bom ficar em dívida com ninguém.

Finalmente, gostaria de agradecer a todas as famílias que me ajudaram durante todos esses anos, a minhas famílias postíças da Capote Valente e da Rua dos Pinheiros, a minha família real, meus tios, tias, primos, primas, avôs e avós. Entre eles, um agradecimento especial a minha tia Cristina, meu tio Fernando e a minha avó Stella por tudo. Mais importante, gostaria de agradecer a minha irmã e a meus pais pelo imenso suporte durante todos esses anos, não sei como uma família poderia ser mais especial. Por último, mas não menos importante, gostaria de agradecer a minha família mais recente, te amo minha Lindinha. Sem você nada disso teria acontecido, nem valido a pena.

“ - Olha, dona Evarista - disse-lhe o padre Lopes, vigário do lugar -, veja se seu marido dá um passeio ao Rio de Janeiro . Isso de estudar sempre, sempre, não é bom, vira o juízo.”

(Machado de Assis - 'O Alienista')

Abstract

In standard Real-time Optimization (RTO) implementations, the plant needs to be sufficiently steady in order to update the RTO model parameters reliably. However, this condition is seldom found in practice. Moreover, because the RTO model is only updated when the plant enters a stationary condition, the optimizer is likely to be out of phase with highly perturbed plants. The main contribution of the thesis is the proposal of an alternative RTO approach, called Real-time Optimization with Persistent Adaptation (ROPA), which integrates on-line parameter estimation in the optimization cycle, avoiding the steady-state (SS) detection step. Instead of predicting the SS, the online estimator keeps the model up-to-date with the plant and allows running the economic optimization at any time, even instants after implementing the current RTO decisions. ROPA provides an intermediary solution between static and dynamic optimization schemes. While it approximates the optimal trajectory, ROPA enables the use of well-established static RTO commercial solutions. Furthermore, the new approach is the key for decoupling the model estimation problem in order to achieve plant-wide optimization. Another contribution of the thesis is to provide several case studies in which ROPA is tested and compared with the standard RTO implementation: a Williams-Otto reactor, a Fluid Catalyst Cracking unit and a separation-reaction system. The idea is to illustrate ROPA convergence properties and how the plant-wide optimum is achieved by asynchronously updating the global plant model. The results show that ROPA is able to track the stationary (plant-wide) optimum. In addition, they confirm that the refinement of the prediction capacity, by decreasing the time between two sequential optimization, enhances the disturbance detection of the optimization cycle and leads to a better economic performance.

Keywords: Real-time optimization. Online estimation. Steady-state detection. Plant-wide optimization.

Resumo

Na implementação padrão de otimização em tempo real (RTO, do inglês *real-time optimization*), a planta deve estar suficientemente estável para que os parâmetros do modelo usado no RTO sejam estimados com precisão. Contudo, esta condição é raramente encontrada na prática. Além disso, devido ao fato de o modelo usado no RTO ser atualizado somente quando a planta entra em estado estacionário, é provável que o otimizador esteja fora de fase quando implementado em plantas com alta frequência de distúrbios. A principal contribuição desta tese é o desenvolvimento de uma metodologia alternativa de RTO chamada otimização em tempo real com atualização persistente de parâmetros (ROPA, do inglês *real-time optimization with persistent adaptation*). A nova metodologia integra estimadores em tempo real ao ciclo de otimização, evitando assim a necessidade da etapa de detecção de estado estacionário. Ao invés de identificá-lo, o estimador em tempo real mantém o modelo atualizado com a planta e permite que se execute a otimização econômica em tempos arbitrários, mesmo instantes depois da implementação da decisão ótima calculada anteriormente pelo RTO. ROPA provê uma solução intermediária entre a otimização estática e dinâmica. Ao mesmo tempo que aproxima a trajetória ótima, ela permite o uso de soluções comerciais já estabelecidas de RTO estacionário. Também, a nova metodologia é a chave para o desacoplamento do problema de estimação a fim de se atingir o ótimo global da planta. Uma contribuição adicional da tese é a apresentação de três casos de estudo que testam a ROPA e comparam sua performance à implementação padrão de RTO: um reator Williams-Otto, uma unidade de craqueamento catalítico e um sistema de separação-reação. A ideia principal é ilustrar as propriedades de convergência da nova metodologia e como a atualização assíncrona do modelo global da planta pode ser usada para atingir o ótimo da planta como um todo. Os resultados mostram que a ROPA é capaz de alcançar o ótimo estacionário da planta. Adicionalmente, o refinamento da capacidade de predição através da diminuição do tempo entre duas execuções sequenciais do otimizador melhora a capacidade de detecção de distúrbios do ciclo de otimização assim como a performance econômica.

Palavras-chaves: Otimização em tempo real. Estimação em tempo real. Detecção de estado estacionário. Otimização global da planta.

List of Figures

Figure 1 – Plant decision hierarchy	23
Figure 2 – Real-time Optimization (RTO) steps	24
Figure 3 – Comparison between standard RTO and ROPA	34
Figure 4 – Williams-Otto reactor flowsheet	48
Figure 5 – Case study 1: Deterministic disturbances profiles	51
Figure 6 – Case study 1: Estimated states profiles	56
Figure 7 – Case study 1: Estimated parameter profiles	58
Figure 8 – Case study 1: Comparison of the input moves computed by MPA and ROPA	60
Figure 9 – Case study 1: Profit comparison between MPA and ROPA	62
Figure 10 – Fluid catalyst cracking (FCC) flowsheet	65
Figure 11 – Case study 2: Comparison of the optimal gasoline production and the achieved by ROPA and MPA	70
Figure 12 – Case study 2: Estimated parameters profiles	72
Figure 13 – Case study 2: Steady-state detection	73
Figure 14 – Separation-reaction system flowsheet	77
Figure 15 – Case study 3: Deterministic disturbances profiles	78
Figure 16 – Case study 3: Column model diagram	79
Figure 17 – Case study 3: Steady-state detection MPA	86
Figure 18 – Case study 3: Steady-state detection ROPA	86
Figure 19 – Case study 3: Estimated parameter profiles	87
Figure 20 – Case study 3: Plant-wide instantaneous profit comparison between MPA and ROPA	88
Figure 21 – Case study 3: Plant-wide profit comparison between MPA and ROPA	89
Figure 22 – Fluid catalyst cracking flowsheet (repeated)	108
Figure 23 – FCC validation - measured values	113
Figure 24 – FCC validation - system inputs	113
Figure 25 – FCC validation - estimated parameter values 1	114
Figure 26 – FCC validation - estimated parameter values 2	115
Figure 27 – Modeling options based on available solvers and information	119
Figure 28 – Vapor recompression distillation flowsheet	121

Figure 29 – DDPA validation - compositions	122
Figure 30 – Model structure adaptation algorithm concept	127
Figure 31 – Model structure adaptation algorithm embedded in the optimization cycle	128
Figure 32 – Gas lifted well network flowsheet	129
Figure 33 – Gas lifted well network - plant surface	131
Figure 34 – Models structure selection while optimizing the well network	132

List of Tables

Table 1 – Initial condition for the Williams-Otto reactor	50
Table 2 – Adjustable parameter set of the Williams-Otto model	50
Table 3 – Initial condition for the FCC process	67
Table 4 – Adjustable parameter set of the FCC model	67
Table 5 – Initial condition for the separation-reaction system	82
Table 6 – Adjustable parameter set of the separation-reaction system model . . .	82
Table 7 – Comparison of ROPA and MPA implementations	83
Table 8 – Tuning parameters for the simulation of the separation-reaction system	84
Table 9 – Measurements used for validation of the FCC model	111
Table 10 – FCC adjustable model parameters for validation	112
Table 11 – VLE modeling strategies - Distillation Columns	117
Table 12 – Model structure adaptation - marker and model correspondence	131

List of abbreviations and acronyms

CEKF	Constrained extended Kalman filter
CSTR	Continuous stirred-tank reactor
DAE	Differential-algebraic equation
DCS	Distributed control system
DDPA	Dynamic design using proportional action
DRTO	Dynamic real-time optimization
EKF	Extended Kalman filter
EMPC	Economic model predictive control
EMSO	Environment for modeling, simulation and optimization
FCC	Fluid catalytic cracking
MA	Modifier adaptation
MHE	Moving horizon estimator
MPA	Model parameter adaptation
MPC	Model predictive control
ODE	Ordinary differential equation
PI	Proportional-integral
rEKF	Reduced extended Kalman filter
ROPA	Real-time optimization with persistent adaptation
RTE	Real-time evolution
RTO	Real-time optimization
SS	Steady-state
SSD	Steady-state detection

VLE	Vapor liquid equilibrium
VRD	Vapor recompression distillation

List of symbols

General notation

$\mathbf{y}_p$	plant input-output mapping
$\mathbf{u}$	system inputs, dimension = n_u (i.e $\mathbf{u} \in \mathbb{R}^{n_u}$)
$\mathbf{d}_p$	deterministic disturbances, dimension = n_d
$\boldsymbol{\epsilon}_p$	random disturbances, dimension = n_n
k	whenever used as subscript, it indicates the variable at time k
$\mathbf{f}_{ss}$	steady-state model
$\mathbf{x}$	system states, dimension = n_x
$\mathbf{p}$	system parameters, dimension = n_p
$\mathbf{y}$	system outputs, dimension = n_y
$\mathbf{h}$	output function
$\boldsymbol{\nu}_{ss}$	measurement noise - Gaussian random noise with zero mean and constant covariance matrix $\mathbf{R}_p$
$\mathbf{f}_{dyn}$	dynamic discrete model
$\boldsymbol{\omega}_k$	process noise - Gaussian random noise with zero mean and constant covariance matrix $\mathbf{Q}$
$\boldsymbol{\nu}_k$	measurement noise - Gaussian random noise with zero mean and constant covariance matrix $\mathbf{R}$
z	process measurement used in steady-state detection
z_f	filtered value of the measurement used in steady-state detection
λ_1	filter factor for z_f
$\delta_{1,f}$	variance of the measurement z with its filtered trend z_f
λ_2	filter factor for $\delta_{1,f}$

$\delta_{2,f}$	variance between two consecutive measurements z
λ_3	filter factor for $\delta_{2,f}$
R_{SS}	ratio of variances
$\mathbf{p}^{adj}$	adjustable parameter set, which is a subset of the model parameters
$\mathbf{p}^{nom}$	non-adjustable parameter set, which is a subset of the model parameters ($\mathbf{p}^{adj} \cup \mathbf{p}^{nom} = \mathbf{p}$)
ϕ_{adap}	objective function of the model adaptation problem
$\mathbf{R}_p$	weights used in ϕ_{adap}
G_{adap}	constraints of the model adaptation problem
$\mathbf{p}^{L,adj}$	adjustable parameter set lower bounds
$\mathbf{p}^{U,adj}$	adjustable parameter set upper bounds
$\star$	whenever used as superscript, it indicates the solution of the optimization problem
ϕ_{profit}	objective function of the economic optimization problem
G	constraints of the economic optimization problem
$\mathcal{U}$	operational constraints
$\mathbf{u}^L$	inputs lower bounds
$\mathbf{u}^U$	inputs upper bounds
$\mathbf{y}^L$	outputs lower bounds
$\mathbf{y}^U$	outputs upper bounds
$\mathbf{x}^e$	extended system states, dimension = $n_x + n_p$
$\boldsymbol{\omega}^p$	parameter noise - Gaussian random noise with zero mean and constant covariance matrix Q^p
$\mathbf{f}^e$	extended dynamic model

$\mathbf{h}^e$	extended output function
$\boldsymbol{\omega}^e$	extended noise array
$\hat{\mathbf{x}}^e$	estimated extended states
$\mathbf{F}$	linearized extended dynamic model
$\mathbf{H}$	linearized extended output function
$\mathbf{P}$	estimate covariance matrix
$\mathbf{K}$	filter gain
$\mathbf{V}_{\mathbf{P}}$	eigenvector matrix of P
$\mathbf{v}_{\mathbf{P}}$	eigenvector of P
$\boldsymbol{\Lambda}_{\mathbf{P}}$	eigenvalue matrix of P
$\lambda_{\mathbf{P}}$	eigenvalue of P
$\mathbf{P}_q$	reduced estimate covariance matrix
λ_{thres}	threshold for lower-rank approximation of the estimate covariance matrix used in rEKF

Chapter 3: Williams-Otto reactor

F_R	reactor outlet mass flow rate [kg/s]
F_A	reactor inlet mass flow rate of component A [kg/s]
F_B	reactor inlet mass flow rate of component B [kg/s]
X_A	mass fraction of component A [-]
X_B	mass fraction of component B [-]
X_C	mass fraction of component C [-]
X_E	mass fraction of component E [-]
X_P	mass fraction of component P [-]

X_G	mass fraction of component G [-]
t	time [s]
M_t	reactor holdup [kg]
k_i	reaction rate constant of the i^{th} reaction [1/s]
A_i	frequency factor of the i^{th} reaction [1/s]
E_i	activation energy of the i^{th} reaction [K]
T_R	reactor temperature [°C]
Φ_i	rearranged parameters of the i^{th} reaction [-]
Ψ_i	rearranged parameters of the i^{th} reaction [-]
T_{ref}	reference temperature for parameter rearrangement [K]
$\mathbf{K}_{opt}$	optimal decision filter factor [-]
Δt_{ROPA}	time between two consecutive ROPA executions [min]
<i>Model Predictive Controller (MPC) implementation in case study 1</i>	
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	linearized matrices of the MPC model
—	whenever a bar is used above symbols, it indicates the values are computed as deviation from the nominal point
$\mathbf{U}^{m_c}$	control input sequence
m_c	input prediction horizon
p_c	output prediction horizon
Q_y, R_y	MPC objective function weights
$\bar{\mathbf{u}}^L$	input (as deviation from the nominal value) lower bounds
$\bar{\mathbf{u}}^U$	input (as deviation from the nominal value) upper bounds
$\Delta \bar{\mathbf{u}}^{max}$	maximum input variation

Chapter 4: Fluid Catalytic Cracking unit

C_{cat}	catalytic carbon concentration [wt%]
H_{ra}	reactor catalyst holdup [ton]
R_{rc}	catalyst recirculation rate [ton/min]
R_{cf}	catalytic carbon formation [Mlb/hr]
C_{sc}	total carbon concentration on spent catalyst [wt%]
C_{rc}	total carbon concentration on regenerated catalyst [wt%]
P_{ra}	reactor pressure [psia]
k_{cc}	pre-exponential factor of catalytic carbon formation reaction [Mlb/(hr psia ton wt% ^{-1.06})]
E_{cc}	activation energy of catalytic carbon formation reaction [BTU/lbmol]
R_g	gas constant [BTU/(lbmol °R)]
T_{ra}	reactor temperature [°F]
S_c	specific heat of catalyst [Btu/(lb °F)]
T_{rg}	regenerator temperature [°F]
S_f	specific heat of feed [Btu/(lb °F)]
D_{tf}	feed density [lb/gal]
R_{tf}	feed flowrate [Mbbl/day]
T_{fp}	feed temperature [°F]
ΔH_{fv}	heat of feed vaporization [Btu/lb]
ΔH_{cr}	heat of cracking [Btu/lb]
R_{or}	feed (gasoil) cracking rate [Mlb/hr]
k_{cr}	pre-exponential factor of carbon cracking reaction [Mbbl/(day psia ton wt% ^{-1.15})]

E_{cr}	activation energy of carbon cracking reaction [BTU/lbmol]
H_{rg}	regenerator catalyst holdup [ton]
R_{cb}	coke burning rate [Mlb/hr]
R_{ai}	air flowrate [Mlb/hr]
C_1	semi-empirical constant for the coke burning rate [lb oxygen/lb carbon]
O_{fg}	oxygen concentration [mol%]
P_{rg}	regenerator pressure [psia]
C_2	semi-empirical constant for oxygen concentration [Mlb oxygen/(hr psia ton cat.)(hr/Mlb) ²]
K_{cr}	oxygen formation coefficient [Mbbl oxygen/(hr psia ton cat.)]
k_{or}	pre-exponential factor of oxygen formation [Mlb/(hr psia ton)]
E_{or}	activation energy of oxygen formation [BTU/lbmol]
S_a	specific heat of air [Btu/(lb °F)]
T_{ai}	air temperature [°F]
ΔH_{rg}	heat of carbon burning reaction [Btu/lb]
C_4	semi-empirical constant for flue gas temperature estimation, [°F/ vol%]
T_{fg}	flue gas temperature [°F]
$GLNV$	gasoline volumetric yield [vol%]
F_1, F_2	auxiliary variables to calculate gasoline yield [-]
FSF	feed characterization parameter [-]
$CONVV$	coke volumetric conversion [vol%]
SEV	cracking reaction severity [%]
A_{est}	estimated value of the cracking reaction rate [-]

$catoil$ catalyst/feed mass flow relation [-]

Chapter 5: Column reactor system

$x_A(i)$ liquid molar fractions of A at the i th column stage [-]

$y_A(i)$ vapor molar fractions of A at the i th column stage [-]

α relative volatility between A and B [-]

$F_{v,i}$ vapor molar flow leaving stage i [mol/min]

$F_{l,i}$ liquid molar flow leaving stage i [mol/min]

F_{fc} column feed molar flow [mol/min]

F_d distillate (top product) molar flow [mol/min]

F_b bottom product molar flow [mol/min]

Q_v volumetric flow of vapor leaving the reboiler (stage 1) [L/min]

Q_d distillate volumetric flow [L/min]

Q_{fc} column feed volumetric flow [L/min]

ρ liquid molar density [mol/L]

ρ_{gas} gas molar density [mol/L]

$x_{A,feed}$ liquid molar fraction of the feed [-]

$m_c(i)$ liquid molar holdup at the i th column stage [mol]

t time [min]

Q_{out} reactor outlet volumetric flow [L/min]

Q_{fr} reactor feed (pure C) volumetric flow [L/min]

C_A concentration of A in the reactor [mol/L]

C_C concentration of C in the reactor [mol/L]

C_D concentration of D in the reactor [mol/L]

C_E	concentration of E in the reactor [mol/L]
V_R	reactor volume [L]
r_1	reaction rate of reaction 1 [mol/(L min)]
r_2	reaction rate of reaction 2 [mol/(L min)]
k_1	rate constant of reaction 1 [L/(mol min)]
k_2	rate constant of reaction 2 [L/(mol min)]

Contents

1	Introduction and literature review	22
1.1	Economic optimization in the process industry	22
1.2	Real-time optimization	23
1.3	Frequent implementation issues	25
1.4	Motivation and objectives	27
1.5	Thesis overview	29
1.6	Publication list	30
2	Real-time optimization with persistent adaptation	32
2.1	Fundamentals	32
2.2	Bibliographic review	33
2.3	ROPA and standard RTO comparison	36
2.4	Conclusion	45
3	Case study 1: Williams-Otto reactor	47
3.1	Process description	47
3.2	Process model	47
3.3	Case study	50
3.4	Results and discussion	55
3.5	Conclusion	63
4	Case study 2: Fluid catalytic cracking unit	64
4.1	Process description	64
4.2	Process model	66
4.3	Case study	67
4.4	Results and discussion	70
4.5	Conclusion	74
5	Case study 3: Plant-wide optimization	76
5.1	Process description	76
5.2	Process model	78
5.3	Case study	82

5.4	Results and discussion	84
5.5	Conclusion	90
6	Concluding remarks and future work recommen- dations	92
6.1	Complementary work	93
6.2	Directions for further work	94
	Bibliography	96
	APPENDICES	105
	APPENDIX A–Fluid catalyst cracking model and validation	106
	APPENDIX B–Dynamic design using proportional action (DDPA)	116
	APPENDIX C–Online model structure adapta- tion via modifier adaptation . . .	124

1 Introduction and literature review

This chapter describes the current practice of Real-Time Optimization (RTO) and discusses the common issues and vulnerabilities associated with the RTO structure. Given the challenges to successful RTO implementations, the motivation and objectives of the thesis are introduced. Next, related work developed during the doctoral studies and a list of publications are presented.

1.1 Economic optimization in the process industry

Economic optimization has always been an issue for the process industry due to the complex dynamics of the process systems. The economic optimization problem is further affected by strong interactions between system variables and, also, by unpredictable market disturbances, such as demand and/or supply fluctuation (MIZOGUCHI; MARLIN; HRYMAK, 1995).

In the early 60s, to optimize a plant economically, the process industry applied concepts of optimal control in process automation (PONTRYAGIN et al., 1962), without satisfactory results (MAGALHÃES, 2010). In the 70s, advanced control strategies appeared in the literature, such as model predictive controllers (MPC). They were developed in order to meet specific control needs of power plants and petroleum refineries, like: maintaining the process at specific targets in the presence of disturbances, and coping with the multi-dimensional aspects of processes operating at constraints (QIN; BADGWELL, 2003). Both features of the advanced controllers allowed the economical optimization of chemical process in real time (CUTLER; PERRY, 1983). Since the 90s, several publications have provided a good overview of theoretical and practical issues related to Real-Time Optimization (RTO) applications, such as stability, numerical solution, implementation drawbacks, etc., spanning a wide variety of processes (e.g. olefin units (YOON; DASGUPTA; MIJARES, 1996; GEORGIOU et al., 1998; LEGGETT; TEETER; STOUT, 2000), ethylene production (ELICECHE et al., 1995; VETTENRANTA et al., 2006; WANG et al., 2011), fluidized catalytic cracking units (MCFARLANE et al., 1993; LID; STRAND, 1997; ELLIS; LI; RIGGS, 1998; SILDIR et al., 2013), gasoline production units (BADER; GUESNEUX, 2007), refineries (Van WIJK; POPE, 1992), bio ethanol production (OCHOA; REPKE; WOZNY, 2010), polymerization reactions (HELBIG; ABEL; MARQUARDT, 1998; KIPARISSIDES et al., 2002), offshore gas production

(WESTLAKE, 1998; CRAMER; PETERS, 2002), metallurgical industry (WOODWARD et al., 2007), and paper industry (MERCANGÖZ; DOYLE, 2008)).

In the current practice, RTO is part of the plant decision hierarchy. This structure, which has no claim of optimality (DARBY et al., 2011), is summarized in many articles (CUTLER; PERRY, 1983; DARBY et al., 2011; QUELHAS; JESUS; PINTO, 2013) and is shown in Figure 1. Briefly, in the hierarchy: the **planning** layer determines the economic forecast and production goals. **Scheduling** focuses on the timing of actions to execute the chosen plan. The **RTO** and **MPC** layers implement the decisions of the upper layers in real-time. The lowest layer, the **distributed control system** (DCS), is responsible for keeping the plant at the desired setpoints and rejecting short term disturbances.

Figure 1 – Plant decision hierarchy (Source: Darby et al. (2011))

1.2 Real-time optimization

The RTO layer is responsible for implementing business decision in the operation. On the other hand, RTO does not manipulate any dynamic degrees of freedom (HUANG, 2010), which is the responsibility of the controllers. The standard RTO scheme, also known as Model Parameter Adaptation (MPA), relies on the basic assumption that the model and disturbance transients can be neglected by the controller, which is able to reach and maintain steady-state between the optimizer executions (ADETOLA, 2008). The standard RTO scheme diagram can be seen in Figure 2.

MPA core tasks are organized in two different optimization layers (QUELHAS; JESUS; PINTO, 2013); the first layer (in Figure 2: *Model update* block) uses process information to update the model parameters of a nonlinear rigorous plant model. In most RTO

Figure 2 – Block diagram of the standard RTO scheme, called model parameter adaptation (MPA) (Source: own elaboration)

applications, the rigorous model includes detailed component mass and energy balances as well as rigorous thermodynamic relations and kinetic reaction schemes (DARBY et al., 2011). The model parameters are adapted in order to improve model accuracy. Plant information, which is received as plant measurements, is used to estimate the model parameters or a subset of them. The idea here is to represent the current operating point (QUELHAS; JESUS; PINTO, 2013).

After adapting the model, the second optimization layer uses the updated rigorous steady state model to solve an economic-based optimization problem (in Figure 2: *Steady state optimization* block). Typically, the objective function of the economic optimization is profit (product revenues *minus* feed costs *minus* variable costs) (DARBY et al., 2011).

In addition to the MPA core tasks, steady state detection and model reconciliation can also be performed. The former is necessary due to the fact that RTO relies on steady state models. Hence, if the model adaptation routine updates the parameters of the stationary model directly with transient data, it is unlikely that the model predictions are accurate, affecting the optimality of the RTO solution (MARCHETTI; FERRAMOSCA; GONZÁLEZ, 2014). The detection of plant stationary periods (*Steady state detection* block in Figure 2) is accomplished by monitoring key process variables. There are several ways to detect steady-state reported in the literature, e.g. F-like test (CAO; RHINEHART, 1995; CAO; RHINEHART, 1997; SHROWTI; VILANKAR; RHINEHART, 2010); Wavelet transform

(JIANG et al., 2003; KORBE et al., 2014; JIANG; CHEN; HE, 2000); Polynomial-based (LE ROUX et al., 2008); and ARX-based (RINCÓN; LE ROUX; LIMA, 2015).

Typically, the choice of the SS detection method varies with the process characteristics and the practitioner experience. Once the steady-state condition is detected, the process data are submitted to gross error detection routines in order to remove errors from instrument malfunction (BESL et al., 1998). Afterwards, data reconciliation is performed (*Reconciliation* block in Figure 2). The goal is to correct the model inputs and outputs, adjusting the measurements to satisfy mass and energy balances (BESL et al., 1998). Note that, reconciliation and gross error detection are not carried out in some applications, depending on the process characteristics and measurement quality.

1.3 Frequent implementation issues

In the beginning of the decade, there were around 300 reported RTO implementations on chemical and petrochemical industry (DARBY et al., 2011). An RTO implementation is justified by: its capacity to identify an unconstrained optimum. Even in cases that the optimum lies at the constraints, RTO is able to track changes on the optimal active constraint set; Depending on the status of the manipulated variables of the controller, RTO can identify changes on the degrees of freedom of the optimization problem; day-to-night variations in the operation conditions; and changing in economics, like product and feedstock prices (DARBY et al., 2011).

However, despite the possible benefits, there are several issues involved in implementing RTO. Mainly, these drawbacks come from the plant decision hierarchy and the standard RTO structure shown in Figures 1 and 2, respectively. Therefore, they are inherent to the concept that defines standard RTO implementations.

Difficulties with the hierarchical structure mainly include the coordination between layers in order to avoid inconsistency and conflict. For example: despite determining the optimum operating point of the plant, RTO does not manipulate any variables of the process as mentioned. The RTO optimal solution is passed as setpoints/targets changes to the advanced control layer (or directly to the DCS), which is responsible for keeping the process at the desired operating point. Hence, there is a possibility that one or more of the setpoints/targets cannot be implemented due to model-mismatch between RTO and the controller (models in the layers have different steady-state gains, for example) (HUANG;

BIEGLER; PATWARDHAN, 2010). Furthermore, the objective functions of the controller and RTO layers are different, regulation/tracking for the controller *vs.* economic-based cost for RTO. Due to the fact that the two layers do not possess the same information, conflict between them might happen and the predicted operational point may be suboptimal (ADETOLA, 2008).

In addition to the issues connected to the hierarchical structure, there are several technical problems related to the standard RTO implementation, like: updating steady-state models to a dynamic plant; the steady state wait; and determining RTO scope and intermediary stream prices. The three highlighted points are discussed in details because they are the main focus of the thesis. For a complete discussion of the multiple challenges regarding RTO implementations, please refer to Friedman (1995), Darby et al. (2011), Quelhas, Jesus and Pinto (2013).

Updating steady-state models to a dynamic plant

In the standard RTO implementation (Model Parameter Adaptation - MPA), the representation of the current plant state in the model is achieved by parameter estimation. Hence, this step has a direct effect on the solution computed by the optimization layer. A poor estimate influences the model accuracy and the closeness of the RTO solution to the true optimum (ENGELL, 2007). As the standard RTO relies on steady state models, the parameters should only be estimated when the plant is at steady-state.

However, depending on the process or on the nature of the disturbances affecting the process, the steady-state periods are very difficult to detect (or even nonexistent). As mentioned earlier, SS detection methods rely on statistical or heuristic tests to determine if the process has reached steady-state. However, even after meeting the methods' criteria, it is hard to determine if the process is actually at steady-state (RHINEHART, 2013).

The SS detection problem gets even more complex if the size of the process increases. In these cases, choosing a set of measurements that represents the plant state is not trivial. Moreover, the measurement set can contain signals with slower responses, which are not in phase with other measurements, misleading the steady-state detection (HUANG; HARINATH; BIEGLER, 2011; RHINEHART, 2013). Additionally, if the period between two disturbances is lower than the process time constants and/or disturbances are constantly affecting a small

section of the process, the global set of measurements is unlikely to pass a steady-state acceptance test.

The steady-state waiting

Even in cases where the steady-state can be perfectly forecast, there is the inherent delay caused by the implementation of the RTO steady-state solution. For example, the optimal solution is calculated at time k . During the implementation phase, there is no guarantee that the previously updated model is still in phase with the plant. The optimal operation point may have changed due to new disturbances affecting the plant. Hence, in this case, when the optimal solution of time k is implemented in the plant, it would not reflect the optimal operation of the plant at the current time instant. Clearly, implementing the optimal solution in several small steps can mitigate this effect, however the inherent steady-state delay remains (FRIEDMAN, 1995).

Determining RTO scope

The best case scenario is that the RTO scope covers the entire unit. However, due to the modeling efforts and maintenance aspects of such large-scale models, the common approach is to implement RTO in specific units/sections of the plant (DARBY et al., 2011). Nevertheless, optimizing a single unit or subsystem does not guarantee that the plant-wide optimum is reached.

For steady-state optimization, it is possible to take advantage of decomposition techniques in order to have a single RTO implementation to a specific section of the plant. They allow a unit to be optimized in isolation and, even in these cases, the optimization results reflect the plant-wide optimum (FRIEDMAN, 1995). Despite appealing, the problem is shifted to pricing and determining the composition of the intermediary streams, which are challenging tasks to be carried out (FRIEDMAN, 1995; ACEVEDO et al., 2015).

1.4 Motivation and objectives

The thesis main contribution is the development of an alternative RTO scheme called Real-time Optimization with Persistent Adaptation (ROPA), which integrates

on-line parameter estimation in the optimization cycle. As a result, ROPA avoids the inherent steady-state delay in the RTO cycle and the problems related to updating steady-state models to a dynamic plant. Essentially, in the ROPA cycle, the parameters are estimated with transient data via online estimators. Then, the parameters are applied in the steady-state model, which is used to calculate the economic optimum setpoints for the plant. Despite not being optimal in a strict sense, these setpoints re-calculations represent a continuous improvement in operation. Thus, ROPA concept involves a persistent pursuit of a steady-state optimum, which most likely will be reached in a broader time horizon. ROPA provides an intermediary solution between static and dynamic optimization schemes. On one hand, it constantly improves plant operations (approximating the dynamic optimal trajectory). On the other hand, as ROPA economic optimization step is based on steady-state models, it relies on the well-established software and literature for SS optimization.

Moreover, ROPA is the key for decoupling the estimation problem in order to achieve plant-wide optimization. For example, a steady-state model can be developed for the entire unit, and then partitioned, allowing individual sections of the model to be updated asynchronously using either online or steady-state estimation. The choice of the estimation method depends on the characteristics of each individual section, like disturbance frequency. As the model is continually updated to the plant, the plant-wide economic optimization can be scheduled at any desired rate.

The new methodology benefits are illustrated with three case studies: (1) a Williams-Otto reactor, which is used for showing the convergence of ROPA to the steady-state optimum; (2) a Fluid Catalytic Cracking (FCC) unit, in which ROPA is applied in order to show some specific advantages of the method and how it behaves with larger, more complex processes. The FCC model used in the case study is compared and validated with actual data from the FCC located at Refinaria Henrique Lage (REVAP), São José dos Campos, Brazil and owned by Petrobras S.A.; and (3) a system composed by a distillation column and a reactor that is used to study the use of ROPA to obtain the plant-wide optimum.

In the three case studies, ROPA performance is compared to the standard RTO method, Model Parameter Adaptation (MPA). The comparison is based on profit loss and deviation from the actual optimal decisions. The results show that the increase of the model updating frequency (i.e. decrease of the time between two sequential optimization

runs) improves the prediction capacity of the optimization cycle, which leads to a better economic performance by enhancing the disturbance detection of the optimization cycle. Furthermore, by applying ROPA to subsections of the model, the plant-wide optimum is successfully obtained without using decomposition techniques.

1.5 Thesis overview

The thesis is organized as follows; In Chapter 2, both ROPA and MPA implementations are discussed in details. The idea is to show the new method advantages and compare it to the current practice. In addition, the notation used throughout the thesis is introduced. Afterwards, the three case studies are presented: Chapter 3 discusses the Williams-Otto reactor; Chapter 4 focuses on the Fluid Catalytic Cracking process. The FCC model equations and the comparison to actual plant data are presented in Appendix A; and in Chapter 5, the advantages of using ROPA to achieve the plant-wide optimum are explained and illustrated; Finally, Chapter 6 concludes the thesis. Also, it discusses possible directions for further works.

Complementary work

In addition to the main work, two additional topics were addressed during the thesis development. Each of these works has its own objectives, contributions and is structured to function as an individual paper. Both are briefly discussed in Appendices B and C.

Appendix B provides a new methodology, called Dynamic Design using Proportional Action (DDPA), that simplifies the dynamics of vapor-liquid equilibrium systems, specifically distillation column trays. The relation between pressure/vapor flowrate and molar holdup/liquid flowrate are modeled as simple linear relations similar to a proportional controller with arbitrarily large gain. Despite the simplifying hypotheses, DDPA provides an appropriate degree of accuracy in order to simulate high purity distillation columns, specially because it allows the use of rigorous thermodynamic packages in equation-oriented simulation environments. A high-purity distillation column is modeled using DDPA and the model is validated with real process data obtained in an industrial-scale depropenizer column located at Refinaria de Paulínia, Paulínia, São Paulo, Brazil and owned by Petrobras S.A..

Appendix C devises a method for selecting the model structure online, allowing the model to evolve over time while running the optimization cycle. The idea is to propose an online methodology that chooses, among a population of knowledge-based models, the model structure that is more consistent with the current available data. Moreover, the methodology is applied in combination with Modifier Adaptation methods, which allows that the optimum computed by the model-based optimization problem matches the plant optimum upon convergence. By allowing the model structure to change over time, the model is able to adapt its structure and complexity to the experimental data. The methodology was tested in a gas lifted oil well network. The results show that, despite of a significant amount of noise, the correct model structure is chosen among several models structure candidates and the plant optimum is reached without constraint violations.

1.6 Publication list

Journal Paper

- **2018 - Journal to be defined.** MATIAS J.O.A., Le ROUX G.A.C. and JÄSCHKE, J.. Journal Paper (submission planned): *Online model structure adaptation via modifier adaptation*
- **2018 - Journal of Process Control.** MATIAS J.O.A. and Le ROUX G.A.C.. Journal Paper: *Real-time Optimization with Persistent Parameter Adaptation Using Online Parameter Estimation*
- **2017 - Industrial & Engineering Chemistry Research.** MATIAS J.O.A., EBRAHIMPOUR M. and Le ROUX G.A.C.. Journal Paper: *Development and Validation of a Reduced-Index Dynamic Model of an Industrial High-Purity Column*

Conferences

- **2018 - COBEQ - Brazilian Congress of Chemical Engineering - São Paulo /Brazil.** MATIAS J.O.A. and Le ROUX G.A.C.. Extended Abstract (accepted): *Otimização em Tempo Real com Adaptação Persistente de Parâmetros Aplicada a Modelo de Unidade de Craqueamento* (Real-time Optimization with Persistent Parameter Adaptation Applied to a Catalytic Cracking Unit Model)

- **2018 - INTPART Brazil-Norway Operation Subsea Workshop - Trondheim/Norway.** MATIAS J.O.A., Le ROUX G.A.C. and JÄSCHKE, J.. Talk at technical section: *Online model structure adaptation via modifier adaptation*
- **2018 - IFAC-OOGP - Automatic Control Conference in Offshore Oil and Gas Production - Esbjerg/Denmark.** MATIAS J.O.A., Le ROUX G.A.C. and JÄSCHKE, J.. Extended Abstract: *Modifier Adaptation for Real-time Optimization of a Gas Lifted Well*
- **2016 - COBEQ - Brazilian Congress of Chemical Engineering - Fortaleza/Brazil.** MATIAS J.O.A., EBRAHIMPOUR M., LIPORACE, F.S. and Le ROUX G.A.C.. Extended Abstract: *Desenvolvimento e Validação de Modelo Dinâmico de Índice Reduzido de uma Coluna de Destilação Industrial de Alta Pureza* (Development and Validation of a Reduced Index Dynamic Model of a High Purity Distillation Column)
- **2016 - COBEQ - Brazilian Congress of Chemical Engineering - Fortaleza/Brazil.** MARTINEZ E.E., GRACIANO J.E.A., LIZA J.C., MATIAS J.O.A., LIPORACE, F.S. and Le ROUX G.A.C.. Extended Abstract: *Desenvolvimento de Software Baseado em Agentes para Implementação de Otimização em Tempo Real* (Development of an Agent Based Software for RTO Implementation)
- **2015 - AIChE's Conference - Salt Lake City/USA.** MATIAS J.O.A., EBRAHIMPOUR M., LIPORACE, F.S., VIANNA Jr. A.S. and Le ROUX G.A.C.. Talk at technical session: *Development and Validation of a Reduced-Index Dynamic Model of an Industrial High Purity Column*
- **2014 - COBEQ - Brazilian Congress of Chemical Engineering - Florianopolis/Brazil.** EBRAHIMPOUR M., MATIAS J.O.A., GRACIANO J.E.A., MARTINEZ E.E., PENA A.M.A., RINCON F.D., MENDOZA D.F., QUELHAS A.D., SOARES R.DE P., LIPORACE F.S., ODLOAK D., VIANNA Jr. A.S. and Le ROUX G.A.C.. Extended Abstract: *Desenvolvimento de Protótipo RTO* (Development of an RTO prototype)

2 Real-time optimization with persistent adaptation

This chapter introduces the Real-time Optimization with Persistent Adaptation (ROPA). The focus is to formalize the new method basic concepts as well as to show the innovation behind ROPA. For this purpose, a brief overview of optimization methods in real time is presented. Then, the standard RTO and ROPA implementations are discussed and compared. The chapter ends with the formal assessment of the ROPA convergence properties.

2.1 Fundamentals

In Real-time optimization (RTO), the plant is optimized based on rigorous nonlinear models and economic-based objectives (MARCHETTI; FERRAMOSCA; GONZÁLEZ, 2014). In order to operate in the actual optimal condition, cost/price updates as well as process disturbances need to be taken into account during the optimization (FRIEDMAN, 1995).

In Model Parameter Adaptation (MPA), this task is achieved by updating the model parameters in order to reflect the current plant condition. As discussed previously, in MPA the model is updated only when the process reaches a stationary condition and detecting steady-states is a challenging task to be carried out (RHINEHART, 2013). Clearly, if the steady-state (SS) periods are poorly detected and the model is updated with erroneous information (i.e. the stationary model parameters are directly updated with transient data), the calculated optimal condition does not correspond to the actual plant optimum (MARCHETTI; FERRAMOSCA; GONZÁLEZ, 2014).

Additionally, there is the question if a unit can be optimized in isolation (FRIEDMAN, 1995). Optimizing a single unit is critical, not only by the difficulties in determining intermediate prices and compositions but also because the optimum of a single unit most likely does not reflect the plant-wide optimum (ACEVEDO et al., 2015). The ideal situation is to obtain a model for the complete process and, then, optimize it in order to achieve the plant optimum. In addition to the modeling and model maintenance issues, the SS detection problem is even more complex for large systems. The larger the size of the process, the more challenging is to detect if the process has reached the SS to trigger the model updating step in MPA. For example, if disturbances are constantly affecting a small

section of the process, the global set of measurements is unlikely to pass a steady-state acceptance test.

An alternative, proposed by Prior and Lopez (1999), is to develop a stationary model for the whole plant and then divide it in smaller subsections, which are update asynchronously, whenever they reach SS condition. This modular approach allows the whole system to be optimized at any desired rate, solving the plant-wide optimization problem. However, as the model updates still rely on stationary parameter estimation, this strategy does not surpass the SS detection problem. For example, the sections of the plant that are affected by frequent disturbances are prone to have a low frequency of parameter updates, which leads RTO to optimize an unsuited model, decreasing the potential benefits of the economic optimization (SEQUEIRA; GRAELLS; PUIGJANER, 2002).

A second alternative is to apply Real-time Optimization with Persistent Adaptation (ROPA) to the system. ROPA continuously updates the parameters of the steady-state model via online estimation methods, avoiding the SS wait. Note that the parameters are estimated in an online fashion but they are treated as if they represent a steady-state condition. Such configuration allows the optimizer to run continuously at any desired rate, restricted only by the process settling time and computational constraints. Furthermore, by applying ROPA to specific sections of the plant, the plant-wide estimation problem is decoupled, which allows the optimization of the entire plant. For example, by partitioning the plant model, online estimation can be applied to the subsections of the plant model that have low parameter update frequency (e.g. highly perturbed sections) and then the optimization problem can be solved for the complete model. This asynchronous model update allows all sections of the model to be updated to the plant and to always reflect the plant current state. Figure 3 compares ROPA with the standard RTO approach, Model Parameter Adaptation.

2.2 Bibliographic review

ROPA is in line with other methods that avoid the SS wait or reduce the sampling time of RTO, like: the inclusion of an intermediary optimization layer (YING; JOSEPH, 1999; GONZÁLEZ; ODLOAK, 2009); the frequent optimization without reestimating the parameters (BESL et al., 1998; PRIOR; LOPEZ, 1999); real time evolution (RTE) (SEQUEIRA;

Figure 3 – Comparison between standard RTO and the ROPA (Source: own elaboration)

GRAELLS; PUIGJANER, 2002); and modifier adaptation (MA) with transient gradient estimation (FRANÇOIS; BONVIN, 2013).

If compared to the methods above, the proposed methodology has three strengths. First, ROPA updates the parameters continuously, enabling the model to reflect the current plant state. The approaches that update the RTO solution without reestimating the parameters (frequent optimization and RTE) can optimize an unsuited model leading to aggressive and profitless updates of the setpoints, even with the recursive improvement provided by the RTE. Second, ROPA does not need plant gradient estimates, which are very challenging to obtain and represent the main bottleneck of MA methods (FRANÇOIS; BONVIN, 2013). Finally, if compared to the approach that includes an intermediary optimization layer, ROPA does not demand implementing advanced controllers, which may not be necessary or appropriate for some processes (SEQUEIRA; GRAELLS; PUIGJANER, 2002). Any control strategy can be used with ROPA.

More recently, in addition to the static optimization methods discussed above, there has been growing interest in schemes that compute optimal trajectories, like Dynamic Real Time Optimization (DRTO) and Economic MPC (EMPC), which are also alternatives to avoid the steady-state detection step.

Instead of calculating stationary setpoints like in static RTO, DRTO calculates a reference trajectory and solves the control problem simultaneously (HELBIG; ABEL; MARQUARDT, 2000). For large-scale application, this regulation-optimization problem is computationally expensive, even if the new solution is recalculated based on a previous horizon (BINDER et al., 2001). Modeling is also a very important step for the DRTO, the process model must have a minimum quality in its predictions and should cover a wide range of process dynamics (KADAM; MARQUARDT, 2007).

In EMPC (AMRIT; RAWLINGS; ANGELI, 2011; HUANG; HARINATH; BIEGLER, 2011; CHEN et al., 2012), the economic optimization and control problem are also solved simultaneously. However, in EMPC, the regulation-optimization problem is merged in a single layer by replacing the traditional quadratic cost of the MPC with a general cost function. Albeit this substitution improves closed-loop economic performance, the change in the standard cost function requires the redesign of the control architecture (ELLIS; CHRISTOFIDES, 2014). Also, it is necessary to apply dynamic optimization algorithms that are able to cope with the computational burden of the economic receding horizon control problem with reasonable time and satisfactory accuracy (GRÜNE, 2013).

Despite appealing, there are few examples of DRTO and EMPC implementations using large-scale rigorous first-principles models. The lack of large-scale models is a problem in cases involving complex processes composed by several units, like a refinery. Because optimizing a single unit or subsystem does not guarantee that the plant global optimum is reached (FRIEDMAN, 1995). Hence, there is a clear trade-off between potential benefits and current limitations of DRTO and EMPC.

In this context, ROPA appears as an intermediary solution between static RTO and DRTO/EMPC, exploiting the best characteristics of both methods. ROPA approximates the optimal trajectory by constantly improving plant operations, while relies on the well-established literature and software of stationary economic optimization, like ROMEo and Aspenplus (QUELHAS; JESUS; PINTO, 2013). The lower computational burden is also a benefit of this proposal.

In order to provide a complete assessment of the differences between the new proposal and the standard RTO, the next sections provide details of the ROPA and MPA implementations used in this thesis. After introducing some mathematical notation, the MPA steps of Figure 3a are discussed. Next, an overview of the online estimation methods that can be applied in the ROPA cycle is shown and the online estimators used in this

thesis (extended Kalman filter (EKF) and reduced extended Kalman filter (rEKF)) are presented in details. Finally, the convergence conditions for the ROPA cycle are assessed formally.

2.3 ROPA and standard RTO comparison

Mathematical preliminaries

The plant is represented by the following steady-state input-output mapping:

$$\mathbf{y}_{p,k}(\mathbf{u}_k, \mathbf{d}_{p,k}, \boldsymbol{\epsilon}_{p,k}) \in \mathbb{R}^{n_y} \quad (1)$$

in which $\mathbf{u}_k \in \mathbb{R}^{n_u}$ are the system inputs, $\mathbf{d}_{p,k} \in \mathbb{R}^{n_d}$ are the deterministic disturbances, and $\boldsymbol{\epsilon}_{p,k} \in \mathbb{R}^{n_n}$ the random disturbances. The subscript k indicates the variable at time t_k assuming a zero-order holder over the interval $[t_k, t_{k+1})$.

Steady-state and dynamic models are developed for the RTO and the online estimation layers. They are associated with the subscripts ss and dyn , respectively. The steady state model is:

$$\begin{aligned} \mathbf{0} &= \mathbf{f}_{ss}(\mathbf{x}, \mathbf{u}, \mathbf{p}) \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{p}) + \boldsymbol{\nu}_{ss} \end{aligned} \quad (2)$$

where, in addition to the previous notation, $\mathbf{x} \in \mathbb{R}^{n_x}$ are the model state variables and $\mathbf{p} \in \mathbb{R}^{n_p}$ is the set of model parameters. $\mathbf{y} \in \mathbb{R}^{n_y}$ are the model outputs, which coincide with the plant measurements in this case. $\mathbf{f}_{ss} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$ is a nonlinear function. $\mathbf{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_y}$ is the output function, which is also nonlinear. The lack of time subscripts indicates steady-state values. The measurement noise, $\boldsymbol{\nu}_{ss}$, is added to the model outputs and is modeled as white Gaussian random noise with zero mean and constant covariance matrices $\mathbf{R}_p \in \mathbb{R}^{n_y, n_y}$. The dynamic model is represented by:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{f}_{dyn}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{p}_k) + \boldsymbol{\omega}_k \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k, \mathbf{p}_k) + \boldsymbol{\nu}_k \end{aligned} \quad (3)$$

In the dynamic case, process and measurement noise, $\boldsymbol{\omega}_k$ and $\boldsymbol{\nu}_k$, are added to the process model. Both are modeled as white Gaussian random noises with zero mean and constant covariance matrices $\mathbf{Q} \in \mathbb{R}^{n_x, n_x}$ and $\mathbf{R} \in \mathbb{R}^{n_y, n_y}$, respectively. Also, the state transition function $\mathbf{f}_{dyn}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{p}_k)$ is a mapping over the interval $[t_k, t_{k+1})$, which represents the solution of the differential model during this period. $\mathbf{f}_{dyn}$ is assumed to be

at least once differentiable in all points of the valid operation range. Moreover, $\mathbf{f}_{dyn}$ has the same dimension as $\mathbf{f}_{ss}$ and both share the same states.

Model parameter adaptation (MPA)

MPA, which is the standard RTO approach (QUELHAS; JESUS; PINTO, 2013), is implemented in the case studies in order to be used as the basis of comparison between the current RTO practice with ROPA. Although the MPA cycle can have more elements, like gross error detection, only the three steps shown in Figure 3a are implemented in the case studies, namely: Steady-state detection; model adaptation (parameter estimation); and steady-state optimization.

Steady-state detection

The steady-state detection step is based on Rhinehart (2013), which estimates the measurements variance by two methods and then compare them in order to determine if the system is at steady-state. The first method calculates an estimate of the variance between the current measurement and a filtered trend of the same measurement (z_k and $z_{f,k-1}$). In turn, the second method calculates the variance between sequential data measurements (z_k and z_{k-1}). Whenever the process reaches a steady state, the variances calculated by both methods are ideally equal to each other. Therefore, if the process is not at steady state, the variance ratio (R_{SS}) is significantly larger than unity. The equations of the SS detection method are:

$$\begin{aligned} z_{f,k} &= \lambda_1 z_k + (1 - \lambda_1) z_{f,k-1} \\ \delta_{1,f,k}^2 &= \lambda_2 (z_k - z_{f,k-1})^2 + (1 - \lambda_2) \delta_{1,f,k-1}^2 \\ \delta_{2,f,k}^2 &= \lambda_3 (z_k - z_{k-1})^2 + (1 - \lambda_3) \delta_{2,f,k-1}^2 \\ R_{SS} &= (2 - \lambda_1) \delta_{1,f,k}^2 / \delta_{2,f,k}^2 \end{aligned} \tag{4}$$

in which, z_k is the given measured variable ($z_k \in \mathbf{y}_{p,k}$), at time t_k ; z_f is its filtered value; $\delta_{1,f}^2$ is the filtered variance calculated by the first method, and $\delta_{2,f}^2$ by the second; and R_{SS} is the ratio of variances. If R_{SS} is larger than a given threshold (R_{crit}), the measurement fails the SS detection test. In the case where two or more measurements are used to indicate the SS, all of them need to pass the test in order to indicate that the process is at

steady-state. All the filtered values are obtained with a conventional first-order filter in which λ_i with $i = \{1, 2, 3\}$ are the filter factors.

Model adaptation

Once the plant has reached steady state, the model adaptation algorithm starts to run. The parameters are estimated by minimizing a weighted sum of squared errors between the measured and predicted outputs (CHACHUAT; SRINIVASAN; BONVIN, 2009).

$$\begin{aligned} \mathbf{p}^{adj,*} = \underset{\mathbf{p}^{adj}}{\operatorname{argmin}} \quad & \phi_{adap} := \|\mathbf{y}_p - \mathbf{y}\|_{\mathbf{R}_p}^2 \\ \text{s. t. } G_{adap} := \quad & \begin{bmatrix} \mathbf{0} = \mathbf{f}_{ss}(\mathbf{x}, \mathbf{u}, [\mathbf{p}^{nom}, \mathbf{p}^{adj}]^T) \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, [\mathbf{p}^{nom}, \mathbf{p}^{adj}]^T) \\ \mathbf{p}^{L,adj} \leq \mathbf{p}^{adj} \leq \mathbf{p}^{U,adj} \end{bmatrix} \end{aligned} \quad (5)$$

in which, $\mathbf{p}^{adj}$ are the adjustable parameters, which is a subset of the model parameters. The parameters outside $\mathbf{p}^{adj}$ set remain in their nominal value $\mathbf{p}^{nom}$. ϕ_{adap} is the adaptation problem objective function, and $\mathbf{R}_p$ are the weights for squared error function. G_{adap} are the constraints of the model adaptation problem. It consists of the model equations and adjustable parameter bounds, $\mathbf{p}^{L,adj}$ and $\mathbf{p}^{U,adj}$. The other variables were previously defined in Section 2.3. The values of inputs $\mathbf{u}$ and outputs $\mathbf{y}_p$ are obtained from plant data. Note that the steady-state model adaptation described here is used only for the MPA cycle. ROPA uses on-line estimators for the model adaptation step.

Steady-state optimization

After the steady-state plant model is updated with the estimated parameters ($\mathbf{p}^{adj,*}$), an economic-based optimization problem is solved:

$$\begin{aligned} \mathbf{u}^* = \underset{\mathbf{u}}{\operatorname{argmin}} \quad & \phi_{profit}(\mathbf{u}, \mathbf{y}) \\ \text{s. t. } G := \quad & \begin{bmatrix} \mathbf{0} = \mathbf{f}_{ss}(\mathbf{x}, \mathbf{u}, [\mathbf{p}^{nom}, \mathbf{p}^{adj,*}]^T) \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, [\mathbf{p}^{nom}, \mathbf{p}^{adj,*}]^T) \\ \mathbf{u} \in \mathcal{U} \end{bmatrix} \end{aligned} \quad (6)$$

where, ϕ_{profit} is a scalar economic objective to be minimized and G is the set of constraints of the optimization problem. G is composed by the model equations and operational

inequality constraints, which are defined by $\mathcal{U} = \{\mathbf{u} \in \mathbb{R}^{n_u} : \mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U \wedge \mathbf{y}^L \leq \mathbf{y} \leq \mathbf{y}^U\}$. Note that $\mathbf{y}$ are a function of $\mathbf{u}$, therefore the constraints are condensed in only one set. $\mathbf{y}^L$, $\mathbf{y}^U$, $\mathbf{u}^L$ and $\mathbf{u}^U$ are the lower and upper bounds of the outputs and inputs. Note that, despite the fact that the optimization problem is formalized in a way that it seeks to find the minimum of the objective function, in practice the problem is set in such way that it maximizes the objective function, which is the unit profit.

After calculating the optimal $\mathbf{u}^*$, the corresponding optimal $\mathbf{y}^*$ are obtained with the SS model. Then, $\mathbf{y}^*$ are used as set-points for the controller, which implements the optimal decision in the process.

Real-time optimization with persistent adaptation

The ROPA cycle, which was shown in Figure 3b, integrates online estimators into the RTO cycle. The main difference lies on the fact that $\mathbf{p}^{adj,*}$ (in the economic-based optimization problem presented in Equation (6)) does not come from the model adaptation problem, but is computed by the online estimators using transient measurements. Therefore, ROPA does not depend on detecting the plant steady-state and, as a consequence, the SS optimization problem can be triggered at an arbitrary time. The period between two consecutive ROPA executions is referred as Δt_{ROPA} .

Despite the fact that the parameters are estimated with transient measurements, they are treated as if they represent a steady-state condition. The fundamental concept of ROPA is based on the fact that the methodology does not seek continuous optimization, in a strict sense, rather it seeks to continuously improve the economic optimal decisions. By recomputing the SS optimal solution, ROPA approximates the optimal trajectory and drives the plant to its optimum steady-state. Thus, ROPA involves a persistent pursuit of an steady-state optimum, which most likely will be reached in a broader time horizon.

Note that applying dynamic models in the estimation step may increase the number of estimated parameters in comparison to SS models (e.g.: mass holdup of trays). Therefore, there is a trade-off between using dynamic models for estimation and the benefits that they can bring when included in the optimization cycle. These trade-offs of the ROPA methodology should be taken into account during the implementation phase. On the other hand, if ROPA is compared to methods that calculate optimal trajectories, in which both need dynamic models for the estimation phase, ROPA has a significant advantage. The

number of estimated parameters can be decreased due to the possibility of applying online estimation to smaller sections of the model in a partitioned model updating strategy, which is explained in further details in Chapter 5.

Online estimators

In order to avoid the problems of the steady-state wait, online estimation is implemented. The online estimator needs to infer the most likely state/parameter estimate based on a dynamic model (a priori information) and the available sensor measurements (RAWLINGS; BAKSHI, 2006). The applied estimators need to provide good estimates and to be robust to: process disturbance; high levels of noise; and modeling errors (LIMA; RAWLINGS, 2011).

If the system is linear and there are no constraints on the estimated values, the Kalman filter is the natural choice. However, this condition is seldom found in practical applications. The extended Kalman filter (EKF) appears as an alternative. It is based on linearization of the nonlinear plant model and has been efficiently applied for a variety of systems (GUDI; SHAH; GRAY, 1994; MIDDLEBROOKS, 2001; PRASAD et al., 2002). On one hand, EKF takes advantage of the Kalman filter's computational efficiency and recursive strategy. On the other hand, its performance when dealing with highly nonlinear systems estimator may have limitations (BIEGLER; YANG; FISCHER, 2015).

Despite the fact that the poor performance in face of nonlinear system is not a consensus (SCHNEIDER; GEORGAKIS, 2013), different proposals that potentially improve EKF performance include have been proposed. For example: modifications of the basic structure, like reduced rank EKF (rEKF) (PHAM; VERRON; ROUBAUD, 1998); *ad hoc* state clipping strategies (HASELTINE; RAWLINGS, 2005); and spawning related estimation methods, for example unscented Kalman filter (JULIER; UHLMANN, 2004).

Another class of estimators that can be applied is the non recursive estimators. For this class, the current states/parameters can be estimated by an optimization problem using a moving window of measurements. Examples are: Moving Horizon Estimation (MHE) (MUSKE; RAWLINGS, 1995; RAO; RAWLINGS; MAYNE, 2003), and Constrained Extended Kalman Filter (CEKF) (ENGELL, 2007; SALAU; TRIERWEILER; SECCHI, 2014). Both strategies allow the estimation for nonlinear models and the inclusion of inequality constraints.

Regarding the proposed methodology, ROPA, it is possible to apply all of the estimators presented above (or any other that the practitioner has experience with). The integration works with every estimator, given that it is able to provide good quality online estimates of the parameters and states.

Albeit the goal of the thesis is to present and discuss the integrated scheme, the evaluation of the performance of different on-line estimators in the optimization cycle is out of scope. The main objective is to show that the on-line estimators can be integrated in the optimization cycle and, as a result, they are able to avoid the SS wait of the RTO cycle. In the case studies, the extended Kalman filter (EKF) and the reduced extended Kalman filter (rEKF) are chosen as the online estimation method. Both are described in the following sections. In order to facilitate the understanding, the EKF implementation is described first and then, the rEKF equations are derived.

Extended Kalman filter (EKF)

To outline the extended Kalman filter, the dynamic model, Equation (3), is considered. In order to estimate both states and parameters, the dynamic model is rearranged by creating an augmented state, $\mathbf{x}_k^e = [\mathbf{x}_k, \mathbf{p}_k]^T$ (i.e. the parameters are considered as additional states). The augmented dynamic system is given by:

$$\begin{aligned} \mathbf{x}_{k+1}^e &= \begin{bmatrix} \mathbf{f}_{dyn}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{p}_k) \\ \mathbf{p}_k \end{bmatrix} + \begin{bmatrix} \boldsymbol{\omega}_k \\ \boldsymbol{\omega}_k^p \end{bmatrix} = \mathbf{f}^e(\mathbf{x}_k^e, \mathbf{u}_k) + \boldsymbol{\omega}_k^e \\ \mathbf{y}_k &= \mathbf{h}_k^e(\mathbf{x}_k^e) + \boldsymbol{\nu}_k \end{aligned} \quad (7)$$

in which, $\boldsymbol{\omega}_k^p \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^p)$. The next step is the system linearization for applying the Kalman filter equations. First-order expansions around the extended state estimate $\hat{\mathbf{x}}_{k|k}^e$ and $\hat{\mathbf{x}}_{k|k-1}^e$ are carried out. The subscript notation $a|b$ is used to represent an estimate at time a based on information available at time b .

$$\begin{aligned} \mathbf{f}^e(\mathbf{x}_k^e, \mathbf{u}_k) &\approx \mathbf{f}^e(\hat{\mathbf{x}}_{k|k}^e, \mathbf{u}_k) + \mathbf{F}_k(\mathbf{x}_k^e - \hat{\mathbf{x}}_{k|k}^e) \\ \mathbf{h}^e(\mathbf{x}_k^e) &\approx \mathbf{h}^e(\hat{\mathbf{x}}_{k|k-1}^e) + \mathbf{H}_k(\mathbf{x}_k^e - \hat{\mathbf{x}}_{k|k-1}^e) \end{aligned} \quad (8)$$

where,

$$\mathbf{F}_k = \left. \frac{\partial \mathbf{f}^e(\mathbf{x}_k^e, \mathbf{u}_k)}{\partial \mathbf{x}_k^{e,T}} \right|_{\hat{\mathbf{x}}_{k|k}^e} = \begin{bmatrix} \frac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{p}_k)}{\partial \mathbf{x}_k^T} & \frac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{p}_k)}{\partial \mathbf{p}_k^T} \\ \mathbf{0} & \mathbf{I}_{n_p} \end{bmatrix}_{\hat{\mathbf{x}}_{k|k}^e} \quad (9)$$

$$\mathbf{H}_k = \left. \frac{\partial \mathbf{f}^e(\mathbf{x}_k^e)}{\partial \mathbf{x}_k^{e,T}} \right|_{\hat{\mathbf{x}}_{k|k-1}^e} = \left[\frac{\partial \mathbf{h}(\mathbf{x}_k, \mathbf{p}_k)}{\partial \mathbf{x}_k^T} \quad \frac{\partial \mathbf{h}(\mathbf{x}_k, \mathbf{p}_k)}{\partial \mathbf{p}_k^T} \right]_{\hat{\mathbf{x}}_{k|k-1}^e} \quad (10)$$

In which, $\mathbf{I}$ indicates the identity matrix, with the subscript specifying the dimension. The partial derivatives of the state transition functions are obtained using the sensitivity equations and are calculated with SUNDIALS (HINDMARSH et al., 2005).

After the system is linearized, the Kalman filter equations can be directly applied. However, instead of using the linear approximation in the prediction step, the nonlinear model is used:

$$\hat{\mathbf{x}}_{k+1|k}^e = \mathbf{f}^e(\hat{\mathbf{x}}_{k|k}^e, \mathbf{u}_k) \quad (11)$$

The predict and update equations are the usual Kalman filter equations:

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^T + \mathbf{Q}^e \quad (12)$$

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1|k} \mathbf{H}_{k+1}^T [\mathbf{H}_{k+1} \mathbf{P}_{k+1|k} \mathbf{H}_{k+1}^T + \mathbf{R}]^{-1} \quad (13)$$

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1} \mathbf{H}_{k+1} \mathbf{P}_{k+1|k} \quad (14)$$

with,

$$\mathbf{Q}^e = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^p \end{bmatrix} \quad (15)$$

Finally, the extended states are estimated using the prediction error.

$$\hat{\mathbf{x}}_{k+1|k+1}^e = \hat{\mathbf{x}}_{k+1|k}^e + \mathbf{K}_{k+1} [\mathbf{y}_{k+1} - \mathbf{h}^e(\hat{\mathbf{x}}_{k+1|k}^e)] \quad (16)$$

Reduced extended Kalman filter (rEKF)

rEKF is an extension of the EKF methodology. The major difference lies in the prediction phase. Instead of updating the predicted estimate covariance matrix $\mathbf{P}_{k+1|k}$ as in Equation (12), this methodology uses an approximation of the estimate covariance matrix.

This variant of the EKF is specially useful if applied in systems with identifiability issues, which rise when distinct parameters map to the same system input/output pair

(MIAO et al., 2011). In cases that the parameters are computed as states, the observability and identifiability concepts are equivalent.

Hence, in such cases, it is not possible to identify the true parameter based on data alone. Furthermore, identifiability problems can result in ill-conditioning of the estimate covariance matrix causing poor predictions by the estimator and affecting the optimal solution of the RTO cycle (MARCHETTI; FERRAMOSCA; GONZÁLEZ, 2014; GRACIANO; MENDOZA; LE ROUX, 2014)

The rEKF implementation is based on Salau, Trierweiler and Secchi (2014). In this paper, the authors apply reduced extended Kalman filter in systems tending to multiple steady states. Although the paper does not discuss identifiability directly, rEKF can prevent an ill-conditioned estimate covariance matrix by discarding unreliable information.

In the rEKF implementation, after calculating the estimates covariance matrix with Equation (12), $\mathbf{P}_{k+1|k}$ is decomposed in terms of its eigenvalues and eigenvectors:

$$\begin{aligned} \mathbf{P}_{k+1|k} &= \mathbf{V}_{\mathbf{P},k} \mathbf{\Lambda}_{\mathbf{P},k} \mathbf{V}_{\mathbf{P},k}^T \\ &= \begin{bmatrix} | & & & | \\ \mathbf{v}_{\mathbf{P},1} & \mathbf{v}_{\mathbf{P},2} & \cdots & \mathbf{v}_{\mathbf{P},n_x+n_p} \\ | & & & | \end{bmatrix} \begin{bmatrix} \lambda_{\mathbf{P},1} & 0 & \cdots & 0 \\ 0 & \lambda_{\mathbf{P},2} & \cdots & 0 \\ . & . & \ddots & . \\ 0 & 0 & \cdots & \lambda_{\mathbf{P},n_x+n_p} \end{bmatrix} \begin{bmatrix} | & & & | \\ \mathbf{v}_{\mathbf{P},1} & \mathbf{v}_{\mathbf{P},2} & \cdots & \mathbf{v}_{\mathbf{P},n_x+n_p} \\ | & & & | \end{bmatrix}^T \end{aligned} \quad (17)$$

in which, $\mathbf{P}_{k+1|k}$ is a $(n_x + n_p) \times (n_x + n_p)$ square matrix. n_x and n_p are the dimensions of the states and parameters. $\mathbf{v}_{\mathbf{P},i}$ are the eigenvectors of the estimate covariance matrix and $\lambda_{\mathbf{P},i}$, the associated eigenvalues. As $\mathbf{P}_{k+1|k}$ is symmetric, the matrix multiplication is rearranged:

$$\mathbf{P}_{k+1|k} = \sum_{i=1}^{n_x+n_p} \lambda_{\mathbf{P},i} \mathbf{v}_{\mathbf{P},i} \mathbf{v}_{\mathbf{P},i}^T \quad (18)$$

Next, the eigenvalues $\lambda_{\mathbf{P},i}$ are sorted in ascending order. Only the eigenvectors whose associated eigenvalues are smaller than a threshold value, λ_{thres} , are used to calculate the estimate covariance matrix. Hence, only a lower-rank approximation of $\mathbf{P}_{k+1|k}$ is used to calculate the filter gain, Equation (13):

$$\mathbf{P}_{q,k+1|k} = \sum_{i=1}^q \lambda_{\mathbf{P},i} \mathbf{v}_{\mathbf{P},i} \mathbf{v}_{\mathbf{P},i}^T \quad (19)$$

in which, q is defined as:

$$\{q = \max\{0, 1, \dots, n_x + n_p\} : |\lambda_{\mathbf{P},q}| \leq \lambda_{thres}\} \quad (20)$$

The implementation in the case studies differs in two points from the implementation outlined in Salau, Trierweiler and Secchi (2014). The estimates covariance matrix, $\mathbf{P}_{k+1|k}$, not the lower-rank approximation, $\mathbf{P}_{q,k+1|k}$, is used in the updating phase, Equation (14). Because, during the estimation process, regions containing relevant information for the parameter/state estimation may be reached and this information should not be wasted. The second change is that the number of directions, q , used on the lower-order approximation is not fixed, it changes based on λ_{thres} .

ROPA convergence analysis

In order to outline the convergence conditions of the ROPA cycle, let us introduce the following assumptions regarding the system:

Assumption 1. *Stationary system properties in Equations (5) and (6)*

1. *for all feasible inputs $\mathbf{u} \in \mathcal{U}$, the steady-state non linear model has an unique solution for a given parametrization;*
2. *ϕ_{profit} and G are continuously differentiable for all $\mathbf{u} \in \mathcal{U}$. Also, these derivatives are bounded in the feasible space. Note that $\{\mathbf{f}_{ss}, \mathbf{h}\} \in G$;*
3. *the model is structurally identifiable. i.e. the mapping $\mathbf{p}^{adj} \mapsto \mathbf{y}$ is one-to-one;*
4. *there is no plant-model mismatch (GRACIANO; MENDOZA; LE ROUX, 2014), the model can be considered perfect;*
5. *$\mathbf{u}^*$ is a strict local minimum of the plant optimization problem.*

The first assumption implies that for a given value of the inputs and parameters, the steady-state mapping, which represents the steady-state solution of the system composed by $\mathbf{f}_{ss}$ and $\mathbf{h}$, is unique in the feasible region. The second assumption assures that the functions are continuous and smooth in $\mathcal{U}$. From the third and fourth assumptions, it follows that there only exists a unique adjustable parameter set ($\mathbf{p}^{adj,*}$) that matches the plant information and, moreover, it corresponds to the true plant parameter values.

Regarding the online estimator, it needs to be able to produce unbiased values for the adjustable parameters, $\mathbf{p}^{adj,*}$. In the case of the extended Kalman filter (EKF) and the reduced extended Kalman filter (rEKF), which are the estimators applied in the case studies, it is necessary to have good initial guesses or weak nonlinearity of the functions $\mathbf{f}_{dyn}$ and $\mathbf{h}$ in Equation (3). In addition, observability of the affine system should

be guaranteed along the estimated trajectory. Note that the extended system (states + parameters) should be observable. Given these conditions, Song and Grizzle (1992) proved the boundedness of the error covariances, which implies the convergence of the EKF estimates. Note that, the rEKF enforces the boundness with Equation (19). The convergence analysis of the EKF is out of scope of this thesis, for further details please refer to Ljung (1979), Song and Grizzle (1992).

The convergence of the ROPA scheme can be stated by the following theorem:

Theorem 1. *Assuming Assumption 1 holds, then if $\mathbf{p}^{adj} \rightarrow \mathbf{p}^{adj,*}$ implies that $(\mathbf{u}, \mathbf{y}) \rightarrow (\mathbf{u}^*, \mathbf{y}^*)$.*

Proof. Since $\mathbf{u}^*$ is a strict local minimum of the optimization problem and there is no plant model mismatch, Equation (6) with the model parameters chosen as $\mathbf{p}^{adj,*}$ converges to $\mathbf{u}^*$. Therefore, if the online estimator is able to produce unbiased estimates of true plant parameters ($\mathbf{p}^{adj,*}$), the model-based optimization problem converges to $\mathbf{u}^*$. Moreover, as there is no steady-state multiplicity and the system is identifiable, there exists only one possible model parametrization that satisfies the condition above. Hence, $\mathbf{p}^{adj} \rightarrow \mathbf{p}_{adeq}^{adj} \implies (\mathbf{u}, \mathbf{y}) \rightarrow (\mathbf{u}^*, \mathbf{y}^*)$. $\square$

2.4 Conclusion

The chapter presents the fundamental properties of the new methodology that associates online estimators to the static (economic) optimizer. As a consequence, it eliminates the steady-state wait, which is one of the main drawbacks of the standard RTO cycle. In this new RTO approach, which is named Real-time Optimization with Persistent Adaptation (ROPA), the online estimator provides an estimate of the parameters at every time step, enabling the optimization cycle to be executed at an arbitrary time depending only on the decision of the practitioner, computational limitations and process dynamics.

In this chapter, the current RTO implementation has been discussed with focus on the advantages that ROPA brings to the process optimization. Several other static methods proposed in the literature that avoid the steady-state wait have been mentioned and compared to the new methodology. Almost all of them suffer from the fact that they use an outdated model in the optimization layer, leading to aggressive and profitless updates of the RTO setpoints.

In the final sections of this chapter, the online estimators used in the thesis are discussed. In addition, the convergence of ROPA is assessed formally in order to improve the technical analysis of the method. In the following chapters, ROPA is applied to three case studies in order to highlight the advantages of the new methodology.

3 Case study 1: Williams-Otto reactor

This chapter discusses the implementation of Real-time Optimization with Persistent Adaptation (ROPA) in a Williams-Otto reactor. The proposed methodology is based on the claim that by continually optimizing a steady-state model, which is updated with the online estimators, the computed solution tends to the steady-state optimum. The case study is used to assess this convergence property. The results show that ROPA is able to track the stationary plant optimum. Furthermore, the refinement of the prediction capacity by decreasing the time between two sequential optimization enhances the disturbance detection of the cycle and leads to a better economic performance.

The main findings presented in this chapter correspond to the paper *Real-time Optimization with Persistent Parameter Adaptation using Online Parameter Estimation*, published in the Journal of Process Control, Volume 68, August 2018, Pages 195-204.

3.1 Process description

The Williams-Otto plant is widely used for RTO and control studies (FORBES; MARLIN, 1996; ZHANG; FORBES, 2000; GRACIANO; MENDOZA; LE ROUX, 2013). Here, the modifications proposed by Roberts (1979) are applied: only the reactor from the Williams-Otto plant is used and the recycle stream flowrate is set to zero. A flow diagram for the reactor can be seen in Figure 4.

The two feed streams are composed by pure A and B . The other components of the reaction are: the products of interest P and E ; the undesired product G ; and the intermediate C . Only the composition of E and P are measured at the reactor outlet. The set of reactions is:

3.2 Process model

A steady-state (SS) and a dynamic model of the Williams-Otto reactor are created. The SS model is used in the RTO/ROPA layer and the dynamic model as a representation

Figure 4 – Williams-Otto reactor flowsheet. The manipulated (F_B and T_R) and measured (X_E , X_P and F_A) variables are highlighted. Note that only the fractions of the valuable products are measured (Source: own elaboration).

of the plant and for linearization purposes (for the application of the online estimator). Only the dynamic model is presented here. The stationary model can be easily derived.

The process is represented by an ideal continuous stirred-tank reactor (CSTR) with perfect level control. The equipment cooling is not limiting (i.e. the reactor temperature can be changed instantaneously and without cost), discarding the energy balance. In the dynamic model, both assumptions lead to an isothermal CSTR model in which the product concentration is directly controlled by the feed flow rate (BEQUETTE, 2003).

$$0 = F_R - F_A - F_B \quad (21a)$$

$$\frac{\partial X_A}{\partial t} = f_{dyn,1}(\mathbf{X}, \mathbf{u}) = \frac{(F_A - F_R X_A)}{M_t} - k_1 X_A X_B \quad (21b)$$

$$\frac{\partial X_B}{\partial t} = f_{dyn,2}(\mathbf{X}, \mathbf{u}) = \frac{(F_B - F_R X_B)}{M_t} - k_1 X_A X_B - k_2 X_B X_C \quad (21c)$$

$$\frac{\partial X_C}{\partial t} = f_{dyn,3}(\mathbf{X}, \mathbf{u}) = \frac{-F_R X_C}{M_t} + 2k_1 X_A X_B - 2k_2 X_B X_C - k_3 X_C X_P \quad (21d)$$

$$\frac{\partial X_E}{\partial t} = f_{dyn,4}(\mathbf{X}, \mathbf{u}) = \frac{-F_R X_E}{M_t} + 2k_2 X_B X_C \quad (21e)$$

$$\frac{\partial X_P}{\partial t} = f_{dyn,5}(\mathbf{X}, \mathbf{u}) = \frac{-F_R X_P}{M_t} + k_2 X_B X_C - 0.5k_3 X_C X_P \quad (21f)$$

$$\frac{\partial X_G}{\partial t} = f_{dyn,6}(\mathbf{X}, \mathbf{u}) = \frac{-F_R X_G}{M_t} + 1.5k_3 X_C X_P \quad (21g)$$

$$k_i = A_i e^{-E_i/T_R}, \quad i = 1, \dots, 3 \quad (21h)$$

where k_i , A_i , and E_i are the reaction rate constant, frequency factor, and activation energy for reaction i ; M_T is the reactor holdup, which is assumed constant; F_j and X_j are the mass flow rate and the mass fraction of component j (X is used for states and mass fraction because, in this case study, they are the same variables. However, to highlight the different context, upper case X is used for mass fraction and lower case x is used for the system states). The flowrate of reactant A, F_A , is fixed and measured. The manipulated variables of the system are: the flow of reactant B, F_B , and the reactor temperature, T_R .

Regarding the model parameters, the selection of adjustable set is based on Forbes and Marlin (1996). This set is composed by the frequency factors and activation energies of reactions. However, in order to avoid a poorly conditioned parameter estimation problem (GRACIANO; MENDOZA; LE ROUX, 2013), a parameter rearrangement step is carried out as in Pritchard and Bacon (1975)

$$\ln(k_i) = \Phi_i + \left(\frac{T_{ref}}{(T_R + 273.15)} - 1 \right) \Psi_i, \quad i = 1, \dots, 3 \quad (22a)$$

$$\Phi_i = \log(A_i) + \Psi_i \quad (22b)$$

$$\Psi_i = -\frac{E_i}{T_{ref}} \quad (22c)$$

where, Φ_i and Ψ_i are the rearranged parameters of the i^{th} reaction and $T_{ref} = 383.15$ K.

Model summary

Given the equations above, the complete system model states x , measurements y , decision variables u , and adjustable parameters p are:

$$\mathbf{x} = [X_A, X_B, X_C, X_E, X_P, X_G]^T \quad (23a)$$

$$\mathbf{y} = [X_E, X_P]^T \quad (23b)$$

$$\mathbf{u} = [F_B, T_R]^T \quad (23c)$$

$$\mathbf{p} = [\Phi_1, \Psi_1, \Phi_2, \Psi_2, \Phi_3, \Psi_3]^T \quad (23d)$$

The initial state of the system as well as the initial value of the parameters are listed in Table 1 and Table 2.

Table 1 – Initial plant condition. The values are based on Marchetti, Ferramosca and González (2014). The lower and upper bounds of the manipulated variables are shown because they are used in the economic optimization problem, Equation (6).

	Variable (symbol) [unit]	Lower bound	Initial value	Upper bound
–	Reactor holdup (M_T) [kg]	–	2105	–
–	Flow of reactant A (F_A) [kg/s]	–	1.827	–
u_1	Flow of reactant B (F_B) [kg/s]	4	4.787	6
u_2	Reactor temperature (T_R) [°C]	80	89.70	100
x_1	Mass fraction of A (X_1) [–]	–	0.0876	–
x_2	Mass fraction of B (X_2) [–]	–	0.3892	–
x_3	Mass fraction of C (X_3) [–]	–	0.0153	–
x_4/y_1	Mass fraction of P (X_4) [–]	–	0.2903	–
x_5/y_2	Mass fraction of E (X_5) [–]	–	0.1093	–
x_6	Mass fraction of G (X_6) [–]	–	0.1083	–

Table 2 – Adjustable parameter set of the Williams-Otto model. The nominal parameter values are based on Marchetti, Ferramosca and González (2014). Here, only the rearranged parameters are shown. The lower and upper bounds are arbitrarily determined based on the nominal values of the nonrearranged parameters.

	Variable (symbol) [unit]	Lower bound	Nominal value	Upper bound
p_1	Φ_1 [–]	-3.385	-3.077	-2.769
p_2	Ψ_1 [–]	-18.26	-17.39	-14.78
p_3	Φ_2 [–]	-1.826	-1.353	-1.217
p_4	Ψ_2 [–]	-22.83	-21.74	-17.39
p_5	Φ_3 [–]	-0.5188	-0.3843	-0.3459
p_6	Ψ_3 [–]	-30.44	-28.99	-27.54

3.3 Case study

The economic optimization objective of the case study is to maximize the unit profit, ϕ_{profit} . E and P are the valuable products and both A and B are purchased. The excess heat generated by the reaction can be removed without cost:

$$\phi_{profit} = 1143.38X_P F_R + 25.92X_E F_R - 76.23F_A - 114.34F_B \quad (24)$$

The variables are related to the model presented in the previous section. The numerical values in the equation above represent the market prices of E and P as well as the purchasing values of the reactants A and B . All the values are in [\$/kg]. Given the operation range (determined by the input constraints), the optimization space is reduced, implying in a strictly convex profit surface with an unique local optimum for the nominal parameter set (FORBES; MARLIN, 1996).

In order to mimic a realistic situation, noise is added to the outputs of the dynamic model, which is considered as the plant. The random values are drawn from the standard

uniform distribution in the interval $(-1, 1)$. The noise is neither correlated between measurements nor in time. For generating the measurement noise at the k^{th} time instant, the following equation is used:

$$y_{n,k} = y_k + \text{error} \times y_k \times \text{randn}() \quad (25)$$

where, $y_{n,k}$ is the measurement with noise at time k ; y_k is its actual value; the *error* is equal to 0.1%; and *randn* draws a random scalar from the standard normal distribution as mentioned earlier.

The system is also affected by deterministic disturbances, which can be divided into measured and unmeasured. The measured disturbance affects the feed flowrate of reactant A, and the unmeasured disturbances change the parameters of the second and third reaction, E_2 and E_3 . The disturbances profiles are presented in Figure 5.

After 6.25 hours operating at the parameter nominal value, the values of E_2 and E_3 suffer a step-like disturbance. Then, starting at 12.5 hours, the parameters return linearly to their initial value, reaching their nominal value around 15.7h. The situation may take place due to catalyst deactivation and subsequent substitution of the spent catalyst. After that, a step change of 0.15 kg/s affects F_A at 18.8 hours.

Figure 5 – Deterministic disturbance profiles. It is important to note that after the parameter rearrangement (Equation (22)), the disturbance in E_2 affects both Φ_2 and Ψ_2 (the same for E_3 , which affects Φ_3 and Ψ_3). The disturbance profile of the rearranged parameters is not shown here for brevity (Source: own elaboration).

In the following sections, extra details for the implementation of Model Parameter Adaptation (MPA) and ROPA are outlined. Also, due to the fact that the control layer is

included in this case study, the derivation of the control problem as well as the controller parameters are shown.

Model parameter adaptation

All the steps of the MPA implementation were previously defined and discussed in Chapter 2. For the SS detection, the filter values are shown below:

$$\boldsymbol{\lambda}_1 = [0.2, 0.3] \quad \boldsymbol{\lambda}_2 = [0.1, 0.2] \quad \boldsymbol{\lambda}_3 = [0.1, 0.2] \quad \mathbf{R}_{crit} = [2.5, 2.5] \quad (26)$$

where, the first value of the array is related to the outlet composition of E and the second, to the outlet composition of P . The critical value for the variance ratio, $\mathbf{R}_{crit}$, is chosen based on the suggestion of Cao and Rhinehart (1997).

The model adaptation algorithm, Equation (5), uses the rearranged parameters discussed above. The upper and lower bounds for the parameters are shown in Table 2. The weighting factor for the estimation problem, $\mathbf{R}_p$, is the identity matrix.

Finally, the optimal decisions of the economic optimization problem, Equation (6), are implemented with an exponential filter. The filtered decision is obtained as $\mathbf{u}_{f,j}^* = \mathbf{K}_{opt} \mathbf{u}_j^* + (\mathbf{I} - \mathbf{K}_{opt}) \mathbf{u}_{f,j-1}^*$, with j as the current optimization execution. The filtering is used to prevent abrupt changes in the optimal decision between optimization executions, which also mitigates the effect of measurement noise (FRANÇOIS; BONVIN, 2013). The filter gain is:

$$\mathbf{K}_{opt} = [0.8, 0.6] \quad (27)$$

where, the first value is related to F_B and the second to T_R . After filtering the optimal decisions, $\mathbf{y}^*$ are calculated using the SS model.

ROPA

As the Williams-Otto reactor presents problem with identifiability (GRACIANO; MENDOZA; LE ROUX, 2013), the reduced extended Kalman filter (rEKF) is chosen over

the extended Kalman filter (EKF) to be used as the online estimator. The filter tuning parameters used for the rEKF are shown below:

$$\mathbf{Q} = \text{diag}([1e^{-1}; 1e^{-1}; 1e^{-1}; 1e^{-2}; 1e^{-2}; 1e^{-2}]) \quad (28a)$$

$$\mathbf{Q}^p = \text{diag}([1e^1; 0.4e^3; 1e^1; 0.2e^3; 1e^0; 0.5e^2]) \quad (28b)$$

$$\mathbf{R} = \text{diag}([1e^1; 1e^1]) \quad (28c)$$

$\mathbf{Q}$ entries correspond to the states ($X_A, X_B, X_C, X_P, X_E, X_G$) and $\mathbf{R}$ entries, to the measured variables (X_E, X_P); $\mathbf{Q}^p$ entries are related to the parameters ($\Phi_1, \Psi_1, \Phi_2, \Psi_2, \Phi_3, \Psi_3$). The values of $\mathbf{Q}$, $\mathbf{Q}^p$ and $\mathbf{R}$ are kept constant during the simulations.

$\mathbf{R}$ entries are determined based on the process noise standard deviation (Equation (25)). On the other hand, $\mathbf{Q}$ entries are determined based on the confidence in the model. Since there is no process-model mismatch, the values are significantly smaller than in $\mathbf{R}$, showing that the filter can “trust” the model predictions. Regarding $\mathbf{Q}^p$, its tuning is much more difficult. They are chosen based on the results of Graciano (2015). The idea is to represent the identifiability issues of the Williams-Otto reactor model, using $\mathbf{Q}^p$ entries with a larger order of magnitude than in $\mathbf{Q}$. This choice reflects the fact that the parameter estimates are heavily affected by noise. Moreover, the entries related to Ψ_1, Ψ_2 and Ψ_3 are larger than the others, which shows that these parameter estimates suffer from a greater variability than the others.

Another important tuning parameter for ROPA is the period between two optimization executions, Δt_{ROPA} . The execution period needs to have a similar order of magnitude as the process settling time, but it should be large enough to capture the disturbance rate and frequency. Additionally, the execution period can change during the simulations depending on the disturbance, which improves even further the ROPA results. Despite being a potential tuning parameter, Δt_{ROPA} is fixed in the case study in 15 min, which is in the same order of magnitude as the reactor residence time.

Model predictive controller

In order to obtain a complete assessment of the influence of the estimation routines on the RTO, model predictive control (MPC) is implemented. The controller receives the optimum solution from the RTO as setpoints for the controlled variables and should be able to provide offset-free performance. The prediction model used in the MPC layer is

discrete, linear and time invariant. The dynamic model, Equation (21), is linearized using a first-order expansion around $(\mathbf{x}_{nom}, \mathbf{u}_{nom})$ with $\mathbf{p}_{nom}$ held constant.

$$\bar{\mathbf{x}}_{k+1} \approx \mathbf{A}\bar{\mathbf{x}}_k + \mathbf{B}\bar{\mathbf{u}}_k, \quad \bar{\mathbf{y}}_k \approx \mathbf{C}\bar{\mathbf{x}}_k \quad (29)$$

where, the bar above symbols indicates deviation from the nominal values (i.e. $\bar{\mathbf{x}}_k = \mathbf{x}_k - \mathbf{x}_{nom}$), which are shown in Table 1. Additionally:

$$\mathbf{A} = \left. \frac{\partial \mathbf{f}_{dyn}(\mathbf{x}, \mathbf{u}, \mathbf{p})}{\partial \mathbf{x}^T} \right|_{nom}, \mathbf{B} = \left. \frac{\partial \mathbf{f}_{dyn}(\mathbf{x}, \mathbf{u}, \mathbf{p})}{\partial \mathbf{u}^T} \right|_{nom}, \mathbf{C} = \left. \frac{\partial \mathbf{h}(\mathbf{x}, \mathbf{u}, \mathbf{p})}{\partial \mathbf{x}^T} \right|_{nom} \quad (30)$$

In order to obtain an offset-free performance, the model is rearranged in complete velocity form (PRETT; GARCIA, 2013):

$$\begin{aligned} \begin{bmatrix} \Delta \bar{\mathbf{x}}_{k+1} \\ \bar{\mathbf{y}}_{k+1} \end{bmatrix} &= \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{CA} & \mathbf{I}_{ny} \end{bmatrix} \begin{bmatrix} \Delta \bar{\mathbf{x}}_k \\ \bar{\mathbf{y}}_k \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{CB} \end{bmatrix} \Delta \bar{\mathbf{u}}_k \\ \bar{\mathbf{y}}_k &= \begin{bmatrix} \mathbf{0} & \mathbf{I}_{ny} \end{bmatrix} \begin{bmatrix} \Delta \bar{\mathbf{x}}_k \\ \bar{\mathbf{y}}_k \end{bmatrix} \end{aligned} \quad (31)$$

in which, $\mathbf{u}_k = \mathbf{u}_{k-1} + \Delta \mathbf{u}_k$. For simplification, System (31) can be written as:

$$\bar{\mathbf{x}}_{inc,k+1} = \mathbf{A}_{inc}\bar{\mathbf{x}}_{inc,k} + \mathbf{B}_{inc}\Delta \bar{\mathbf{u}}_k, \quad \bar{\mathbf{y}}_k = \mathbf{C}_{inc}\bar{\mathbf{x}}_{inc,k} \quad (32)$$

Then, the rearranged model is used to calculate the optimal input sequence $\bar{\mathbf{U}}^{m_c, \star} = [\bar{\mathbf{u}}_{k|k}^*, \bar{\mathbf{u}}_{k+1|k}^*, \bar{\mathbf{u}}_{k+2|k}^*, \dots, \bar{\mathbf{u}}_{k+m_c|k}^*]$ based on prediction of the process behavior for the next p_c time steps, with $p_c \geq m_c$. Note that the control input remains constant after the time instant $k + m_c$ until the end of the prediction horizon (i.e. $k + m_c + p_c$). The optimal sequence is the argument that minimizes the following optimization problem:

$$\begin{aligned} \bar{\mathbf{U}}^{m_c, \star} &= \underset{\bar{\mathbf{U}}^{m_c}}{\operatorname{argmin}} \sum_{i=1}^{p_c} \|\bar{\mathbf{y}}_{k+i|k} - \bar{\mathbf{y}}_k^{sp}\|_{\mathbf{Q}_y}^2 + \sum_{i=0}^{m_c-1} \|\Delta \bar{\mathbf{u}}_{k+i}\|_{\mathbf{R}_u}^2 \\ &\text{s. t.} \\ \bar{\mathbf{x}}_{inc,k+i+1} &= \mathbf{A}_{inc}\bar{\mathbf{x}}_{inc,k+i} + \mathbf{B}_{inc}\Delta \bar{\mathbf{u}}_{k+i}, \quad i = 0, \dots, p_c - 1 \\ \bar{\mathbf{y}}_{k+i} &= \mathbf{C}_{inc}\bar{\mathbf{x}}_{inc,k+i}, \quad i = 1, \dots, p_c \\ \bar{\mathbf{u}}^L &\leq \bar{\mathbf{u}}_{k+i} \leq \bar{\mathbf{u}}^U, \quad i = 0, \dots, m_c - 1 \\ -\Delta \bar{\mathbf{u}}^{max} &\leq \Delta \bar{\mathbf{u}}_{k+i} \leq \Delta \bar{\mathbf{u}}^{max}, \quad i = 0, \dots, m_c - 1 \end{aligned} \quad (33)$$

where, $\bar{\mathbf{U}}^{m_c}$ is the input sequence defined previously; $\mathbf{Q}_y$ and $\mathbf{R}_u$ are the weight matrices for the setpoint deviation and the control increments, respectively. Both are positive definite

matrices with appropriate dimensions. $\bar{\mathbf{u}}^L$ and $\bar{\mathbf{u}}^U$ are the lower and upper bounds for the inputs (in deviation form), which are coherent with the bounds used in the RTO layer. $\bar{\mathbf{y}}_k^{sp}$ is the solution, in deviation form, of the optimization layer that was referred to as $\mathbf{y}^*$ in Equation (6). Then, after computing the input sequence from the open loop solution, the controller applies the first value of $\bar{\mathbf{U}}^{m_c,*}$ in order to create a feedback law (GRÜNE, 2013).

In this case study, the controller has 2 manipulated variables, F_B and T_R , and 2 controlled variables, X_E and X_P . Also, it has an output prediction horizon $p_c = 5$ and input horizon $m_c = 3$. The control sampling time is 1 min. The additional tuning parameters are: $\mathbf{Q}_y = \text{diag}([2.5e^3; 1e^3])$, $\mathbf{R}_u = \text{diag}([1e^0; 1e^0])$ and $\Delta\bar{\mathbf{u}}^{max} = [8e^{-3}, 8e^{-2}]^T$. State feedback for the MPC controller is achieved using the state estimated by the rEKF in both ROPA and MPA cases. The same estimator is chosen in order to avoid a major influence of the controller in the comparison between the methodologies. Additionally, the disturbances shown in Figure 5 affect the controller response because they move the process far from the nominal point.

3.4 Results and discussion

The performance of standard RTO (represented by MPA) and ROPA are analyzed by comparing the states and parameters predictions, as well as the economic results (optimal decision, costs and profits). ROPA promotes the association of online estimations with static optimization. Thus, the SS model is always up-to-date with the plant. Due to the fact that the SS wait is no longer necessary, the ROPA cycle can be executed every Δt_{ROPA} instants, decreasing the period between two consecutive executions of the optimization cycle and enhancing its prediction capacity under the presence of process/economic disturbances.

All the results shown in this section are obtained using the Monte Carlo Method (BARD, 1974). This assessment is carried out in order to obtain more information about the state/parameter estimates distribution. For example, given an initial parameter guess and noise distributions, it is possible to approximate the expected value of the state/parameter distribution, which allows to understand how the variability of the states/estimates might affect the economic performance. 100 repetitions of the simulation are executed for MPA and ROPA. The n^{th} repetition uses the same parameter initial guess and sequence of measurement noise for both methodologies. The results shown in this section are the average values over all the replications.

Figure 6 – Comparison of the estimated states profiles for ROPA and the standard RTO implementation, MPA. The states X_E and X_P are not shown because they are directly measured in the plant (Source: own elaboration).

The first analysis compares the state estimation quality of ROPA and MPA cycles. As discussed earlier, rEKF is used as the state estimator in both cases. The state observer is necessary because of the MPC implementation (i.e. it is necessary to have an estimate of the states at every MPC sampling time, which leads to an almost continuous profile in Figure 6). However, in MPA only the states are estimated by the rEKF, the parameters are updated as described in Equation (5).

Figure 6 shows the estimated (full line) and actual state (dashed line) for both cases. The system state trajectories are presented in different figures for MPA and ROPA because each method makes different optimal decisions in face of the disturbances. As a consequence, the results of the plant state profiles are different for each method. However, the estimated and actual states profiles for both cases are almost overlapped.

Regarding the state estimation, the behavior of the filter indicates that the covariance matrix estimate tends fast to its stationary value, suggesting that the filter is working properly and that the state covariance matrix is well tuned. Despite not being the main focus of the paper, the online estimator should be able to predict the system states satisfactorily, because the state prediction is used as feedback to the MPC layer and the control is directly responsible for implementing optimal decisions in the plant. Hence, the MPC performance is intrinsically connected with the quality of implementation, which directly affects the optimality of the RTO/ROPA decisions (MARCHETTI; FERRAMOSCA; GONZÁLEZ, 2014)

To assess the parameters estimation, the predicted parameter values are compared to the actual “plant” values in Figures 7a to 7c. The setting of the plots is slightly different from the previous ones. The plant value is still shown as a dashed line. However, as the parameter disturbances are the same for the MPA and ROPA simulations, the estimated profiles are plotted in the same figures (black for ROPA and blue for MPA). In addition, the constraints used in the MPA model adaptation algorithm, Equation (5), are shown.

Clearly, the parameter estimates generated by MPA are unreliable. After the first disturbance, the SS detection method takes almost 1h to trigger the MPA cycle. Then, the Φ_3 and Ψ_3 values estimated by the model adaptation algorithm diverge from the plant values to a great extent, which can be seen in Figure 7c. The same happens after the ramp disturbance ceases, around 16h. This divergence comes from the identifiability issues of the Williams-Otto reactor, which has been previously identified by Graciano (2015).

Although the identifiability analysis is not the main goal of this case study, the ill-conditioning of estimation problem clearly affects the performance of the optimizer. Mathematically, the identifiability issues arise if the Hessian matrix of the problem is ill-conditioned (MCLEAN; MCAULEY, 2012). In these cases, the parameter adaptation algorithm adjusts the parameter values but these adjustments fail to improve the objective function of the problem. The model predicted values are insensitive to changes in some combination of parameters.

If constraints are used to decrease the feasible region of the estimation problem, the optimization algorithm is prone to obtain estimate values closer to their lower and upper bounds, generating parameter inflation (GRACIANO; MENDOZA; LE ROUX, 2014). This behavior has the potential to generate unreliable parameters estimates for the economic optimization and helps to justify the choice of rEKF over other online estimation methods that use constraints (like CEFK and MHE) in this case study.

Instead of adjusting all the parameters in $\mathbf{p}^{adj}$ as in the MPA implementation, the rEKF method adjusts only a subspace of model parameters and maintains the other parameters at their previously adjusted values or at the nominal values. This behavior can be seen in Figures 7a to 7c. Ψ_1 and Ψ_2 are slightly adjusted from their initial arbitrary value and, then, the value is kept constant. In the other hand, Ψ_3 value is not adjusted in any of the simulations. The choice of the adjustable parameter subspace is online and addressed by spectral decomposition (Equation (19) and (20)). Due to the ill-conditioning of the Hessian matrix, which is directly seen in the rank-deficiency of the estimate covariance

(a) Reaction 1

(b) Reaction 2

(c) Reaction 3

Figure 7 – Comparison of the estimated parameter profiles for ROPA and the standard RTO implementation, MPA (Source: own elaboration).

matrix $\mathbf{P}_{k|k}$, rEKF reduces the search directions of the parameter estimation problem only to the principal components directions of $\mathbf{P}_{k|k}$, choosing the subset of parameters with the smallest variance (i.e. smallest eigenvalues).

The superior performance of rEKF is clearer in Φ_2 and Φ_3 profiles. The predicted parameters are much closer to the actual values. Moreover, the Φ_3 estimated values in the MPA cycle diverge from the plant values in such way that the disturbance affecting Φ_3 cannot be seen in Figure 7c. In addition, the gap between the disturbance occurrence and the response of ROPA is much smaller than MPA, which highlights a critical advantage of ROPA. The methodology does not need to wait that the system reaches steady-state to update the model. For example, in Figures 7b and 7b, it is possible to notice that after the first step disturbance, the MPA cycle needs multiple runs to estimate the parameters correctly. Therefore, it takes around 2.5 hours to update the model parameters to the actual value. On the other hand, the ROPA cycle completes the same task in approximately 1h. For the ramp disturbance, ROPA estimates follow the actual plant parameter profile, while MPA estimates are updated only after the disturbance ceases, which takes almost 3h.

Nonetheless, assessing the method's performance only with the parameters is misleading as the system presents identifiability issues. In these systems, different sets of parameters can map to the same set of states. Hence, the analysis of the parameter profiles may not give insights about the model capacity to reflect the plant state (for example: in the case that $\mathbf{p}_1^{adj} \mapsto \mathbf{y}_1$, $\mathbf{p}_2^{adj} \mapsto \mathbf{y}_1$, it is clear that the model updated with $\mathbf{p}_1^{adj}$ and $\mathbf{p}_2^{adj}$ can achieve similar economic performance for both sets of parameters). Therefore, this analysis is not sufficient in order to conclude if the continuous optimization of the steady-state model leads to the convergence of the overall system to the plant steady-state optimum.

A more proper indicator is the optimal decisions and the economic performance of the closed-loop optimization, which can be seen in Figure 8 and 9, respectively. The figures scheme follows the same pattern of the estimated parameter plots, except that the input constraints are not plotted. The optimal input value is calculated by simply solving the economic-based optimization problem with the “true” parameter values shown in the previous figures. The analysis of the optimal inputs calculated by both methods also highlights some important features of the proposed framework.

Figure 8 – Comparison between the economic performance of ROPA and MPA using the input moves calculated by the economic optimization problem (Source: own elaboration).

First, even though the SS optimization problem of ROPA is solved every sampling time using parameters estimated with transient measurements (resulting in the step-like shape of the ROPA optimal decision profiles in Figure 8), the methodology is able to drive the plant to its optimum steady-state. Note that, the optimal trajectory is approximated by recomputing the SS optimal solution continually. This result supports the fundamental idea behind ROPA, which is based on the persistent pursuit of an steady-state optimum, which most likely will be reached in a broader time horizon. Second, the input sequence behavior of ROPA is smoother than MPA in general. The average ROPA decision moves towards the optimal values step-by-step, while MPA average decisions change abruptly whenever the system reaches steady-state and the optimization cycle starts to run, even with the input exponential filter (Equation (27)). Third, the average MPA decision drifts from the optimal condition due to a considerable amount of parametric uncertainty. The model adaptation step of MPA fails to use the measurements to update the model, driving the plant far from optimality.

Such behavior is further intensified by the low number of economic optimization executions. In all static RTO frameworks (MPA and ROPA included), several iterations are

generally necessary for achieving convergence to the plant optimal. Hence, it is desirable to have the largest number of optimization runs during the simulations. Again, the key advantage of ROPA appears. ROPA can run continually, during any plant condition, depending only on the process and computational limitations, while MPA only runs when the system reaches steady-state. ROPA obtains an almost continuous adjustment of set-point values, according to the current plant disturbances, which avoids the wait for a steady-state condition and the inherent delay on optimization due to SS detection. In the ROPA simulations, the optimization runs every 15 min (resulting in 100 executions per simulation). In turn, MPA executes, in average, only 13 times per simulation. Such decrease in the period between optimizations provides more information to the optimizer, allowing more frequent updates in the optimal decisions.

In contrast to calculating an optimal trajectory, ROPA improves the plant operation continually based on a hypothetical future steady-state condition. Despite the fact that no optimality can be guaranteed in the solution of the SS optimization problem using dynamic data, the plant operation will eventually reach the steady-state (given the conditions for convergence stated in Chapter 2). Hence, ROPA does not seek exclusively to optimize the process, but rather to improve it step-by-step pursuing the SS optimum.

In the same manner as in the parameter responses, the effect of the MPA steady-state wait delay can be seen in the optimal decisions adjustment. Regarding the ramp disturbance between 12.5h and 15.7h, MPA starts to run in average only around 16h. A considerable amount of time after the disturbance occurrence, bearing in mind the process settling time. After a few iterations, the inputs calculated by the MPA start to move towards the optimal values, however the presence of a second disturbance at 18.8h affects the performance driving the RTO solutions far from the optimum. In turn, ROPA improves the input values constantly, reaching the steady-state optimal faster.

A second indicator used for evaluating ROPA is the overall economic performance. Regarding the instantaneous profit (Figure 9), the response of ROPA is faster for both measured and unmeasured disturbances. Also, ROPA performance is better for both forms of disturbance, step and ramp.

During the ramp disturbance, the MPA does not take any actions, which can be confirmed in Figure 8. Due to the nature of the disturbance, the profit of the unit increases. However, after the parameter disturbance ceases, which happens around 14h, the profit obtained with the MPA diverges from the optimal profit. The main reason is the values of

Figure 9 – Comparison between the economic performance of ROPA and MPA using the instantaneous profit (Source: own elaboration).

the parameters estimated by the model adaptation step, as seen in Figures 7a to 7c. They do not correspond to the actual plant values, which affects the optimality of the RTO solution. Therefore, the optimizer drives the plant far away from its optimal operation. After a few MPA runs, the optimal decisions calculated by the optimizer start to tend to the optimum. In practical applications, such behavior could be avoided by implementing a simple algorithm that checks if the new optimization solution improves or not the economic objective function.

Additionally, between 15.7h and 18.8h, a bias can be observed in the optimal value for both methodologies. The reasons are the plant/controller model mismatch and the absence of a input target in the controller implementation. The MPC is able to provide offset-free performance (the optimal values for the controlled variables are reached), however, to do so, the controller uses values for the manipulated variables that are slightly different to the optimal ones, resulting in a suboptimal performance. Note that the instantaneous profit can be larger than the steady-state profit during some periods. However, it tends to the stationary value whenever plant tends to steady-state.

3.5 Conclusion

The results obtained in this case study illustrate the convergence properties of the ROPA methodology. In addition, they are in good agreement with the claim that the integration of static optimization with online estimators improves the prediction capacity and, most important, the overall economic results. Furthermore, ROPA keeps the standard RTO formulation almost unchanged and the possibility to use static optimization represents a great advantage due to the vast literature available (e.g. modeling of SS processes; reported implementations of static RTO; static optimizers; and reliable commercial solutions for RTO).

Clearly, ROPA could destabilize the plant if it is not well tuned. The execution period (time between two successive ROPA executions), the filter of the optimal optimization decision, and the rEKF parameters play a major role in the stabilization of the plant dynamics and, in an actual application, these tuning parameters values should not be neglected.

Finally, despite the fact that all types of on-line estimators can be used in the ROPA formulation as long as they are able to provide state and parameter estimates, the use of rEKF provides an alternative to mitigate the identifiability problems of the system. Albeit identifiability is not the main issue of the paper, such ability provides substantial benefits to the optimization cycle (GRACIANO; MENDOZA; LE ROUX, 2014).

4 Case study 2: Fluid catalytic cracking unit

In this chapter, both ROPA and the MPA are applied to the conversor of a Fluid Catalytic Cracking (FCC) unit. This process plays a critical role in the refinery economic performance because it converts low-value high-molecular weight compounds into more valuable products, like gasoline and olefinic gases (ZANIN; TVRZSKA; ODLOAK, 2002). Although it is an attractive candidate for RTO applications, detecting steady-state (SS) periods during operation is a very difficult task to be carried out. Mainly because system parameters (like reaction constants) change significantly in time and external disturbances affect the process continuously (MORO; ODLOAK, 1995).

Thus, there are potential benefits to avoid the SS wait in the RTO cycle of an FCC unit and ROPA appears as an interesting candidate to optimize this process. In order to make the case study more realistic (validating the reaction rates and semi-empirical relations used in the model), the open-loop FCC model response is compared with actual plant data obtained from the conversor of the FCC unit of Refinery Henrique Lage owned by Petrobras S.A.. The model validation is shown in details in Appendix A. The case study results confirm the claim that by continually optimizing a SS model, even in transient periods, the optimization cycle is able to track the actual steady-state optimum. Moreover, ROPA decreases in approximately 60% the time to achieve the optimal gasoline production of the unit in comparison to MPA.

4.1 Process description

The conversor of the Fluid Catalytic Cracking (FCC) can be divided in two subsystems, riser/reactor and regenerator. The feed, which is mainly composed by gasoil and atmospheric residual, is mixed with hot regenerated catalyst at the riser/reactor entrance. The contact with the hot catalyst vaporizes the feed and promotes the cracking reactions, generating lower boiling temperature products like gasoline, light cycle-oil, butanes and gas. As an undesired consequence, the endothermic cracking reactions deactivate the catalyst by forming coke on its surface. The spent catalyst is fed to the regenerator, where air is used to promote the combustion of coke to CO and CO_2 . Both gases and the excess air of the combustion reaction leave the regenerator as flue gas.

Figure 10 – Flowsheet of the conversor of a Fluid Catalytic Cracking unit (Source: own elaboration)

The catalyst cycle is completed by feeding the regenerated catalyst to the riser. This loop takes an important role in the energy balance inside the unit, because the coke-burning reactions supply the heat necessary for the endothermic cracking reaction. This interdependence generates highly coupled mass and energy balances in the system. Figure 10 shows a simplified flowsheet of the model (left) and a scheme illustrating the model inputs, outputs and control loops (right).

The model has 5 degrees of freedom, however there is a control loop that adjusts the air blower flowrate (R_{ai}) in order to keep the oxygen concentration of the flue gas at a given setpoint. Hence, only 4 variables can be freely manipulated: catalyst recirculation rate (R_{rc}), feed rate (R_{tf}), feed temperature (T_{fp}), and air temperature (T_{ai}). Both temperatures can be manipulated due to a preheat system. The system measured variables are: the temperatures of the riser/reactor (T_{ra}) and regenerator (T_{rg}), as well as the flue gas temperature (T_{fg}) and oxygen concentration (O_{fg}).

In order to stabilize the unit, a proportional-integral (PI) controller is implemented. It uses the value of the air flowrate, R_{ai} , for maintaining the oxygen percentage in the flue gas at the optimum setpoint, which is important for safety reasons. The manipulated variables as well as the setpoint for the PI controller are determined by the steady-state optimizer (ROPA or MPA), which maximizes the production of gasoline while restricting the variables within allowable ranges. Note that, no MPC is used in this case study.

4.2 Process model

The FCC model is based on Kurihara (1967). Due to its simplicity, this model has been widely used in the literature, specifically in control studies (DENN, 1986). The model equations are presented in Appendix A.

Model summary

The complete system model states x , measurements y , decision variables u , and adjustable parameters p are:

$$\mathbf{x} = [C_{cat}, C_{cs}, C_{rc}, T_{ra}, T_{rg}]^T \quad (34a)$$

$$\mathbf{y} = [O_{fg}, T_{fg}, T_{ra}, T_{rg}]^T \quad (34b)$$

$$\mathbf{u} = [T_{fp}, R_{tf}, T_{ai}, R_{rc}]^T \quad (34c)$$

$$\mathbf{p} = [k_{cc}, E_{cc}, k_{cr}, E_{cr}, \Delta H_{cr}, k_{or}, E_{or}, \Delta H_{rg}, C_1, C_4, \Delta H_{fv}, H_{ra}]^T \quad (34d)$$

The variables are also explained in Appendix A. Briefly: the states are composed by the total carbon concentration on the surface of the catalyst in the riser (C_{cs}) and in the regenerator (C_{rc}), temperatures in the riser (T_{ra}) and in the regenerator (T_{rg}), and a state representing the catalytic carbon concentration (C_{cat}), which is the carbon deposited on the surface due to the cracking reactions (i.e. it does not considered the carbon that is already deposited on the surface of the regenerated catalyst). The adjustable parameter set contains: the parameters of the carbon deposition on the catalyst (k_{cc}, E_{cc}), of the cracking reaction (k_{cr}, E_{cr}), and of the oxygen formation (k_{or}, E_{or}); Heat of feed vaporization (ΔH_{fv}), of carbon burning (ΔH_{rg}), and of cracking (ΔH_{cr}); two semi-empirical constants for calculating coke burning rate (C_1) and flue gas temperature (C_4); and, finally, the catalyst holdup in the reactor (H_{ra}). The inputs (manipulated variables) and the system measurements were explained in the previous section.

The initial state of the system, as well as the initial value of the parameters are listed in Table 3 and Table 4. Note that the validation step presented in Appendix A uses a different plant condition, specifically different parameters and inputs. The plant data is used only to validate the model dynamics and prediction capacity. The plant condition used in the case study comes from Kurihara (1967).

Table 3 – Initial plant condition. The values are based on Kurihara (1967). The lower and upper bounds of the manipulated variables are shown because they are used in the economic optimization problem, Equation (6). The oxygen concentration and flue gas temperature upper bound constraints are also included because they are considered in the optimization problem as process safety constraints.

	Variable (symbol) [unit]	Lower bound	Initial value	Upper bound
u_1	Feed preheater temperature (T_{fp}) [$^{\circ}\text{F}$]	520	580.0	580
u_2	Total feed rate (R_{tf}) [Mbbbl/day]	80	83.00	83
u_3	Air inlet temperature (T_{ai}) [$^{\circ}\text{F}$]	400	500.0	500
u_4	Catalyst circulation rate (R_{rc}) [ton/min]	50	77.61	80
u_5	Air flowrate (R_{ai}) [Mlb/hr]	400	579.3	1000
x_1	Catalytic carbon on spent catalyst (C_{cat}) [wt%]	-	0.327	-
x_2	Total carbon on spent catalyst (C_{sc}) [wt%]	-	0.7956	-
x_3/y_1	Riser temperature (T_{ra}) [$^{\circ}\text{F}$]	-	940.3	-
x_4	Carbon on regenerated catalyst (C_{rc}) [wt%]	-	0.7923	-
x_5/y_2	Regenerator temperature (T_{rg}) [$^{\circ}\text{F}$]	-	1100	-
y_3	Flue gas temperature (T_{fg}) [$^{\circ}\text{F}$]	-	1103	1400
y_4	Flue gas oxygen concentration (O_{fg}) [vol%]	-	0.7231	1.5

Table 4 – Adjustable parameter set of the FCC model. The nominal parameter values are based on Kurihara (1967). The lower and upper bounds are arbitrarily determined based on the nominal values.

	Variable (symbol) [unit]	Lower bound	Nominal value	Upper bound
p_1	Carbon formation reaction (k_{cc}) [Mlb/(hr psia ton (wt% ^{1.06}))]	50	103.3	150
p_2	Carbon formation reaction (E_{cc}) [BTU/lbmole]	1e2	2.925e4	5e4
p_3	Carbon cracking reaction (k_{cr}) [Mbbbl/(day psia ton (wt% ^{1.15}))]	5e5	1.038e6	1.5e6
p_4	Carbon cracking reaction (E_{cr}) [BTU/lbmole]	1e2	3.048e4	5e4
p_5	Heat of cracking (ΔH_{cr}) [BTU/lb]	0	44.69	200
p_6	Carbon burning reaction (k_{or}) [Mlb/(hr psia ton)]	5e11	1.003e12	1.5e12
p_7	Carbon burning reaction (E_{or}) [BTU/lbmole]	5e4	9.016e4	1.5e5
p_8	Heat of regeneration (ΔH_{rg}) [BTU/lb]	1e2	1.526e4	5e4
p_9	Stoichiometric coefficient (C_1) [lb oxygen/lb carbon]	1.5	3.851	10
p_{10}	Flue gas temperature coefficient (C_4) [$^{\circ}\text{F}/\text{vol}\%O_2$]	0	3.729	10
p_{11}	Heat of feed vaporization (ΔH_{fv}) [BTU/lb]	0	35.76	50
p_{12}	Reactor catalyst holdup (H_{ra}) [ton]	50	104.1	150

4.3 Case study

The economic optimization problem aims to maximize the production of gasoline instead of a standard profit function. However, the objective function is referred as profit in order to use the same nomenclature and notation as in the other case studies.

$$\phi_{profit} = P_{gasoline}(T_{ra}, T_{rg}, T_{fd}, R_{tf}) \quad (35)$$

in which, $P_{gasoline}$ is an empirical correlation that computes the gasoline outflow. The objective function is defined in Appendix A and is based on Zanin (2001). In addition to maximize gasoline production, the other degrees of freedom of the optimization problem are used for maintaining the process within the operation range, which is extremely important due to safety reasons. The main process safety constraints are related to the maximum

value of flue gas temperature and of the oxygen in flue gas. The first constraint protects the cyclones, which remove particles from the flue gas, from excessive afterburning of carbon monoxide. The second avoids extreme temperatures at the regenerator due to heat generated by the coke combustion.

Despite using actual data to validate the model, the dynamic model represents the plant in the case study. The filter configuration and parameter values presented in Appendix A are not applied here. The validation is only carried out in order to confirm that the reaction rates and semi-empirical relations are able to reproduce the behavior of a real process. In the case study, the plant measurements are obtained as in the previous case study (Equation (25)), where noise is added to the outputs of the dynamic model representing the plant. The *error* is in the range of 0.1% the current value of the measurement and *randn* value is drawn from the standard uniform distribution in the interval $(-1, 1)$. Additionally, the system is also affected by deterministic disturbances. The parameters of the carbon formation reaction change in a ramp-like disturbance. Starting at the nominal point shown in Table 4, the parameter values gradually change at 2100 s until they reach the values of $k_{cc} = 104.1$ and $E_{cc} = 2.889e4$ at 4300 s. Clearly, the changes in the reaction parameter affect the optimal gasoline production of the system.

In the following sections, extra details of the implementation of Model Parameter Adaptation and ROPA are given.

Model parameter adaptation

All the steps of the MPA implementation were previously discussed in Chapter 2. The implemented steady state detection method is the same as in the previous case study. However, here, the variables that indicate steady-state are the system temperatures. The filter values for each measured variable and the critical value for the variance ratio are shown below. The values are related to T_{ra} , T_{rg} and T_{fg} , respectively.

$$\begin{aligned} \lambda_1 &= [0.02, 0.03, 0.03] & \lambda_2 &= [0.01, 0.03, 0.03] & \lambda_3 &= [0.01, 0.03, 0.03] \\ \mathbf{R}_{crit} &= [5, 5, 5] \end{aligned} \tag{36}$$

The model adaptation algorithm, Equation 5, uses the set of *normalized* adjustable parameters to minimize the estimation criterion (i.e. the values shown in Table 4 are rearranged in order to have an order of magnitude near to 1). The upper and lower bounds

for the parameters are shown in the same table. The weighting factor for the estimation problem, $\mathbf{R}_p$, is once again the identity matrix.

The optimal decisions of the economic optimization problem, Equation 6, are also implemented with an exponential filter. The filter gain is the same for all inputs $\mathbf{K}_{opt} = 0.9$. Differently from the previous case study, the optimal values are directly implemented on the dynamic model that represents the plant (i.e. there is no MPC layer) with exception of R_{ai} . As shown in Figure 10, there is a control loop in the process. The PI controller uses the value of the air flowrate, R_{ai} , for maintaining the oxygen percentage in the flue gas at the optimum setpoint. Hence, instead of implementing the value of R_{ai} in the plant, the setpoint of O_{fg} is calculated with the SS model and this value is fed to the PI controller, which, in turn, implements the value of R_{ai} that matches the given setpoint.

ROPA

To estimate the parameters online, this implementation of ROPA uses the reduced rank extended Kalman filter (rEKF) (PHAM; VERRON; ROUBAUD, 1998). The rEKF is chosen over the EKF due to the large number of parameters. Instead of determining the adjustable parameter set beforehand, the rEKF is able to change this set according to the received plant information. This characteristic is in line with the Rotational Discrimination Method, which is the most promising estimation method for actual RTO applications (GRACIANO; MENDOZA; LE ROUX, 2014).

The Rotational Discrimination method decomposes the parameter space, using spectral decomposition, similarly to the rEKF implementation shown in Chapter 2. In a few words, it projects the least-square search space onto a subspace in which the estimation problem objective function decreases the most. This strategy is advantageous for RTO implementations because: it requires less a priori information regarding the determination of the adjustable parameter set; its implementation is simple; and it is able to avoid data overfitting (GRACIANO; MENDOZA; LE ROUX, 2014).

The same advantages are obtained by using rEKF in ROPA. For implementing this estimator, the system is linearized as in the previous case study. The states, measurement and parameter covariance matrix used throughout this chapter are:

$$\begin{aligned}\mathbf{Q} &= \text{diag}([1e^{-5}; 1e^{-5}; 1e^{-2}; 1e^{-5}; 1e^{-5}]) \\ \mathbf{Q}^p &= \text{diag}([1e^{-2}; 1e^{-2}; 1e^{-2}; 1e^{-2}; 1e^{-2}; 1e^{-2}; 1e^{-2}; 1e^{-2}; 1e^{-4}; 1e^{-3}; 1e^{-2}; 1e^{-2}]) \\ \mathbf{R} &= \text{diag}([1e^0; 1e^1; 1e^0; 1e^1])\end{aligned}\quad (37)$$

The values in the matrices diagonals correspond to entries of $\mathbf{x}$, $\mathbf{p}$ and $\mathbf{y}$, which were introduced in Equation (34). The period between two executions of ROPA, Δt_{ROPA} , is fixed in 30s, which is in the same order of magnitude of the catalyst residence time in the riser/reactor system.

4.4 Results and discussion

First, the dynamic model is simulated and compared to real plant data in order to verify the quality of the model response. After this validation step, which is shown in Appendix A, the economic performance of ROPA and MPA are compared using the gasoline production, while a dynamic model represents the plant in Figure 11.

Figure 11 – Comparison of the optimal gasoline production and the achieved by ROPA and MPA (Source: own elaboration)

Figure 11 shows the gasoline production profiles obtained with ROPA (continuous green line) and MPA (continuous blue line). They are compared to the optimum values (dashed black line). The optimal gasoline production is calculated by simply solving the economic-based optimization problem with the “true” parameter values. Both method-

ologies are simulated starting from the initial suboptimal point shown in Table 3 and they use the same sequence of noise, which was precomputed. Note that, the Monte Carlo Method is not applied here.

After a few iterations, the economic performance achieved by the ROPA cycle diverges from the optimal due to the fact that the parameter estimates deviate from the “true” plant values. The parameters are incorrectly estimated mainly because the initialization values of the estimate covariance matrix, $\mathbf{P}_k$. Note that, $\mathbf{P}_k$ is very difficult to be tuned in practice and has a major influence on the filter performance (SCHNEIDER; GEORGAKIS, 2013). Therefore, such situation is very likely to occur in real applications. However, ROPA runs continually every Δt_{ROPA} instants. Hence, it gradually corrects the values of the parameter estimates by updating the values of $\mathbf{P}_k$ with more plant information. The parameters estimates tend to their “true” values after a few ROPA iterations. The parameter profiles are shown in Figure 12. If the same scenario takes place in the MPA cycle (i.e. an erroneous parameter update), the system needs to reach another steady-state before any corrections on the parameter estimates and, consequently, on the optimal decisions.

At 2100 seconds, a ramp-like disturbance occurs in the cracking reaction kinetic. ROPA detects it while the disturbance is affecting the plant and already starts to update the model parameters and optimal decision, which avoids a significant loss of gasoline production. Around 6000 seconds, ROPA has already reached the optimal steady-state gasoline production while MPA takes approximately 10000 seconds. Note that, the gasoline production obtained with ROPA is higher than the optimal stationary production in some periods. The reason is the same as in the previous case study. The instantaneous production can be higher than the steady-state optimal production during a given period but, in a broader time horizon, the SS optimum is eventually reached. In addition, despite the fact that MPA does not adjust the optimal inputs, the gasoline production increases because of the nature of the parameter disturbance.

Regarding the parameters, Figure 12 shows the parameter profiles. Instead of plotting all parameters shown in Table 4, only the kinetic parameters of the carbon formation reaction are shown. The configuration of the figures is the same as in Figure 11, the parameters estimated by ROPA (continuous green line) and MPA (continuous blue line) are compared to the “actual” plant value (dashed black line).

Figure 12 – Comparison of the estimated values of the kinetic parameters of the carbon formation reaction. The parameters estimated by ROPA and MPA are compared to the “actual” plant values. Here, the parameter values are normalized in order to have similar orders of magnitude. (Source: own elaboration).

The figure illustrates an important advantages of ROPA over MPA. First, despite having a good parameter initial guess, the first MPA model adaptation run is unreliable, which leads to aggressive updates in the parameters estimates. This problem is related to identifiability issues, which stem from two different causes (GRACIANO; MENDOZA; LE ROUX, 2014): the structural form of the model and parameter estimation objective function; and insufficient measurements (i.e. few states are measured).

The first cause is related to the structural identifiability. These issues appear in the cases that a given set of system inputs/outputs does not map to a single parameter vector (LJUNG, 1979) (i.e. the parameter estimation problem does not have a local minimum at an isolated point). Therefore, several combination of parameters can lead to similar values of the objective functions, which can cause a great variability on the parameter estimates. On the other hand, the second cause, which is called practical identifiability, is related to the quantity/quality of information provided by the measurements. Both structural and practical identifiability issues can affect the parameter prediction to a great extent. Such behavior is worsened by the lower number of MPA executions and the fact that the SS detection method indicates an inexistent SS period around 4000 seconds (Figure 13).

In this case, the parameters are updated directly with transient data, which affects the quality of the estimates. In addition, in Figure 12, it is possible to notice the effects of rEKF in the system. Due to unreliable information, the online estimator does not update the value of k_{cc} , maintaining it on its nominal value. In turn, the parameters estimated by MPA tend to their constraint, possibly, due to the parameter inflation phenomena (GRACIANO; MENDOZA; LE ROUX, 2014).

Finally, Figure 13 shows the set of measurements used for calculating if the plant is at steady-state and the profile of the SS detection flag for MPA. In the case that the system is at steady-state, the flag is 1, otherwise it is 0.

Figure 13 – Steady-state detection results for MPA. The set of measurements that are analyzed to indicate the steady-state is $z_{ss} = [T_{rs}, T_{rg}, T_{fg}]^T$. (Source: own elaboration).

The SS detection method is able to indicate the initial and final steady-state periods. However, during the transient phase, it falsely indicates short SS periods. Despite undesired, such results are also very likely to occur in real applications. The main effect of this error is that the model adaptation problem (Equation 5), which is an optimization based on steady-state models, is fed with transient data. As a result, the data points cannot be represented by the SS model, which may influence the quality of the parameter estimates and, consequently, the model accuracy. This erroneous parameter update can be seen in

Figure 12. Clearly, the poorly performance of the SS detection method can be mitigated by adjusting the filter parameters.

4.5 Conclusion

Differently from the Chapter 3, ROPA is compared to MPA in a more realistic scenario. The model, which represents the plant, is more complex than in the previous case study and its nonlinearities pose a challenging optimization problem (ZANIN; TVRZSKA; ODLOAK, 2002). The results confirm that ROPA is able to track the steady-state optimum by continuously solving a steady-state economic problem adapted with transient measurements. However, if the transient data are directly used to update the steady-state model, the estimates are prone to diverge from the “true” parameter values. On the other hand, if online estimators are used, the results indicate that it is possible to avoid a significant loss of profit (gasoline production in this case). In addition, this chapter has shown a different advantage of ROPA. The possibility of correcting poor parameter estimates by increasing the optimization frequency.

The MPA cycle would also profit from the same advantage if there were many steady-state periods to trigger the model adaptation step. However, as discussed previously, in process like an FCC unit, the SS periods are very difficult to detect (or even inexistent). For example, in the data used for model validation, there is no clear characterization of steady-state periods during 3500 seconds of operation. Hence, the benefits to avoid the steady-state (SS) wait in the optimization cycle are clear in this process, which indicates a potential opportunity for applying the ROPA methodology to optimize the system instead of the standard RTO approach.

The next step for ROPA applications in FCC units is the exploitation of more complex models. Despite being able to reproduce the plant dynamics and temperatures, this model is much more simple than the FCC models used in commercial RTOs. These models are developed to represent with high accuracy the reaction effects, catalyst recirculation, feed quality, etc.. In real applications, the trade-offs between model complexity and filter accuracy need to be assessed. On one hand, the model should be complex enough to capture the main variables that affect the unit economic performance. On the other hand, the model complexity should allow the filter to infer the system states efficiently, within the sampling time of the controller. For example, solving a Moving Horizon Estimator

using a complex model is a challenging task to be carried out (RAO; RAWLINGS; MAYNE, 2003).

5 Case study 3: Plant-wide optimization

The best way to guarantee that the plant-wide optimum is achieved is by optimizing the whole plant. However, if the standard RTO methodology, Model Parameter Adaptation, is applied to the entire plant, a set of measurements needs to be determined in order to represent the current plant state (stationary or transient). Furthermore, the whole system needs to be at steady-state (SS) in order to initiate the RTO cycle. Depending on how the steady-state is defined, this condition is seldom found in practice.

The main goal of this case study is to illustrate the claim that ROPA is the key to decouple the estimation problem to obtain the plant-wide optimum. In order to do so, the stationary model of the complete plant is partitioned according to the disturbance pattern, which allows individual section of the model to be updated asynchronously. By decoupling the plant-wide estimation problem, online estimation is applied to the sections with higher disturbance frequency. As the model is continually up-to-date to the plant, the economic optimization for the entire plant can be scheduled at any desired rate, depending only on process settling time and computational constraints. The case study results indicate that ROPA is able to drive the complete system to the optimal plant-wide steady-state. In comparison to MPA, ROPA reacts much faster to the disturbances, improving the overall economic performance.

5.1 Process description

ROPA is implemented in a process composed by two stages. A separation stage, which is represented by a distillation column, and a reaction stage, represented by a reactor. A simplified flowsheet is shown in Figure 14. In order to evaluate the results of the decoupled estimation problem, both ROPA and the standard RTO (MPA) approach are applied to optimize the complete system. The disturbances affecting the system are shown in Figure 15. Clearly, the disturbances have different frequencies. While the column disturbance (changes in relative volatility, α) affects the process twice during the simulation, the reactor stage is affected by a sinusoidal and a ramp disturbance (changes in reaction rate constants, k_1 and k_2). Despite the fact that these disturbances are theoretical, they can reflect changes in column temperature and mixing problems inside the reactor, for example.

Figure 14 – Complete system containing a separation and a reaction stage. The model inputs (Q_v , Q_d and Q_{fr}) and adjustable parameters (α , k_1 and k_2) are highlighted. In addition, the process measurements are indicated with composition and flow meters (Source: own elaboration)

The different nature and frequency of disturbance influences the SS detection (SSD) methods. If the steady-state of the entire plant needs to be detected, it is clear that the disturbance affecting exclusively the reactor interferes in the global SS indicator. For example, between 200 and 500 minutes the column is already in steady-state but, due to the reactor disturbances, the complete system is not at steady-state and the RTO cycle does not initiate.

However, if the SS detection is applied only to the column, it is able to indicate that this process is at a steady-state. Therefore, as an alternative to the classical MPA scheme, the complete system is partitioned into two subsections, the reactor and column sections. For the reaction section, an online estimator (EKF) is used for estimating the parameters. On the other hand, SS detection methods are applied to the column system and whenever the system reaches steady-state, the parameters of the section are estimated. By using this asynchronous estimation strategy, the complete model is always updated to the plant. Thus, the plant can be optimized more frequently and the complete model can be used for achieving the plant-wide optimum without relying on decomposition techniques. The complete system model is presented in the next section.

Figure 15 – Disturbance profiles. α is the disturbance affecting the column and k_1, k_2 are the disturbances affecting the reactor (Source: own elaboration).

5.2 Process model

Distillation column

In order to simplify the model, some assumptions are made regarding the column behavior. Only two components enter the column, A and B . A is the lighter component, which has the lower boiling point, and most of A leaves the system as distillate. In turn, B is heavier and the major part of B leaves the column in the bottom product. The components are assumed to be similar, therefore the hypothesis of constant relative volatility holds. Also, the heats of vaporization do not differ significantly. Thus, constant molar flows are assumed along the column, which implies that the liquid accumulation is constant in the stages. Vapor holdup is considered negligible due to the low operating pressure. The distillation column has 4 stages: reboiler (stage 1), feed stage (stage 2); stage above the feed (stage 3); and total condenser (stage 4). The equilibrium is calculated by:

$$y_A(i) = \frac{\alpha x_A(i)}{1 + (\alpha - 1)x_A(i)} \quad i = 1, 2, 3. \quad (38)$$

where, $x_A(i)$ and $y_A(i)$ are the liquid and vapor molar fractions of A , respectively. α , which is the model adjustable parameter, is the relative volatility between A and B . Considering

that the molar heat of vaporization is equal for A and B , the mass and energy balances for all the stages lead to the following relations:

$$F_{v,2} = F_{v,3} = F_v \quad (39a)$$

$$F_{l,3} = F_l \quad (39b)$$

$$F_{l,2} = F_{l,3} + F_{fc} \quad (39c)$$

$$F_b = F_{l,2} - F_v \quad (39d)$$

$$F_d = F_{v,3} - F_l \quad (39e)$$

$$F_d = Q_d \rho, \quad F_v = Q_v \rho_{gas}, \quad F_{fc} = Q_{fc} \rho \quad (39f)$$

where, $F_{v,i}$ and $F_{l,i}$ are the vapor and liquid molar flows leaving stage i . F_d is the distillate (top product) and F_b , the bottom product molar flow. F_{fc} is the feed molar flow. ρ and ρ_{gas} are the molar densities, which are assumed to be constant. Q_v , Q_d and Q_{fc} are the volumetric flowrates of the vapor flow leaving the reboiler, distillate and feed, respectively. Q_v and Q_d are the manipulated variables for the distillation column and Q_{fc} is specified.

Figure 16 – Diagram of the column model described by Equations (39) and (40) (Source: own elaboration).

It is implicit in the relations above that the molar holdup is constant. Thus, the component molar balances for the system can be written as:

$$m_c(1) \frac{dx_A(1)}{dt} = F_{l,2}x_A(2) - F_v y_A(1) - F_b x_A(1) \quad (40a)$$

$$m_c(2) \frac{dx_A(2)}{dt} = F_{l,3}x_A(3) + F_v y_A(1) - F_{v,2}y_A(2) - F_{l,2}x_A(2) + F_{fc}x_{A,feed} \quad (40b)$$

$$m_c(3) \frac{dx_A(3)}{dt} = F_{l,4}x_A(4) + F_{v,2}y_A(2) - F_{v,3}y_A(3) - F_{l,3}x_A(3) \quad (40c)$$

$$m_c(4) \frac{dx_A(4)}{dt} = F_{v,3}y_A(3) - F_{l,4}x_A(4) - F_d x_A(4) \quad (40d)$$

where, $m_c(i)$ are the liquid molar holdup of stage i .

Reactor

The distillation column downstream process is a tank reactor. Three other components enter the system: C , D and E . A stream containing A and B (distillate from the column) is fed to the reactor as well as a second stream, Q_{fr} , containing only C . The product of interest is D and E is the undesired component. The set of reactions taking place in the system is:

Note that B is considered as an inert. The reactor model also has some simplifying assumptions. An isothermal continuous stirred-tank reactor with perfect control level is considered. Thus, the reaction rates are controlled by manipulating the system residence time. The reaction set provides an interesting situation for optimization. By increasing the residence time, the rate of reaction 1 increases producing more D , which is the product of interest. However, it increases the production of E by a larger extent, given that reaction 2 is a second order reaction in relation to C . Therefore, there exists an opportunity to apply RTO strategies in the reactor to maximize production of D without producing large quantities of E . The reactor mass balance is represented by:

$$Q_{out} = Q_d + Q_{fr} \quad (41)$$

where, Q_{out} is the reactor outlet, Q_d is the distillate defined previously, and Q_{fr} is the reactor feed, which contains pure C . Q_{fr} is the manipulated variable of the reactor system.

All flowrates are volumetric. Once again, due to the assumptions, the energy balance is not necessary. The system can be described by the following equations:

$$\frac{dC_A}{dt} = \frac{Q_d(\rho x_A(4)) - Q_{out}C_A}{V_R} - r_1 \quad (42a)$$

$$\frac{dC_C}{dt} = \frac{Q_{fr}C_{C,in} - Q_{out}C_C}{V_R} - r_1 - 2r_2 \quad (42b)$$

$$\frac{dC_D}{dt} = \frac{-Q_{out}C_D}{V_R} + r_1 \quad (42c)$$

$$\frac{dC_E}{dt} = \frac{-Q_{out}C_E}{V_R} + r_2 \quad (42d)$$

where, C_j is the concentration of the component j leaving the reactor, V_R is the reactor volume, ρ is the liquid molar density of the mixture, and $x_A(4)$ is the fraction of A leaving the distillation column at the top product. The reaction rates, r_1 and r_2 are defined as:

$$r_1 = k_1 C_A C_C \quad (43a)$$

$$r_2 = k_2 C_C^2 \quad (43b)$$

where, k_1 and k_2 are the rate constants of the two chemical reactions. Both are the model adjustable parameters.

Model Summary

Given the equations above, the complete system model states x , measurements y , decision variables u , and adjustable parameters p are:

$$\mathbf{x} = [x_A(1), x_A(2), x_A(3), x_A(4), C_A, C_C, C_D, C_E]^T \quad (44a)$$

$$\mathbf{y} = [x_A(4), C_A, C_C, C_D, C_E]^T \quad (44b)$$

$$\mathbf{u} = [Q_v, Q_d, Q_{fr}]^T \quad (44c)$$

$$\mathbf{p} = [\alpha, k_1, k_2]^T \quad (44d)$$

The inputs, initial state of the system and nominal parameter values are listed in Tables 5 and 6.

Table 5 – Initial plant condition for the complete system. The lower and upper bounds of the manipulated variables are shown because they are used in the economic optimization problem, Equation (6).

	Variable (symbol) [unit]	Lower bound	Initial value	Upper bound
–	Column feed flowrate (Q_{fc}) [L/min]	–	210.0	–
–	Feed fraction of A ($x_{A,feed}$) [–]	–	0.5	–
–	Concentration of C at Q_{fr} ($C_{C,in}$) [mol/L]	–	5.5	–
–	Liquid molar holdup of tray i ($m_c(i)$ $i = 1, 2, 3, 4$) [mol]	–	1	–
u_1	Reboiler vapor outflow (Q_v) [L/min]	0	90.00	150
u_2	Distillate flowrate (Q_d) [L/min]	0	12.00	30
u_3	Reactor feed flowrate (Q_{fr}) [L/min]	0	13.00	45
x_1	Liquid molar fractions of A at the 1 st column stage ($x_A(1)$) [–]	–	0.1424	–
x_2	Liquid molar fractions of A at the 2 nd column stage ($x_A(2)$) [–]	–	0.3991	–
x_3	Liquid molar fractions of A at the 3 rd column stage ($x_A(3)$) [–]	–	0.7328	–
x_4/y_1	Liquid molar fractions of A at the 4 th column stage ($x_A(4)$) [–]	–	0.9291	–
x_5/y_2	Concentration of A in the reactor (C_A) [mol/L]	–	1.100	–
x_6/y_3	Concentration of C in the reactor (C_C) [mol/L]	–	0.1500	–
x_7/y_4	Concentration of D in the reactor (C_D) [mol/L]	–	2.000	–
x_8/y_5	Concentration of E in the reactor (C_E) [mol/L]	–	0.5000	–

Table 6 – Adjustable parameter set of the complete system model. The lower and upper bounds are arbitrarily determined based on the nominal values of the parameters.

	Variable (symbol) [unit]	Lower bound	Nominal value	Upper bound
p_1	Relative volatility between A and B (α) [–]	2	6	8
p_2	Rate constant of reaction 1 (k_1) [L/mol min]	0.5	0.75	3
p_3	Rate constant of reaction 2 (k_2) [L/mol min]	0.5	1.5	3

5.3 Case study

The objective of the optimization is to maximize the complete system profit, which is measured by the following equation:

$$\phi = 1Q_{out}C_D + 0.1Q_b - (0.05Q_d + 0.05Q_v + 0.05Q_{fr}); \quad (45)$$

All the variables have been previously defined. The numeric values are related to the market price of the two valuable products, which are B and D . In this scenario, the purity of B at the bottom product is not important and the reactor outflow of B has no market price. On the other hand, as D has to be separated from the other reaction subproducts, the concentration C_D influences the market price of the reactor outflow stream. The costs of the unit are related to the consumption of vapor at the reboiler and cooling water at the condenser, which are considered as proportional to Q_v and Q_d . Additionally, the feed of C to the reactor has its own related cost. The column feed, Q_{fc} , also needs to be purchased, however, as it is constant during the simulation, it does not appear on Equation (45). There is one operational constraint that plays an important role in the optimization problem. The reactor has a maximum processing capacity of 25

[L/min]. Due to the nature of the economic objective function, this constraint is always active. Hence, the problem does not have three degrees of freedom, but two.

Given that objective function and operational constraint, the process cannot be decoupled. In order to optimize only the reactor, it would be necessary to determine the distillate composition and its intermediary price, which is not an easy task because this stream has no market price (FRIEDMAN, 1995). Furthermore, depending on how this intermediary price is determined, the optimization of the column does not reflect the plant-wide optimum. This scenario is the motivation to use ROPA together with a partitioned model, applying the online estimation to a subsections of the system.

In order to make the simulation scenario more realistic, noise is added to the plant measurements, $\mathbf{y}$. The noise sequence is obtained by adding zero-mean noise within the range of 0.1% of the measurement value (as described in the first case study in Chapter 3). The noise is neither correlated between measurements nor in time. The system is also affected by deterministic disturbances, which were already shown in Figure 15. The plant measurements are obtained as the solution of the dynamic model described above. The steady-state model used in the optimization and model adaptation steps is directly derived from the dynamic model, i.e. there is no plant-model mismatch. No control layer is added to the system, the optimization results are directly fed to the process as values for the manipulated variables.

Model parameter adaptation and ROPA

Table 7 shows a comparison between the implementation of ROPA and MPA in this case study.

Table 7 – Comparison of ROPA and MPA implementations

	MPA	ROPA
<i>Control volume</i>	Complete system (reactor + column)	Complete system (reactor + column) Online estimation: Reactor
<i>Optimization</i>	Steady-state (SS) optimization with Equation (45) as objective function	
<i>Parameter estimation</i>	SS estimation of α , k_1 and k_2	SS estimation to adjust α . k_1 and k_2 are estimated online
SS Detection	Entire system needs to be in steady-state to trigger the optimization cycle $z_{ss} = [x_A(4), C_A, C_C]^T$	Only the column steady-state is analyzed in order to update the column's parameter. The optimization cycle is trigger every Δt_{ROPA} min, independently of the column condition. $z_{ss} = x_A(4)$

The SS parameter estimation, optimization, and SS detection methods are the same as in the previous case studies. On the other hand, instead of using rEKF as the online estimator, EKF is chosen. The filter uses the inputs and measured values at the outlet of the reactor (C_A , C_C , C_D and C_E) to estimate the reactor parameters (k_1 and k_2). Table 8 shows the tuning parameters for MPA and ROPA.

Table 8 – Parameters of the extended Kalman filter (EFK) and the steady-state detection (SSD) method. MPA uses the measurements of $x_A(4)$, C_A and C_C to indicate SS, while only the column SS behavior ($x_A(4)$) is analyzed in the ROPA cycle. Thus, only the first value of λ_1 , λ_2 , λ_3 and $\mathbf{R}_{crit}$ is used in ROPA.

	Variable (symbol) [unit]	value
Δt_{ROPA}	ROPA frequency [min]	30
λ_1	mean filter for SSD	[0.2;0.1;0.2]
λ_2	variance (trend) filter for SSD	[0.1;0.05;0.1]
λ_3	variance (time) filter for SSD	[0.1;0.05;0.1]
$\mathbf{R}_{crit}$	critical threshold for variance relation	[2.5;3.9;2.5]
$\mathbf{K}_{opt}$	filter for the optimization inputs	0.5
$\mathbf{Q}$	process noise variance	diag([0.1,0.1,0.1,0.1])
$\mathbf{Q}^p$	parameter noise variance	diag([10,10])
$\mathbf{R}$	measurement noise variance	diag([1,1,1,1])

5.4 Results and discussion

The first analysis addresses the steady-state detection and, as a direct consequence, the number of model updates. As described in Table 7, both MPA and ROPA use the steady-state model adaptation algorithm (Equation (5)) in order to update the model. However, while in MPA the complete system model is updated using this strategy, in ROPA only the column parameters are estimated using steady-state measurements. The reactor model is updated via EKF using transient data.

Figures 17 and 18 show the profile of the steady-state detection flag for MPA and ROPA, respectively. If the flag is 1, the system is considered at steady-state. 0 indicates a transient state. In addition, the set of measurements that is used to represent if the plant is at steady-state is shown. In MPA, the three measurements are applied in the steady-state detection algorithm. In turn, only the column measurement, $x_A(4)$, is used to indicate if the column is at steady-state in ROPA. In Figure 18, the other measurements C_A and C_C are presented in order to illustrate that, despite the fact that the entire unit is at a transient state, the column is at steady-state. Therefore, its parameters can be estimated using the

steady-state estimation strategy. Note that the measurements profiles are different for the MPA and ROPA simulations, because each method computes different optimal inputs in face of the disturbances.

Figure 17 – Steady-steady detection results for MPA. In this case, the entire system needs to be in steady-state in order to trigger the optimization cycle. The set of measurements that are analyzed to indicate the steady-state ($z_{ss} = [x_A(4), C_A, C_C]^T$) are also plotted. (Source: own elaboration).

Figure 18 – Steady-steady detection results for ROPA. In this case, only the column parameters are updated using the steady-state model adaptation (Equation (5)), hence only the column measurements are analyzed $z_{ss} = [x_A(4)]$. However, the values of C_A and C_C are also presented in order to show that they do not have influence on the column SS detection (Source: own elaboration).

Due to the nature of the reactor disturbances, MPA is able to detect only four short steady-state periods during the simulation, leading to four optimization runs. On

the other hand, as ROPA SS detection takes into account just $x_A(4)$, it is able to track the column disturbances faster and the SS detection flag represents its behavior accurately (i.e. it shows if the column is at a steady or a transient state). As a consequence, the column model is updated more frequently, which allows the optimization to run every Δt_{ROPA} instants. Such difference in the updating police has a direct effect on the parameter estimates profiles.

Figure 19 presents the estimated values of the system parameters by the two methodologies. The ROPA estimated parameters are plotted in red, while the MPA estimated parameters are plotted in blue. In addition, the actual parameter values are shown as a dashed black line.

Figure 19 – Comparison of the parameter estimates. The red line represents the values estimated by ROPA and the blue line, the values estimated by MPA. Despite being able to determine the correct values for the column, MPA has an inherent delay for estimating the parameters due to the steady-state wait. In turn, the reactor parameter changes are not detected by MPA (Source: own elaboration).

Figure 19 shows that ROPA is already able to update the column parameters around 150 minutes. Due to the influence of the reactor measurements in the SS detection algorithm, MPA takes about 180 minutes to accomplish the same task. Next, the reactor

is affected by the sinusoidal disturbance starting at 200 minutes. As shown in Figure 17, this disturbance hinders MPA to initiate the optimization because the whole system needs to be at steady-state to trigger the cycle. Thus, the reactor parameter changes are not detected by MPA. On the other hand, by applying the EKF, ROPA detects the reactor disturbance and maintains the parameter values almost up-to-date to the plant.

MPA updates the model and optimizes the system again only after the reactor disturbance ceases. Despite the fact that there are no great adjustments in the column parameter, MPA is able to determine the correct values of α , k_1 and k_2 . Then, two disturbances affect the system after 700 min, α returns to its nominal value and a ramp disturbance starts to act on the reactor. ROPA is already able to detect the column disturbance at 750 min. In turn, MPA does not update the column parameter until 1130 minutes. This inherent SS-wait delay in MPA has a significant effect on the unit profit as it can be seen in Figures 20 and 21.

Figure 20 – Comparison of the instantaneous profit. The red line represents the values obtained with ROPA and the blue line, the values obtained with MPA. The dashed line is the optimum steady-state profit (Source: own elaboration).

Clearly, the profit achieved by the two methodologies is directly influenced by the SS-wait. Regarding the instantaneous profit (Figure 20), both methodologies start at a suboptimal point. However, ROPA identifies the disturbance and acts faster in comparison to MPA. Thus, even though ROPA uses steady-state estimating methods for α , the

Figure 21 – Comparison of the profit difference obtained with the different methodologies. Profit difference is calculated by subtracting the value of the instantaneous profit obtained with ROPA from the profit obtained with MPA. A line for the 0 (i.e. no difference) is added to improve readability. (Source: own elaboration).

decoupling of the plant-wide estimation strategy increases the updating frequency of the column, which improves the economic results.

For the sinusoidal disturbance, the SS-wait hinders the MPA cycle to start. As MPA is unable to adapt the model parameters, the performance is always suboptimal in this region. For the ramp disturbance, the effects of the delay are not so extreme, but ROPA has also a better economic performance. MPA is able to achieve the optimal performance only around the end of the simulation at 1130 minutes, when the process reaches the steady-state and the disturbances stop affecting the process. The same results are supported by the differences between ROPA and MPA profit (Figure 21). ROPA outperforms MPA in almost every instant during the simulation. The profit difference decreases every time that the MPA cycle starts, reaching 0 towards the end of the simulation.

The cumulative profit loss associated with the SS-wait, which is inherent with the MPA implementation, is representative for this process. Note that ROPA also suffers with a delay. It needs to wait Δt_{ROPA} instants to act. However, as this period is chosen based on the process settling time, the delay effects are smaller. Both Figures 20 and 21 support the claim that by decoupling the estimation problem, the disturbance detection is improved and ROPA is able to drive the complete system to the optimal plant-wide steady-state much faster. Clearly, determining which section of the complete process is update with the

online estimator is a key feature for a successful implementation. If the scope increases, the modeling and computational burdens also increase. Process knowledge should be used in order to determine the sections with higher disturbance frequency.

5.5 Conclusion

The final case study demonstrates the most important advantage of the proposed methodology, which is the decoupling of the plant-wide estimation problem. By dividing the estimation problem into smaller subproblems, ROPA assures that the whole plant model is up-to-date to the plant data, allowing the complete plant model to be used in the economic optimization problem, Equation (6). As a direct consequence, the optimization results reflect the plant-wide optimum.

Clearly, one issue is how to divide the estimation problem. In the case study, the parameter variation is known. Hence, it is easier to determine the scopes of the online estimation and steady-state estimation. On the other hand, in actual implementations, how to divide the problem in subproblems is not so intuitive. One alternative is to analyze the past plant measurement offline and determine the subsections by applying different SS detection methods to the data in order to identify the sections with longer, more frequent and more marked steady-states periods. Process knowledge is also a helpful tool to identify the subsections.

A second alternative to achieve plant-wide optimal results is the application of decomposition techniques (FRIEDMAN, 1995). These techniques are based on estimating intermediary products prices using shadow prices of the planning layer, which is part of the plant decision hierarchy shown in Figure 1. Despite appealing, there is no wide acceptance of these techniques in the RTO community (ACEVEDO et al., 2015). Furthermore, in order to optimize a single unit, not only the intermediary prices should be determined but also the intermediary product composition, which, depending on the process, further complicates this task.

The case study has also shown that ROPA is able to track disturbances of different natures (ramp, step and sinusoidal), and, as the subsystems are divided based on the disturbance frequency, the column model is updated more frequently in the ROPA implementation, even though it also uses steady-state updating methods. Due to the

more frequent optimization, ROPA response to the plant disturbances is faster and the cumulative profit obtained is larger than with MPA.

As discussed in Chapter 2, both methods that calculate optimal trajectories, like DRTO and EMPC, and ROPA use dynamic models to estimate system states and parameters. Unfortunately, there is a lack of reported applications of large-scale dynamic models in estimation problems. In this context, an extra advantage of the partitioned methodology appears. The estimation problem of a large-scale model can be decoupled and, instead of estimating all model parameters online, only a subset of them is updated in this manner, which decreases the size of the estimated parameter set. Hence, fewer phenomena occurring in the process need to be modeled dynamically. Despite the fact that this strategy has not been tested, the success of the partitioned strategy presented in this chapter encourages developments towards this direction.

6 Concluding remarks and future work recommendations

A new methodology of Real-time Optimization (RTO) named Real-time Optimization with Persistent Adaptation (ROPA), which associates online estimators to the static economic optimizer was proposed in this thesis. As a consequence, the steady-state (SS) wait, which is one of the main drawbacks of the current RTO implementation, is reduced. Despite the numerous benefits of RTO, it still finds split acceptance in industry (DARBY et al., 2011). Therefore, addressing the challenges for deploying and maintaining RTO is a meaningful contribution, not only for the academic literature but also for industrial purposes. Note that, although based on sophisticated concepts, the proposed solution derives from a simple idea, which makes it even more attractive for both academia and industry.

In general, the adaptation of a steady-state model with transient measurements might appear counterintuitive. However, ROPA underlying assumption is that it is better to treat the transient parameters as if they are stationary and improve the plant operation continuously, instead of waiting for a (possible) steady-state condition. Note that ROPA provides an intermediary solution between static and dynamic optimization schemes, like Dynamic Real Time Optimization (DRTO) and Economic Model Predictive Control (EMPC).

An important advantage of ROPA over schemes that calculate the optimal trajectory is the difficulty associated with creating dynamic models that encompass the whole plant (and estimating its parameters) in order to assure plant-wide optimal results. Because ROPA employs SS models for the economic optimization, a steady-state model can be developed to the entire unit and, then, partitioned. This approach allows individual sections of the model to be updated asynchronously using either online or steady-state estimation. The choice of the estimation method depends on the characteristics of each individual section, like disturbance frequency. As the global plant model is continuously updated to the plant, the plant-wide economic optimization can be scheduled at any desired rate. Thus, ROPA is the key for decoupling the plant-wide estimation problem.

The new methodology benefits are illustrated with three case studies: (1) a Williams-Otto reactor, which is used for showing the convergence properties of ROPA; (2) a Fluid Catalytic Cracking (FCC) unit, in which ROPA is applied to a more realistic process in order to show some practical advantages of the new methodology. The FCC model is

previously compared and validated with actual data from the FCC located at Refinaria Henrique Lage (REVAP), São José dos Campos, Brazil and owned by Petrobras S.A.; and (3) a system composed by a distillation column and a reactor that illustrates the use of ROPA to obtain the plant-wide optimum.

In the three case studies, ROPA performance is compared to the standard RTO method, Model Parameter Adaptation (MPA). The results show that the increase of the model updating frequency (i.e. decrease of the time between two sequential optimization runs) improves the prediction capacity of the optimization cycle, which leads to a better economic performance and enhances the disturbance detection of the optimization cycle. Furthermore, by applying ROPA to subsections of the model, the plant-wide optimum is successfully obtained.

6.1 Complementary work

In Appendix B, the Dynamic Design using Proportional Action (DDPA) method is proposed. The method approximates the dynamics of vapor-liquid equilibrium systems by a simple linear relation, generating a reduced-index model. This allows the use of ODE solvers and, more important, the application of any thermodynamic package embedded in equation-oriented environments. DDPA capability for representing the dynamics of a real process, a depropenizer column, is shown. The simulation results are compared to real plant data during a long operation period, showing that DDPA is able to reproduce the dominant (fast and slow) dynamics of the column with sufficient accuracy. DDPA was first developed in order to allow the application of ROPA to the depropenizer column. However, it was not possible to obtain the sensitivities in this case, which hindered the joint application of DDPA and ROPA.

Additionally, a new method that combines Modifier Adaptation (MA), an RTO variant, with an online model structure adaptation method is proposed in Appendix C. Without deriving a fixed model for the system, the novel approach is able to determine the best model structure (i.e. more consistent with plant data) online by combining different sub-models while optimizing the process to the actual plant optimum. The competing sub-models are compared using the modifiers of MA as criterion. The novel method is tested in a gas lifted oil well network and it is able to determine the overall model structure that best fits the plant in face of noisy measurements, while maximizing the production of

the oil well network. The most important contribution of this work is that, by allowing that the model evolve over time, it can be a starting point for changing the paradigm that the model structure is an immutable entity.

6.2 Directions for further work

- *More case studies* - More examples should be studied in order to test the performance of ROPA in different scenarios. Focus should be on large scale systems and cases where the potential gain of a plant-wide optimization is significant (units competing for resources, for example).
- *Combining DDPA and ROPA* - Future works could involve implementing online estimation (extended Kalman filter, moving horizon estimation, etc.) using the dynamic model developed with DDPA. Then, the online estimator could be used in order to apply ROPA.
- *Model uncertainty* - The effect of model uncertainty regarding the online estimators applied in ROPA was not studied. When selecting the dynamic model to be used in the estimator, the occurrence of plant-model mismatch should be taken into account.
- *Assessing the effect of Δt_{ROPA}* - Another interesting development for ROPA is the comprehension of the effects of different Δt_{ROPA} in the ROPA performance. It is important to understand how the execution period can destabilize the plant and how much the fine tuning of this parameter can improve ROPA performance. Furthermore, it would also be interesting to evaluate the effects of the Δt_{ROPA} changing during the simulations, according to the disturbance (i.e not determining Δt_{ROPA} beforehand, but letting it be set according to the plant conditions).
- *Using the decoupled estimation strategy of Chapter 5 in a large-scale model* - As described in the Conclusion section of Chapter 5, the asynchronous model updating strategy could be tested in order to decrease the number of parameters that are estimated online in a large-scale estimation problem.
- *Plant-wide optimization* - The comparison among decomposing techniques and the ROPA is necessary to assess the benefits of the methodology. Comparing the results will offer insights regarding the benefits of the decoupled estimation problem.
- *Model structure adaptation* - It is necessary to analyze the effect of different objective functions on the model structure adaptation problem. The efficiency of method

depends on the objective function. It is not clear to which extent the methodology is robust to noise measurements and the addition of a large number of blocks. Despite the evidences provided by the case study, the technical analysis regarding robustness should be extended. The convergence of the method needs to be assessed formally.

Bibliography

ACEVEDO, A. M. P.; GRACIANO, J. E. A.; LIPORACE, F. D. S.; VIANNA, A. S.; LE ROUX, G. A. C. Decomposition techniques for the real-time optimization of a propylene production unit. *Computer Aided Chemical Engineering*, v. 37, p. 893–8, 2015. Cited 3 times on pages 27, 32, and 90.

ADETOLA, V. A. *Integrated real-time optimization and model predictive control under parametric uncertainties*. Thesis (PhD) — Queen’s University, 2008. Cited 2 times on pages 23 and 26.

ALBET, J.; LANN, J. M. L.; JOULIA, X.; KOEHRET, B. Operational policies for the start-up of batch reactive distillation column. In: ESCAPE. *Proceedings of the 4th European Symposium on Computer Aided Process Engineering*. [S.l.], 1994. v. 133, p. 63–63. Cited on page 117.

AMRIT, R.; RAWLINGS, J. B.; ANGELI, D. Economic optimization using model predictive control with a terminal cost. *Annual Reviews in Control*, v. 35, n. 2, p. 178–186, 2011. Cited on page 35.

ANDERSSON, J.; ÅKESSON, J.; DIEHL, M. Casadi: A symbolic package for automatic differentiation and optimal control. In: *Recent advances in algorithmic differentiation*. [S.l.]: Springer, 2012. p. 297–307. Cited on page 112.

ASCHER, U. M.; PETZOLD, L. R. *Computer methods for ordinary differential equations and differential-algebraic equations*. [S.l.]: Siam, 1998. Cited on page 120.

BADER, J.-M.; GUESNEUX, S. Use real-time optimization for low-sulfur gasoline production. *Hydrocarbon processing*, v. 86, n. 2, 2007. Cited on page 22.

BARD, Y. *Nonlinear parameter estimation*. [S.l.]: Academic press, 1974. Cited on page 55.

BEERS, K. J. *Numerical methods for chemical engineering: applications in Matlab*. [S.l.]: Cambridge University Press, 2006. Cited 3 times on pages 117, 118, and 119.

BEQUETTE, B. W. *Process control: modeling, design, and simulation*. [S.l.]: Prentice Hall Professional, 2003. Cited 2 times on pages 48 and 125.

BESL, H.; KOSSMAN, W.; CROWE, T.; CARACOTSIOS, M. Nontraditional optimization for isom unit improves profits. *Oil and Gas Journal*, v. 96, n. 19, 1998. Cited 2 times on pages 25 and 33.

BIEGLER, L.; YANG, X.; FISCHER, G. Advances in sensitivity-based nonlinear model predictive control and dynamic real-time optimization. *Journal of Process Control*, v. 30, p. 104–116, 2015. Cited on page 40.

BINDER, T.; BLANK, L.; DAHMEN, W.; MARQUARDT, W. Multiscale concepts for moving horizon optimization. In: *Online Optimization of Large Scale Systems*. [S.l.]: Springer, 2001. p. 341–361. Cited on page 35.

BONVIN, D.; GEORGAKIS, C.; PANTELIDES, C. C.; BAROLO, M.; GROVER, M. A.; RODRIGUES, D.; SCHNEIDER, R.; DOCHAIN, D. Linking models and experiments. *Industrial and Engineering Chemistry Research*, v. 55, n. 25, p. 6891–6903, 2016. Cited on page 125.

BRENAN, K. E.; CAMPBELL, S. L.; PETZOLD, L. R. *Numerical solution of initial-value problems in differential-algebraic equations*. [S.l.]: Siam, 1996. Cited on page 117.

CAO, S.; RHINEHART, R. R. An efficient method for on-line identification of steady state. *Journal of Process Control*, v. 5, p. 363–374, 1995. Cited on page 24.

CAO, S.; RHINEHART, R. R. Critical values for a steady-state identifier. *Journal of Process Control*, v. 7, n. 2, p. 149–152, 1997. Cited 3 times on pages 24, 52, and 111.

CHACHUAT, B.; SRINIVASAN, B.; BONVIN, D. Adaptation strategies for real-time optimization. *Computers and Chemical Engineering*, v. 33, n. 10, p. 1557–1567, 2009. Cited on page 38.

CHEN, X.; HEIDARINEJAD, M.; LIU, J.; CHRISTOFIDES, P. D. Distributed economic MPC: Application to a nonlinear chemical process network. *Journal of Process Control*, v. 22, n. 4, p. 689–699, 2012. Cited on page 35.

CRAMER, R.; PETERS, R. Shell pecten cameroon integrates software for real-time gas lift control, optimization. *Oil and Gas Journal*, v. 100, n. 34, p. 52–54, 2002. Cited on page 23.

CUTLER, C.; PERRY, R. Real time optimization with multivariable control is required to maximize profits. *Computers and Chemical Engineering*, v. 7, n. 5, p. 663–667, 1983. Cited 2 times on pages 22 and 23.

DAOUTIDIS, P. DAEs in model reduction of chemical processes: An overview. In: *Surveys in Differential-Algebraic Equations II*. [S.l.]: Springer, 2015. p. 69–102. Cited on page 118.

DARBY, M. L.; NIKOLAOU, M.; JONES, J.; NICHOLSON, D. RTO: An overview and assessment of current practice. *Journal of Process Control*, v. 21, n. 6, p. 874–884, 2011. Cited 7 times on pages 23, 24, 25, 26, 27, 92, and 129.

DENN, M. M. *Process modeling*. [S.l.]: Longman Scientific and Technical, 1986. Cited on page 66.

EBRAHIMPOUR, M. *Dynamic and Steady-State Modeling of VRD Column in Equation-Oriented Environment*. Dissertation (Master) — Universidade de São Paulo, 2015. Cited 2 times on pages 120 and 121.

ELICECHE, A. M.; PETRACCI, N. C.; HOCH, P.; BRIGNOLE, E. A. Optimal operation of an ethylene plant at variable feed conditions. *Computers and Chemical Engineering*, v. 19, p. 223–228, 1995. Cited on page 22.

ELLIS, M.; CHRISTOFIDES, P. D. Integrating dynamic economic optimization and model predictive control for optimal operation of nonlinear process systems. *Control Engineering Practice*, v. 22, p. 242–251, 2014. Cited on page 35.

ELLIS, R. C.; LI, X.; RIGGS, J. B. Modeling and optimization of a model IV fluidized catalytic cracking unit. *AIChE Journal*, v. 44, n. 9, p. 2068–2079, 1998. Cited on page 22.

ENGELL, S. Feedback control for optimal process operation. *Journal of Process Control*, v. 17, n. 3, p. 203–219, 2007. Cited 2 times on pages 26 and 40.

FORBES, J.; MARLIN, T. E. Design cost: a systematic approach to technology selection for model-based real-time optimization systems. *Computers and Chemical Engineering*, v. 20, n. 6, p. 717–734, 1996. Cited 3 times on pages 47, 49, and 50.

FRANÇOIS, G.; BONVIN, D. Use of transient measurements for the optimization of steady-state performance via modifier adaptation. *Industrial and Engineering Chemistry Research*, v. 53, n. 13, p. 5148–5159, 2013. Cited 2 times on pages 34 and 52.

FRIEDMAN, Y. Z. What’s wrong with unit closed loop optimization? *Hydrocarbon Processing*, v. 74, n. 10, p. 107–116, 1995. Cited 6 times on pages 26, 27, 32, 35, 83, and 90.

GANI, R.; RUIZ, C. A.; CAMERON, I. T. A generalized model for distillation columns I: Model description and applications. *Computers and Chemical Engineering*, v. 10, n. 3, p. 181–198, 1986. Cited on page 117.

GAO, W.; WENZEL, S.; ENGELL, S. A reliable modifier-adaptation strategy for real-time optimization. *Computers and Chemical Engineering*, v. 91, p. 318–328, 2016. Cited 2 times on pages 126 and 130.

GEORGIOU, A.; SAPRE, A. V.; TAYLOR, P.; GALLOWAY, R. E.; CASEY, L. K. Ethylene optimization system reaps operations and maintenance benefits. *Oil and Gas Journal*, v. 96, n. 10, 1998. Cited on page 22.

GONZÁLEZ, A. H.; ODLOAK, D. A stable MPC with zone control. *Journal of Process Control*, v. 19, n. 1, p. 110–122, 2009. Cited on page 33.

GRACIANO, J.; MENDOZA, D.; LE ROUX, G. Performance comparison of parameter estimation techniques for unidentifiable models. *Computers and Chemical Engineering*, v. 64, p. 24–40, 2014. Cited 7 times on pages 43, 44, 57, 63, 69, 72, and 73.

GRACIANO, J. E.; MENDOZA, D. F.; LE ROUX, G. A. C. Assessing the reliability of different real-time optimization methodologies using monte carlo analysis. In: *AIChE. AIChE Annual Meeting Conference*. [S.l.], 2013. Cited 3 times on pages 47, 49, and 52.

GRACIANO, J. E. A. *Real time optimization in chemical process: evaluation of strategies, improvements and industrial application*. Thesis (PhD) — Universidade de São Paulo, 2015. Cited 2 times on pages 53 and 57.

GRÜNE, L. Economic receding horizon control without terminal constraints. *Automatica*, v. 49, n. 3, p. 725–734, 2013. Cited 2 times on pages 35 and 55.

GUDI, R. D.; SHAH, S. L.; GRAY, M. R. Multirate state and parameter estimation in an antibiotic fermentation with delayed measurements. *Biotechnology and Bioengineering*, v. 44, n. 11, p. 1271–1278, 1994. Cited on page 40.

HASELTINE, E. L.; RAWLINGS, J. B. Critical evaluation of extended kalman filtering and moving-horizon estimation. *Industrial and Engineering Chemistry Research*, v. 44, n. 8, p. 2451–2460, 2005. Cited on page 40.

- HELBIG, A.; ABEL, O.; MARQUARDT, W. Model predictive control for online optimization of semi-batch reactors. In: IEEE. *Proceedings of the American Control Conference*. [S.l.], 1998. v. 3, p. 1695–1699. Cited on page 22.
- HELBIG, A.; ABEL, O.; MARQUARDT, W. Structural concepts for optimization based control of transient processes. In: *Nonlinear Model Predictive Control*. [S.l.]: Springer, 2000. p. 295–311. Cited on page 35.
- HINDMARSH, A. C.; BROWN, P. N.; GRANT, K. E.; LEE, S. L.; SERBAN, R.; SHUMAKER, D. E.; WOODWARD, C. S. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*, v. 31, n. 3, p. 363–396, 2005. Cited 2 times on pages 42 and 118.
- HUANG, R. *Nonlinear model predictive control and dynamic real time optimization for large-scale processes*. Thesis (PhD) — Carnegie Mellon University, 2010. Cited on page 23.
- HUANG, R.; BIEGLER, L. T.; PATWARDHAN, S. C. Fast offset-free nonlinear model predictive control based on moving horizon estimation. *Industrial and Engineering Chemistry Research*, v. 49, n. 17, p. 7882–7890, 2010. Cited on page 26.
- HUANG, R.; HARINATH, E.; BIEGLER, L. T. Lyapunov stability of economically oriented NMPC for cyclic processes. *Journal of Process Control*, v. 21, n. 4, p. 501–509, 2011. Cited 2 times on pages 26 and 35.
- JIA, G.; STEPHANOPOULOS, G.; GUNAWAN, R. Incremental parameter estimation of kinetic metabolic network models. *BioMed central systems biology*, v. 6, n. 1, p. 142, 2012. Cited on page 125.
- JIANG, T.; CHEN, B.; HE, X. Industrial application of wavelet transform to the on-line prediction of side draw qualities of crude unit. *Computers and Chemical Engineering*, v. 24, n. 2, p. 507–512, 2000. Cited on page 25.
- JIANG, T.; CHEN, B.; HE, X.; STUART, P. Application of steady-state detection method based on wavelet transform. *Computers and Chemical Engineering*, v. 27, n. 4, p. 569–578, 2003. Cited on page 25.
- JOGWAR, S. S.; DAOUIDIS, P. Vapor recompression distillation: Multi-scale dynamics and control. In: IEEE. *Proceedings of the American Control Conference*. [S.l.], 2009. p. 647–652. Cited 2 times on pages 117 and 120.
- JULIER, S. J.; UHLMANN, J. K. Unscented filtering and nonlinear estimation. In: IEEE. *Proceedings of the IEEE*. [S.l.], 2004. v. 92, n. 3, p. 401–422. Cited on page 40.
- KADAM, J. V.; MARQUARDT, W. *Integration of economical optimization and control for intentionally transient process operation*. [S.l.]: Springer, 2007. Cited on page 35.
- KIPARISSIDES, C.; SEFERLIS, P.; MOURIKAS, G.; MORRIS, A. J. Online optimizing control of molecular weight properties in batch free-radical polymerization reactors. *Industrial and Engineering Chemistry Research*, v. 41, n. 24, p. 6120–6131, 2002. Cited on page 22.

- KORBE, M.; BELLEC, S.; JIANG, T.; STUART, P. Steady state identification for on-line data reconciliation based on wavelet transform and filtering. *Computers and Chemical Engineering*, v. 63, p. 206–218, 2014. Cited on page 25.
- KRISHNAMOORTHY, D.; FOSS, B.; SKOGESTAD, S. Real-time optimization under uncertainty applied to a gas lifted well network. *Processes*, v. 4, n. 4, p. 52, 2016. Cited 3 times on pages 128, 129, and 130.
- KURIHARA, H. *Optimal control of fluid catalytic cracking processes*. Thesis (PhD) — Massachusetts Institute of Technology, 1967. Cited 4 times on pages 66, 67, 107, and 111.
- LE ROUX, G. A. C.; SANTORO, B. F.; SOTELO, F.; TEISSIER, M.; JOULIA, X. Improving steady-state identification. In: ESCAPE. *Proceedings of European Symposium on Computer Aided Process Engineering*. [S.l.], 2008. p. 459–464. Cited on page 25.
- LEGGETT, P. D.; TEETER, E. A.; STOUT, R. M. Olefin plant revamp affects process control, real-time optimization. *Oil and Gas Journal*, v. 98, n. 19, p. 76–79, 2000. Cited on page 22.
- LEVY, R. E.; FOSS, A. S.; GRENS, E. A. Response modes of a binary distillation column. *Industrial and Engineering Chemistry Fundamentals*, v. 8, n. 4, p. 765–776, 1969. Cited on page 121.
- LID, T.; STRAND, S. Real-time optimization of a cat cracker unit. *Computers and Chemical Engineering*, v. 21, p. 887–892, 1997. Cited on page 22.
- LIMA, F. V.; RAWLINGS, J. B. Nonlinear stochastic modeling to improve state estimation in process monitoring and control. *AIChE Journal*, v. 57, p. 996–1007, 2011. Cited on page 40.
- LJUNG, L. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, v. 24, n. 1, p. 36–50, 1979. Cited 2 times on pages 45 and 72.
- LOCKETT, M. J. *Distillation tray fundamentals*. [S.l.]: Cambridge University, 1986. Cited on page 117.
- LOURENÇO, M.; SECCHI, A. R. Solução de sistemas de equações algébrico-diferenciais ordinárias de índice superior. *Revista Liberato*, v. 9, n. 11, 2008. Cited on page 118.
- MAGALHÃES, O. I. *Desenvolvimento de um sistema de otimização dinâmica em tempo real*. Dissertation (Master) — Universidade Federal do Rio de Janeiro, 2010. Cited on page 22.
- MARCHETTI, A.; CHACHUAT, B.; BONVIN, D. Modifier-adaptation methodology for real-time optimization. *Industrial and Engineering Chemistry Research*, v. 48, n. 13, p. 6022–6033, 2009. Cited 2 times on pages 125 and 126.
- MARCHETTI, A. G.; FERRAMOSCA, A.; GONZÁLEZ, A. H. Steady-state target optimization designs for integrating real-time optimization and model predictive control. *Journal of Process Control*, v. 24, n. 1, p. 129–145, 2014. Cited 5 times on pages 24, 32, 43, 50, and 56.

MARCHETTI, A. G.; FRANÇOIS, G.; FAULWASSER, T.; BONVIN, D. Modifier adaptation for real-time optimization—methods and applications. *Processes*, v. 4, n. 4, p. 55, 2016. Cited on page 125.

MATIAS, J. O. A.; EBRAHIMPOUR, M.; LE ROUX, G. A. C. Development and validation of a reduced-index dynamic model of an industrial high-purity column. *Industrial and Engineering Chemistry Research*, v. 57, n. 5, p. 1531–1546, 2018. Cited on page 121.

MATIAS, J. O. A.; LE ROUX, G. A. C.; JÄSCHKE, J. Modifier adaptation for real-time optimization of a gas lifted well network. *IFAC Papers online*, (in press). Cited on page 131.

MCCABE, W. L.; THIELE, E. W. Graphical design of fractionating columns. *Industrial and Engineering Chemistry Research*, v. 17, n. 6, p. 605–611, 1925. Cited on page 117.

MCFARLANE, R. C.; REINEMAN, R. C.; BARTEE, J. F.; GEORGAKIS, C. Dynamic simulator for a model IV fluid catalytic cracking unit. *Computers and Chemical Engineering*, v. 17, n. 3, p. 275–300, 1993. Cited on page 22.

MCLEAN, K. A. P.; MCAULEY, K. B. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Canadian Journal of Chemical Engineering*, v. 90, n. 2, p. 351–366, 2012. Cited on page 57.

MERCANGÖZ, M.; DOYLE, F. J. Real-time optimization of the pulp mill benchmark problem. *Computers and Chemical Engineering*, v. 32, n. 4, p. 789–804, 2008. Cited on page 23.

MIAO, H.; XIA, X.; PERELSON, A. S.; WU, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, v. 53, n. 1, p. 3–39, 2011. Cited on page 43.

MIDDLEBROOKS, S. A. *Modelling and control of silicon and germanium thin film chemical vapor deposition*. Thesis (PhD) — University of Wisconsin–Madison, 2001. Cited on page 40.

MIZOGUCHI, A.; MARLIN, T. E.; HRYMAK, A. N. Operations optimization and control design for a petroleum distillation process. *Canadian Journal of Chemical Engineering*, v. 73, n. 6, p. 896–907, 1995. Cited on page 22.

MORO, L. F. L.; ODLOAK, D. Constrained multivariable control of fluid catalytic cracking converters. *Journal of Process Control*, v. 5, n. 1, p. 29–39, 1995. Cited on page 64.

MUSKE, K. R.; RAWLINGS, J. B. Nonlinear moving horizon state estimation. In: *Methods of model based process control*. [S.l.]: Springer, 1995. p. 349–365. Cited on page 40.

OCHOA, S.; REPKE, J.-U.; WOZNY, G. Integrating real-time optimization and control for optimal operation: Application to the bio-ethanol process. *Biochemical Engineering Journal*, v. 53, n. 1, p. 18–25, 2010. Cited on page 22.

PANTELIDES, C. C. The consistent initialization of differential-algebraic systems. *SIAM Journal on Scientific and Statistical Computing*, v. 9, n. 2, p. 213–231, 1988. Cited on page 118.

PANTELIDES, C. C.; GRITSIS, D.; MORISON, K. R.; SARGENT, R. W. H. The mathematical modelling of transient systems using differential-algebraic equations. *Computers and Chemical Engineering*, v. 12, n. 5, p. 449–454, 1988. Cited on page 118.

PHAM, D. T.; VERRON, J.; ROUBAUD, M. C. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine systems*, v. 16, n. 3, p. 323–340, 1998. Cited 2 times on pages 40 and 69.

PONTRYAGIN, L. S.; BOLTYANSKII, V. G.; GAMKRELIDZE, R. V.; MISHCHENKO, E. *The mathematical theory of optimal processes*. [S.l.]: Interscience, 1962. Cited on page 22.

PRASAD, V.; SCHLEY, M.; RUSSO, L. P.; BEQUETTE, B. W. Product property and production rate control of styrene polymerization. *Journal of Process Control*, v. 12, n. 3, p. 353–372, 2002. Cited on page 40.

PRETT, D. M.; GARCIA, C. E. *Fundamental Process Control Butterworths Series in Chemical Engineering*. [S.l.]: Butterworth-Heinemann, 2013. Cited on page 54.

PRIOR, D.; LOPEZ, S. Grangemouth ethylene plant installs closed-loop optimization solution. *Oil and Gas Journal*, v. 97, n. 42, p. 83–86, 1999. Cited on page 33.

PRITCHARD, D.; BACON, D. Statistical assessment of chemical kinetic models. *Chemical Engineering Science*, v. 30, n. 5-6, p. 567–574, 1975. Cited on page 49.

QIN, S. J.; BADGWELL, T. A. A survey of industrial model predictive control technology. *Control Engineering Practice*, v. 11, n. 7, p. 733–764, 2003. Cited on page 22.

QUELHAS, A. D.; JESUS, N. J. C. de; PINTO, J. C. Common vulnerabilities of RTO implementations in real chemical processes. *Canadian Journal of Chemical Engineering*, v. 91, n. 4, p. 652–668, 2013. Cited 5 times on pages 23, 24, 26, 35, and 37.

RAO, C. V.; RAWLINGS, J. B.; MAYNE, D. Q. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, v. 48, n. 2, p. 246–258, 2003. Cited 2 times on pages 40 and 75.

RAWLINGS, J. B.; BAKSHI, B. R. Particle filtering and moving horizon estimation. *Computers and Chemical Engineering*, v. 30, n. 10, p. 1529–1541, 2006. Cited on page 40.

RHINEHART, R. R. Automated steady and transient state identification in noisy processes. In: IEEE. *Proceedings of the American Control Conference*. [S.l.], 2013. p. 4477–4493. Cited 3 times on pages 26, 32, and 37.

RINCÓN, F. D.; LE ROUX, G. A. C.; LIMA, F. V. A novel ARX-based approach for the steady-state identification analysis of industrial depropanizer column datasets. *Processes*, v. 3, n. 2, p. 257–285, 2015. Cited on page 25.

ROBERTS, P. D. An algorithm for steady-state system optimization and parameter estimation. *International Journal of System Science*, v. 10, n. 7, p. 719–734, 1979. Cited on page 47.

SALAU, N. P. G.; TRIERWEILER, J. O.; SECCHI, A. R. State estimation of chemical engineering systems tending to multiple solutions. *Brazilian Journal of Chemical Engineering*, v. 31, n. 3, p. 771–785, 2014. Cited 3 times on pages 40, 43, and 44.

SCHNEIDER, R.; GEORGAKIS, C. How to not make the extended kalman filter fail. *Industrial and Engineering Chemistry Research*, v. 52, n. 9, p. 3354–3362, 2013. Cited 2 times on pages 40 and 71.

SEQUEIRA, S. E.; GRAELLS, M.; PUIGJANER, L. Real-time evolution for on-line optimization of continuous processes. *Industrial and Engineering Chemistry Research*, v. 41, n. 7, p. 1815–1825, 2002. Cited 2 times on pages 33 and 34.

SHROWTI, N. A.; VILANKAR, K. P.; RHINEHART, R. R. Type-II critical values for a steady-state identifier. *Journal of Process Control*, v. 20, n. 7, p. 885–890, 2010. Cited on page 24.

SILDIR, H.; ARKUN, Y.; CAKAL, B.; GOKCE, D.; KUZU, E. Plant-wide hierarchical optimization and control of an industrial hydrocracking process. *Journal of Process Control*, v. 23, n. 9, p. 1229–1240, 2013. Cited on page 22.

SOARES, R. P.; SECCHI, A. R. EMSO: A new environment for modelling, simulation and optimisation. *Computer Aided Chemical Engineering*, v. 14, p. 947–952, 2003. Cited on page 122.

SONG, Y.; GRIZZLE, J. W. The extended kalman filter as a local asymptotic observer for nonlinear discrete-time systems. In: IEEE. *Proceedings of the American Control Conference*. [S.l.], 1992. p. 3365–3369. Cited on page 45.

Van WIJK, R. A.; POPE, M. R. Advanced process control and on-line optimisation in shell refineries. *Computers and Chemical Engineering*, v. 16, p. S69–S80, 1992. Cited on page 22.

VETTENRANTA, J.; SMEDS, S.; YLI-OPAS, K.; SOURANDER, M.; VANHAMAKI, V.; AALJOKI, K.; BERGMAN, S.; OJALA, M. et al. Dynamic real-time optimization increases ethylene plant profits. *Hydrocarbon Processing*, v. 85, n. 10, p. 59–59, 2006. Cited on page 22.

WANG, H.; WANG, Z.; DU, W.; WANG, D.; QIAN, F.; TANG, Z. Intelligent severity optimization project pays off in two months. *Hydrocarbon Processing*, v. 90, n. 10, p. 39–43, 2011. Cited on page 22.

WESTLAKE, G. Real-time process optimization aids oil production operation. *Oil and Gas Journal*, v. 96, n. 46, 1998. Cited on page 23.

WOODWARD, L.; SRINIVASAN, B.; ROBITAILLE, B.; PERRIER, M. Real-time optimization of an off-gas distribution system of an iron and titanium plant. *Computers and Chemical Engineering*, v. 31, n. 4, p. 384–389, 2007. Cited on page 23.

YEOMANS, H.; GROSSMANN, I. E. A systematic modeling framework of superstructure optimization in process synthesis. *Computers and Chemical Engineering*, v. 23, n. 6, p. 709–731, 1999. Cited on page 127.

YING, C. M.; JOSEPH, B. Performance and stability analysis of LP-MPC and QP-MPC cascade control systems. *AIChE Journal*, v. 45, n. 7, p. 1521–1534, 1999. Cited on page 33.

YOON, S.; DASGUPTA, S.; MIJARES, G. Real-time optimization boosts capacity of korean olefins plant. *Oil and Gas Journal*, v. 94, n. 25, 1996. Cited on page 22.

ZANIN, A. C. *Implementação industrial de um otimizador em tempo real*. Thesis (PhD) — Universidade de São Paulo, 2001. Cited 2 times on pages 67 and 110.

ZANIN, A. C.; TVRZSKA, M. G.; ODLOAK, D. Integrating real-time optimization into the model predictive controller of the FCC system. *Control Engineering Practice*, v. 10, n. 8, p. 819–831, 2002. Cited 2 times on pages 64 and 74.

ZHANG, Y.; FORBES, J. F. Extended design cost: a performance criterion for real-time optimization systems. *Computers and Chemical Engineering*, v. 24, n. 8, p. 1829–1841, 2000. Cited on page 47.

Appendices

APPENDIX A – Fluid catalyst cracking model and validation

This appendix concerns the development of the fluid catalyst cracking (FCC) conversor model used in case study 2 in Chapter 4.

Modeling

The dynamic model encompasses only the conversor (riser/reactor and regenerator systems) of a FCC unit. The other systems, such as feed preheater and fractionator columns, are not included. The process is modeled as an association of two continuous stirred-tank reactors (CSTR), which results in a system of ordinary differential equations (ODE). The model contains energy balances, mass balances for all components and relations for predicting the reaction rates. The considered reactions are products conversion, coke formation at the riser, and coke combustion rate. The temperature and oxygen concentration at the flue gas are calculated using semi-empirical relationships, all of which are derived utilizing experimental studies reported in the literature and are thoroughly explained in Kurihara (1967). A number of simplifications have been put forward in order to obtain the model. All the hypotheses are listed below:

1. The catalyst is maintained in a fluidized state. The properties of the fluidized bed regarding both spent and regenerated catalyst are considered homogeneous in the control volumes;
2. Both the gas formed by the cracking reactions and the air fed to the regenerator pass through the catalyst fluidized bed in a plug-flow manner with negligible time lag. Therefore, the heat transfer between solid and gas can be considered fast and the gas temperature and properties (like density, specific heat) are assumed to be constant throughout the bed;
3. Heat loss from riser/reactor and regenerator walls is negligible;
4. Pressure and catalyst accumulation are constant in both vessels (riser/reactor and regenerator);
5. In the riser, the heat capacity of reactants and products are equal and constant;
6. The system of equations is dimensional (i.e. the variables must be used in specific units, which are specified in the List of symbols in the beginning of the thesis).

Despite its simplicity (i.e only three state variables are used to represent all components of the system: catalytic carbon formed due to cracking C_{cat} ; total carbon on the spent catalyst C_{sc} ; and total carbon on the regenerated catalyst C_{rs}), the model is able to represent with satisfactory precision the actual temperatures of the riser/reactor and regenerator as well as the overall dynamic response of the unit (KURIHARA, 1967). The process dynamics should be accurately represented by the model in order to make it suitable for control studies. In turn, the temperature calculation is important for predicting the valuable products concentration, which dictates the economic performance of the unit.

Figure 22 – Conceptual flowsheet of the FCC model

The model is shown below, each control volume is presented in a different section. The flowsheet, which was introduced in Chapter 4, is shown again to facilitate the reading.

Riser/reactor

Two different species of carbon are taken into account. The catalytic carbon (C_{cat}), which is deposited on the catalyst by the cracking reactions, and the total carbon on the spent catalyst (C_{sc}) (catalytic carbon plus the coke that comes on the regenerated catalyst).

$$\frac{dC_{cat}}{dt} = \frac{1}{H_{ra}}(-60R_{rc}C_{cat} + 50R_{cf}) \quad (46)$$

$$\frac{dC_{sc}}{dt} = \frac{1}{H_{ra}}(60R_{rc}(C_{rc} - C_{sc}) + 50R_{cf}) \quad (47)$$

where, $60 = (60 \text{ min/h})$ and $50 = (1000/M)(100\%/1)(2000 \text{ lb/ton})$. H_{ra} is the riser/reactor catalyst holdup. R_{rc} is the catalyst recirculation rate. C_{rc} is the total carbon concentration on the regenerated catalyst. The rate of formation of catalytic carbon, R_{cf} , represents the carbon deposition on the catalyst and is a measurement of catalyst activity lost.

$$R_{cf} = H_{ra}P_{ra} \frac{k_{cc}}{C_{cat}C_{rc}^{0.06}} \exp\left(\frac{-E_{cc}}{R_g(T_{ra} + 460)}\right) \quad (48)$$

where, $T[^{\circ}R] = T[^{\circ}F] + 460$. The kinetic parameters for the coke formation at the riser are k_{cc} and E_{cc} . R_g is the gas constant. The other variables are the riser/reactor pressure P_{ra} , which is considered constant, and riser/reactor temperature T_{ra} , which is calculated using the energy balance. Note that the endothermic reaction effects are taken into account in the energy balance, explaining the negative sign of the last term in Equation (49). The feed vaporization also requires energy from the system. The heat necessary is provided by

the catalyst coming from the regenerator, which is represented by the first term of the equation.

$$\begin{aligned} \frac{dT_{ra}}{dt} = \frac{1}{S_c H_{ra}} (60 S_c R_{rc} (T_{rg} - T_{ra}) + 7/8 S_f D_{tf} R_{tf} (T_{fp} - T_{ra}) \\ + 7/8 (-\Delta H_{fv}) D_{tf} R_{tf} + 1/2 (-\Delta H_{cr} R_{or}) \end{aligned} \quad (49)$$

where, $60 = (60 \text{ min/h})$, $1/2 = (1000/M)(\text{ton}/2000 \text{ lb})$ and $7/8 = (42 \text{ gal/bbl})(1000/M)(\text{day}/24 \text{ hr})(\text{ton}/2000 \text{ lb})$. S_f , D_{tf} , ΔH_{fv} , R_{tf} , and T_{fp} are the specific heat, density, heat of vaporization, flowrate and temperature of the feed, respectively. T_{rg} is the regenerator temperature. S_c is the specific heat of catalyst. ΔH_{cr} is the heat of cracking. R_{or} is the feed (gasoil) cracking rate:

$$R_{or} = \frac{1.75 H_{ra} P_{ra} D_{tf} K_{cr}}{1 + H_{ra} P_{ra} K_{cr} / R_{tf}} \quad (50)$$

where, $1.75 = (42 \text{ gal/bbl})(\text{day}/24 \text{ hr})$. K_{cr} is the carbon cracking reaction rate and is calculated by:

$$K_{cr} = \frac{k_{cr}}{C_{cat} C_{rc}^{0.15}} \exp \left(\frac{-E_{cr}}{R_g (T_{ra} + 460)} \right) \quad (51)$$

where, $T[^\circ R] = T[^\circ F] + 460$. k_{cr} and E_{cr} are the kinetic parameters of the carbon cracking reaction.

Regenerator

The mass balance takes into account the coke burning rate to calculate the total carbon concentration on the regenerated catalyst (C_{rc}).

$$\frac{dC_{rc}}{dt} = \frac{1}{H_{rg}} (60 R_{rc} (C_{sc} - C_{rc}) - 50 R_{cb}) \quad (52)$$

where, $60 = (60 \text{ min/h})$ and $50 = (1000/M)(100\%/1)(2000 \text{ lb/ton})$. H_{rg} is the regenerator catalyst holdup. The coke burning rate is calculated by a stoichiometric relation:

$$R_{cb} = \frac{R_{ai} (21 - O_{fg})}{100 C_1} \quad (53)$$

in which, $21 = (O_2\% \text{ mol})$ in air and $100 = (100\%/1)$. C_1 is a constant that represents the partial combustion of carbon, i.e. the amount of carbon that is oxidized to carbon monoxide and dioxide. For example: if 50% of the carbon is partially combusted, then $2C + 3/2O_2 \rightarrow CO + CO_2$. In this case, C_1 is estimated by $C_1 = 2.0 \text{ lb oxygen/lb carbon}$. O_{fg} is the oxygen concentration, which is calculated by:

$$O_{fg} = 21 \exp \left(\frac{\frac{-H_{rg} P_{rg} / R_{ai}}{1} + \frac{100}{C_2 R_{ai}^2 + \frac{C_{rc} k_{or} \exp(-E_{or} / R_g (T_{rg} + 460))}{100}}}{\frac{100}{C_2 R_{ai}^2 + \frac{C_{rc} k_{or} \exp(-E_{or} / R_g (T_{rg} + 460))}{100}}} \right) \quad (54)$$

where, $21 = (O_2\% \text{ mol})$ in air, $T[^\circ R] = T[^\circ F] + 460$ and $100 = (100\%/1)$. $C_2 = 4.76e^{-6}$ is a semi-empirical constant. P_{rg} is the regenerator pressure. R_{ai} is the air flowrate. k_{or} and E_{or} are the oxygen formation kinetic parameters. T_{rg} is the regenerator temperature. The energy balance calculates the changes in T_{rg} .

$$\frac{\partial T_{rg}}{\partial t} = \frac{1}{S_c H_{rg}} (60 S_c R_{rc} (T_{rg} - T_{ra}) + 1/2 S_a R_{ai} (T_{ai} - T_{rg}) + 1/2 (-\Delta H_{rg}) R_{cb}) \quad (55)$$

where, $60 = (60 \text{ min/h})$ and $1/2 = (1000/M)(\text{ton}/2000 \text{ lb})$. S_a is the specific heat of feed. T_{ai} is the air temperature. ΔH_{rg} is the heat of carbon burning reaction. Finally, to predict the flue gas temperature, a simple linear empirical relationship is used:

$$T_{fg} = T_{rg} + C_4 O_{fg} \quad (56)$$

where, C_4 is an empirical constant.

Gasoline volumetric yield

The gasoline production is calculated using empirical correlations, which relate the riser/reactor, regenerator and feed temperatures with the gasoline volumetric yield, $GLNV$. The empirical model presented here is introduced by Zanin (2001) and is based on the correlations used by Petrobras S.A..

$$GLNV = F_1 F_2 \quad (57)$$

where, the auxiliary variables, F_1 and F_2 , are calculated as:

$$F_1 = -27.198427 + 7.128543 \ln FSF + 0.555905 \text{ CONVV} - 2.56093e^{-7} \text{ CONVV}^4 - 9.963736e^{-8} \text{ FSF}^4 + 0.008509 \text{ CONVV} \times \text{FSF} \quad (58)$$

$$F_2 = -0.5972 + 0.015746 \text{ FSF} + 14.107127/\text{FSF} - 1.16e^{-4} \text{ FSF}^2 - 3.3024438e^{-6} \left(\frac{\text{CONVV} \times T_{ra}[^\circ C]}{\text{FSF}} \right)^2 - 0.00279 \frac{\text{CONVV} \times T_{ra}[^\circ C]}{\text{FSF}} \quad (59)$$

where, T_{ra} is the riser/reactor temperature. The volumetric conversion of coke, CONVV is calculated as a function of the parameter FSF , which, in turn, is calculated based on feed properties.

$$\begin{aligned} \text{CONVV} = & -0.019164 \text{ FSF}(1 + 0.021289919 \text{ SEV}) - \frac{64.866937}{\text{FSF}}(1 + 0.01248132 \text{ SEV}) \\ & + 1.14585 \text{ SEV} + 0.000997 \text{ SEV}^2 + 3.32586e^{-6} \text{ SEV}^3 \end{aligned} \quad (60)$$

Severity SEV is calculated using the following correlations:

$$SEV = \frac{100A_{est}}{1 + A_{est}} \quad (61)$$

$$A_{est} = \frac{2.5e^5 \text{ catoil}^{0.65}}{R_{tf}^{0.35}} \exp\left(\frac{-E_{cr}}{R_g T_{ra}[K]}\right) \quad (62)$$

where, $catoil$ is the relation between the mass flows of the regenerated catalyst and feed. It is estimated by:

$$\text{catoil} = 2.761 \frac{T_{ra} - T_{fp}}{T_{rg} - T_{ra}} + 1.805 \quad (63)$$

where, T_{fp} is the feed temperature and T_{rg} the regenerator temperature.

Validation

The dynamic model results are compared to plant data in order to verify the quality of the model response. The data are obtained from the conversor of the FCC Unit of Refinery Henrique Lage owned by Petrobras S.A.. The data set contains measurements of variables listed in Table 9 with frequency of 1 minute. Note that the units, which are going to be used on this point forward, are different from the ones used in the model description. The italic names are the measurements (dependent values), while the others are the independent values.

Table 9 – Measurements used for validation and the correspondent model variables

Measurement	Model variable	Unit
<i>Riser/reactor temperature</i>	T_{ra}	$^{\circ}C$
<i>Regenerator temperature</i>	T_{rg}	$^{\circ}C$
<i>Flue gas temperature</i>	T_{fg}	$^{\circ}C$
Feed flowrate	R_{tf}	m^3/day
Feed temperature	T_{fp}	$^{\circ}C$
Air flowrate	R_{ai}	Nm^3/h
Air temperature	T_{ai}	$^{\circ}C$
Catalyst recirculation rate	R_{rc}	ton/h
Riser/reactor pressure	P_{ra}	kgf/cm^2a
Regenerator pressure	P_{rg}	kgf/cm^2a

The system is simulated as follows: First, the stationary periods are identified in the data set from the real plant using the methodology proposed by Cao and Rhinehart (1997), which was described in Chapter 2. Then, one of the steady state (SS) periods is chosen and the values of the variables presented in Table 9 are averaged over this period. Finally, the averaged data is used for estimating the model parameters and states. The model parameters are divided in two sets. The choice is based on process knowledge. The adjustable set is presented in Table 10. The fixed parameters remain in their nominal values, which are presented in Kurihara (1967).

The simulation and parameter/state estimation are performed in Matlab using the IPOPT solver available at the CASADI extension (ANDERSSON; ÅKESSON; DIEHL, 2012). The parameters values are computed as described in Chapter 2, Problem (5). They are estimated by minimizing the squared deviation between the plant measurements and model prediction of the temperatures T_{ra} , T_{rg} and T_{fg} .

Table 10 – FCC adjustable model parameters for validation

Paramter	Model variable
Coke formation reaction parameters	k_{cc} and E_{cc}
Cracking reaction parameters	k_{cr} and E_{cr}
Oxygen prediction parameters	k_{or} and E_{or}
Heat of cracking	ΔH_{cr}
Feed vaporization heat	ΔH_{fv}
Heat of carbon combustion	ΔH_{rg}
Riser/reactor holdup	H_{ra}
Semi-empirical parameters	C_1 and C_4

Note that there are 12 parameters to adjust only 3 measurements. The large number of parameters can lead to overfitting of the plant data. However, after calculating the initial parameters estimates, a reduced extended Kalman filter (rEKF) is applied to the data in order to estimate the parameters and system states dynamically. As explained in Chapter 2, rEKF uses plant measurements to reduce the adjustable parameter set (using the lower-rank approximation of $\mathbf{P}_k$). Hence, rEKF can mitigate a possible model overfitting to the initial data.

The idea here is to simulate the model using the independent variables (inputs plus manipulated variables) values of the plant data and, then, correct the model predictions with the rEKF using the plant measurements. As the inputs are directly obtained from the plant, the control structure discussed in Chapter 4 is not implemented. In order to apply the filter, the covariance matrices are determined as following (the matrices are built as described previously except for $\mathbf{R}$, where the entries correspond to T_{ra} , T_{rg} and T_{fg}):

$$\begin{aligned}
\mathbf{Q} &= \text{diag}([1e^{-3}; 1e^{-3}; 1e^{-1}; 1e^{-2}; 1e^{-1}]) \\
\mathbf{Q}^p &= \text{diag}([1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-3}; 1e^{-4}; 1e^{-3}; 1e^{-3}]) \\
\mathbf{R} &= \text{diag}([1e^2; 1e^3; 1e^2])
\end{aligned} \tag{64}$$

The results of the model and real data comparison are shown in Figure 23. Here, the temperatures profiles calculated by the model (continuous blue line) are compared to the plant values (dashed red line). During the filter tuning, the main goal is to predict the riser/regenerator temperature accurately because the unit optimal operation relies on the maximization of the gasoline production, which is a function of this temperature.

Despite the deviations, the results show that the model can represent the main process dynamics and, most important, calculate the temperature values with acceptable accuracy. Apparently, the filter “learns” slowly the true value of T_{rg} and the predicted

Figure 23 – Comparison of the plant measurements with the model predictions (Source: own elaboration).

values reaches the measured value only around 1600 minutes. In relation to T_{fg} , there is a small difference, around $10^{\circ}C$, between the predicted and measured values. The values of the independent variables (manipulated variables plus process inputs) that generated the temperature profiles are shown in Figure 24.

Figure 24 – Independent variables that generated the temperature profiles of Figure 23 (Source: own elaboration).

Analyzing the independent variables, the cause of the variation on the temperature profiles appears. Around 1300 minutes, the independent variables R_{tf} , R_{rc} , R_{ai} , P_{ra} and P_{rg} change in a step-like form influencing the model response. Additionally, there is an increase and posterior decrease in the air temperature around 3000 minutes, which explains the small variation observed in the regenerator and flue gas temperatures. As an extra feature of this validation, it is possible to analyze the estimated parameter profiles, Figures 25 and 26. Note that the parameter values in these figures are normalized between 0 and 10.

Figure 25 – Analysis of the estimated parameter values of the model using plant data and rEKF (Source: own elaboration).

In the parameter profiles, there is no clear characterization of steady-state periods. Some of the parameters estimates maintain the same values for some periods, but regarding the whole system, there is not a period when all the parameters reach a stationary value. Hence, this indicates a potential opportunity for applying the ROPA methodology to optimize the system instead of the standard RTO approach.

Figure 26 – Analysis of the estimated parameter values of the model using plant data and rEKF (cont.) (Source: own elaboration).

APPENDIX B – Dynamic design using proportional action (DDPA)

1

¹ "Reprinted (adapted) with permission from *Development and Validation of a Reduced-Index Dynamic Model of an Industrial High-Purity Column*. Copyright 2018 American Chemical Society."

This appendix, which is part of the complementary work of the thesis, presents an alternative modeling approach called Dynamic Design Using Proportional Action (DDPA). DDPA is used to simplify the hydraulic and pressure drop calculations in a distillation column model. A depropanizer column is modeled using DDPA and the model is validated with real process data obtained in an industrial-scale column located at Refinaria de Paulínia, Paulínia, São Paulo, Brazil and owned by Petrobras S.A.. The results show that the model with DDPA is able to predict the actual column dynamic behavior properly.

Introduction

Dynamic modeling of distillation columns can be carried out using different levels of complexity for the tray hydrodynamics and vapor-liquid equilibrium (VLE). In cases where a detailed description is necessary, rigorous models should be used. They are characterized by tray-by-tray models using mass and energy balances, and equilibrium relations on each tray. Additionally, equations describing pressure and flow dynamics are necessary. Depending on the assumptions on the pressure and molar holdup, the model equation system is affected differently. Table 11 summarizes the most common modeling approaches regarding pressure and flow dynamics and how they influence the model equations.

Table 11 – VLE modeling strategies - Distillation Columns

Molar Holdup	Pressure	Equation or Assumption	Consequence	Examples
Variable	Variable	- Francis weir formula for flow through weirs - gas pressure drop through plate	DAE model is Index 1 Validity of the correlation and parameter accuracy are not guaranteed.	(GANI; RUIZ; CAMERON, 1986), (LOCKETT, 1986), and most commercial simulators
Fixed	Fixed	- liquid and vapor flow constant in sections - no heat of mixing	DAE model is Index 1 McCabe-Thiele hypothesis. No energy balances in trays. Oversimplification of heat relations.	(MCCABE; THIELE, 1925) (JOGWAR; DAOUTIDIS, 2009)
Fixed	Fixed	- differential equations added to the model: $dM/dt = 0$ and $dP/dt = 0$	DAE model is Index 2 McCabe-Thiele hypothesis are not necessary. Liquid and vapor flow change through the column.	(ALBET et al., 1994)

A differential algebraic equation (DAE) model is basically a combination of differential and algebraic equations. Fundamentally, DAE systems are distinct from ordinary differential equation (ODE) systems (BEERS, 2006) and their index provides a conceptual measurement of this difference. The system index can be defined as the minimum number of times that the algebraic equations needs to be differentiate in order to obtain an equivalent ODE system (BRENAN; CAMPBELL; PETZOLD, 1996). As seen in Table 11, different assumptions for the description of the hydraulic behavior lead to DAE models of different indexes.

If the DAE system is index one, it can be easily rearranged and treated as an ODE system. Then, it can be solved with ODE solvers or conventional DAE solvers (BRENAN; CAMPBELL; PETZOLD, 1996). If the system is index two or more, the model index can be either reduced to one by algorithms that reduce the index automatically (PANTELIDES

et al., 1988; LOURENÇO; SECCHI, 2008), or efficiently solved by stiffly stable solvers, like IDAS included in SUNDIALS (HINDMARSH et al., 2005).

Nevertheless, for modeling distillation columns, both options require higher order derivatives of the thermodynamic properties, which are not normally provided by conventional thermodynamic packages and are quite difficult to obtain manually. Additionally, high index DAE systems bring numerical challenges for the simulation, e.g. finding a set of consistent initial conditions (PANTELIDES, 1988); and numerical inaccuracies in enforcing the algebraic constraints (DAOUIDIS, 2015).

Hence, in practice, it is interesting to have an index one model. According to Table 11, index one models can be obtained either by using inaccurate algebraic relations for pressure drop and tray hydraulics, like Francis weir formula, or by assuming ideal solution behavior and constant holdup. In the first option, the use of phenomenological relations for tray hydraulic and pressure drop equations give rise to errors as a result of unknown project details and parameters. For the second option, the model assumptions force liquid and vapor flows to be constant in each column section such that the energy balance in the stages is not necessary. This assumption is reasonable only if there is no heat of solution and it is unrealistic for many challenging problems in literature.

A different approach for obtaining index one models, which is proposed in this appendix, is to use linear relations between pressure/vapor flowrate and molar holdup/liquid flowrate. This linear relation is similar to a proportional controller with arbitrarily large gain. The proposed approach is called Dynamic Design Using Proportional Action (DDPA). The main advantage is that the tray model developed with the DDPA equations is still index 1, which allows the use of thermodynamic package in order to calculate complex equilibrium relations accurately. Figure 27 summarizes the commitments involved on choosing the tray modeling approach based on available solvers and information.

Mathematical Considerations

To exemplify the DDPA strategy and the sources of the index problem, the following system can be considered (BEERS, 2006):

$$\mathbf{M}(\mathbf{x})\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \quad (65)$$

in which, $\mathbf{M}(\mathbf{x})$ is a state-dependent matrix called *mass matrix*. If this matrix is nonsingular, which indicates an ODE system, the model can be easily rearranged as:

$$\dot{\mathbf{x}} = \mathbf{M}^{-1}\mathbf{f}(\mathbf{x}) \quad (66)$$

Figure 27 – Modeling options based on available solvers and information (Source: own elaboration).

However, if the mass matrix is singular (i.e. the system is composed by a combination of differential and algebraic equations), the rearrangement strategy is different. Considering the following system:

$$\begin{aligned}\dot{\mathbf{y}} &= \mathbf{F}(\mathbf{y}, \mathbf{z}) \\ \mathbf{0} &= \mathbf{G}(\mathbf{y}, \mathbf{z})\end{aligned}\tag{67}$$

Then, the equations can be expressed in the form (65).

$$\mathbf{M}\dot{\mathbf{x}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{y}} \\ \dot{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \mathbf{F}(\mathbf{y}, \mathbf{z}) \\ \mathbf{G}(\mathbf{y}, \mathbf{z}) \end{bmatrix} = \mathbf{f}(\mathbf{x})\tag{68}$$

If $(\partial \mathbf{G} / \partial \mathbf{z}^T)$ is nonsingular, it is easy to determine a set of consistent initial conditions (i.e. one that satisfies all algebraic equations) simply by setting $\mathbf{y}$, and computing $\mathbf{z}$ using Newton's method (BEERS, 2006):

$$\mathbf{0} = \mathbf{G}(\mathbf{y}, \mathbf{z})\tag{69}$$

Also, this nonsingularity allows the system to be rearranged in order to obtain an ODE system.

$$\frac{d}{dt} \begin{bmatrix} \dot{\mathbf{y}} = \mathbf{F}(\mathbf{y}, \mathbf{z}) \\ \mathbf{0} = \mathbf{G}(\mathbf{y}, \mathbf{z}) \end{bmatrix} \Rightarrow \ddot{\mathbf{y}} = \left(\frac{\partial \mathbf{F}}{\partial \mathbf{y}^T} \right) \dot{\mathbf{y}} + \left(\frac{\partial \mathbf{F}}{\partial \mathbf{z}^T} \right) \dot{\mathbf{z}}, \mathbf{0} = \left(\frac{\partial \mathbf{G}}{\partial \mathbf{y}^T} \right) \dot{\mathbf{y}} + \left(\frac{\partial \mathbf{G}}{\partial \mathbf{z}^T} \right) \dot{\mathbf{z}}\tag{70}$$

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \left(\frac{\partial \mathbf{G}}{\partial \mathbf{z}^T} \right) \end{bmatrix} \begin{bmatrix} \dot{\mathbf{y}} \\ \dot{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \mathbf{F}(\mathbf{y}, \mathbf{z}) \\ - \left(\frac{\partial \mathbf{G}}{\partial \mathbf{y}^T} \right) \mathbf{F}(\mathbf{y}, \mathbf{z}) \end{bmatrix}\tag{71}$$

The difficulty in solving the DAE system depends on the singularity of $\partial \mathbf{G} / \partial \mathbf{z}^T$ (ASCHER; PETZOLD, 1998), which relates to the index problem defined previously. If $\partial \mathbf{G} / \partial \mathbf{z}^T$ is singular, the system has an index 2 or higher.

For a more concrete example, consider the modeling of distillation column trays given in Table 11. Assuming constant pressure and liquid holdup is the same as adding $M_L = M_L^{ref}$ and $P = P^{ref}$ to the algebraic equation set. The other algebraic equations are equilibrium equation and thermodynamic relations.

However, with this assumption, vapor (V) and liquid (L) outlet flowrates do not appear in the algebraic set. Because both are algebraic variables (i.e. they do not have explicit differential equations), $\partial \mathbf{G} / \partial \mathbf{z}^T$ becomes singular, generating the index problem. It can be removed by not assuming M_L and P constant, and introducing $L = f(M_L)$ and $V = f(P)$ to the algebraic set. For further discussion, please refer to (EBRAHIMPOUR, 2015).

As discussed, the correlations for calculating the pressure drop (vapor flow rate) and liquid flow rate in the trays outlet have this form and also decrease the system index to 1. But, as explained previously, they usually rely on unknown and arbitrarily assumed parameters and project details. Therefore, the motivation to develop DDPA appears. DDPA intends to approximate the tray hydraulics by a “perfect control” strategy in the form of:

$$L - L^{ref} = K_M(M_L - M_L^{ref}) \quad V - V^{ref} = K_P(P - P^{ref}) \quad (72)$$

Note that in the cases that the gain, K_P or K_M , is large, the changes in L in relation to L^{ref} do not affect the difference $(M_L - M_L^{ref})$ to a great extent. However, the inverse statement is not true. If M_L differs from M_L^{ref} , the value of $(L - L^{ref})$ is drastically affected.

Hence, P and M_L tend to their reference value, satisfying the DDPA equations, while the liquid and vapor flows can change according to the thermodynamic calculations. Another advantage of the DDPA approach is that the relationship between the properties appears explicitly in the model, highlighting the phenomenological relationship between them.

Case study: depropanizer column modeling

To illustrate the effectiveness of DDPA, the methodology is applied to a depropanizer. The separation of close-boiling point components, propane and propene (propylene) in this case, demands a considerable amount of energy. Vapor recompression distillation (VRD) configuration is chosen due to the significant savings from the reboiler-condenser combination (JOGWAR; DAOUTIDIS, 2009), i.e. the vapor coming from the top of the column is used to vaporize the bottom stream. A simplified diagram of the process is shown in Figure 28

Figure 28 – Vapor recompression distillation column with controllers. The stream tags shown here are used in the discussion section to identify the streams (Source: own elaboration).

VRD has some characteristics that justify the specification of pressure and liquid holdup required for applying the DDPA equations, i.e. tightly controlled pressure and composition dynamics that are much slower than the flow dynamics (LEVY; FOSS; GRENS, 1969). However, as discussed in the previous sections, if pressure and holdups are directly specified to a given value, the system becomes index-2 hindering the use of thermodynamic packages. As the propane-propene system is highly non-ideal (EBRAHIMPOUR, 2015), it is necessary to apply rigorous thermodynamic calculations. Therefore, the proposed approach appears as an interesting alternative to model this process. The DDPA equations, which are added to the model, are:

$$L_{norm} - L_{norm}^{ref} = K_M (Level_{L,norm} - Level_{L,norm}^{ref}) \quad (73a)$$

$$V_{norm} - V_{norm}^{ref} = K_P (\Delta P_{norm} - \Delta P_{norm}^{ref}) \quad (73b)$$

in which, V and L are the vapor and liquid molar flows leaving the tray; $Level_L$ is the liquid level of the tray; ΔP is the difference of pressure between the tray and the one above; The subscript *norm* indicates that the variable is normalized and the superscript *ref* is related to the reference values (i.e. the specification of pressure difference and level); K_P and K_M are the gains of DDPA.

Choosing large values of K is equivalent to keep the value of pressure and liquid holdup constant. These large values generate a stiff system that tends to an index-2 system when K tends to infinity. For further information regarding DDPA numerical solution difficulties, the impact of different K values on the solution and details about the column modeling, please refer to Matias, Ebrahimpour and Le Roux (2018).

Validation of the model using the DDPA approach

The dynamic simulation is performed for more than two days of operation in order to verify the quality of model response compared to real plant data, which are obtained from an industrial-scale depropenizer column owned by Petrobras S.A.. This data set contains all the available measurements of the vapor recompression unit with frequency of 1 minute.

In order to compare the model response to plant data, data reconciliation and parameter estimation are carried out previously. The former is performed in order to adjust the data to the mass and energy balances, resulting in a consistent data set for the model (i.e. the data points can be represented by the model mass and energy balance, avoiding that random and gross errors corrupt the parameter estimation step). The latter is performed in order to reproduce bottom and top product specifications and compressor discharge temperature. Both steps and the simulation are carried out in EMSO (SOARES; SECCHI, 2003). After adjusting the model to the initial point, the compositions of propylene at distillate (*stream 4*) and propane at the bottom product (*stream 3*) predicted by the model are compared with real plant data in Figure 29. The stream tag is related to Figure 28.

Figure 29 – Dynamic response of the model bottom and top compositions in comparison to actual plant data (Source: own elaboration).

The model can predict the dynamic behavior of the plant properly. The main differences between model predictions and plant data are in propane composition at the top product. However, before 2500 minutes, plant data of the top composition are clearly affected by instrument failure. Another small deviations can be seen in Figures 29, there are several sources that could cause this loss in the model accuracy, mainly regarding unmeasured disturbances and thermodynamics, for example: uncertainties in composition

measurements and in flow rate measurements; neglecting some minor components in the feed; and inaccuracies in thermodynamic parameters, like binary coefficient for the pair propylene/propane.

However, apart from the minor lack of accuracy, the process dominant time constants of both plant and model match within an adequate precision, which is the main goal of the DDPA validation. Note that the aim is not to validate the thermodynamic package. Especially for control purposes, the response curves and the steady-state changes should be represented rather than the absolute value actual process variable. Hence, we believe that this lack of accuracy does not affect the main conclusions presented in this appendix.

Conclusion

A VRD column is modeled based on the DDPA strategy to produce a reduced-index model. Its capability for representing the column dynamics is demonstrated by comparing the simulated and real plant data during a long operation period. The DDPA approach shows a significant degree of accuracy if compared to the actual plant data.

Hence, DDPA is an interesting modeling approach in order to obtain an index one model, which allows the use of most thermodynamic packages available in equation-oriented simulation environments. In comparison to the use of correlations and phenomenological relations to describe the dynamics of the tray, DDPA does not give rise to errors as a result of unknown project details and parameters. Also, DDPA allows that the liquid and vapor flows inside the column are computed based only on the thermodynamic calculations.

APPENDIX C – Online model structure adaptation via modifier adaptation

Significant research efforts have been dedicated to the development of mathematical models over the last decades (BONVIN et al., 2016). In knowledge-driven models, physical-chemical relationships (like conservation of mass and energy and reaction kinetics) are used to describe the inner workings of the process (BEQUETTE, 2003). However, a variety of models can be used for describing the same phenomenon (BONVIN et al., 2016). In this case, the question arises: how to discriminate between different model types and structures?

There is often not enough information to make proper choices during the model design phase, for example: discriminating between competing metabolic pathways when modeling biological systems (JIA; STEPHANOPOULOS; GUNAWAN, 2012). However, as measured data are available, more information about the system behavior can (and should) be incorporated in order to properly distinguish between rival models structures. The paradigm that the model is a fixed entity and experimental data are used only to update the model parameters fails to address the rapid advances of computational methods.

Among the available methods for Real-time Optimization, Modifier Adaptation (MA) applies correction (modifier) terms to the cost and constraint functions (or directly to the model) in order to avoid plant-model mismatch (MARCHETTI; CHACHUAT; BONVIN, 2009). At a given operating point u , the modifiers indicate the differences between model predictions and plant measurements as well as the differences between model and plant gradients. Thus, the modifiers can be treated as residuals for the model predictions and shape and they can be used as indicators for discriminating between rival model structures.

Motivated by the possibility of unifying MA with model evolution, this appendix proposes a systematic approach to adapt the model structure online based on available data, while optimizing the model in an RTO-fashion using MA. The methodology is tested in a gas lift oil well network. The results show that, despite of a significant amount of noise, the correct model structure is chosen among several candidates while the plant optimum is reached without constraint violations.

Modifier Adaptation (MA)

Before presenting the proposed methodology, the Modifier Adaptation (MA) method is briefly introduced. For more details regarding MA, we refer to Marchetti et al. (2016). In MA methods, instead of updating the model parameters, plant information is incorporated to the model through modifiers (input-affine correction terms), which are applied to the cost and constraint functions of the optimization problem. The idea is to enforce that, upon convergence, the optimum calculated by the model-based optimization matches the plant optimum (i.e. avoiding plant-model mismatch). Marchetti, Chachuat and Bonvin

(2009) present a second variant of MA that consists of making the linear modifications into the process model instead of the cost and constraints:

$$\mathbf{y}_{ad,k}(\mathbf{u}) := \mathbf{y}(\mathbf{u}) + \boldsymbol{\epsilon}_k^y + (\boldsymbol{\lambda}_k^y)^T(\mathbf{u} - \mathbf{u}_k) \quad (74)$$

in which, $\mathbf{u} \in \mathbb{R}^{n_u}$ represents the process inputs; $\mathbf{y}(\mathbf{u}) \in \mathbb{R}^{n_y}$, the steady-state solution of the system composed by $\mathbf{f}_{ss}$ and $\mathbf{h}$ (discussed in Chapter 2); and k is the subscript representing the k th RTO iteration. As discussed previously, the zeroth-order modifiers $\boldsymbol{\epsilon}_k^y$ correspond to bias terms between the predicted values of the model and the plant measurements. The first order modifiers $\boldsymbol{\lambda}_k^y$ represent the difference between the plant gradients and the gradients predicted by the model. Both modifiers are valid only locally. They are calculated by:

$$\boldsymbol{\epsilon}_k^y = \mathbf{y}_p(\mathbf{u}_k) - \mathbf{y}(\mathbf{u}_k) \quad \boldsymbol{\lambda}_k^y = \left(\widehat{\frac{\partial \mathbf{y}_p}{\partial \mathbf{u}}}(\mathbf{u}_k) - \frac{\partial \mathbf{y}}{\partial \mathbf{u}}(\mathbf{u}_k) \right)^T \quad (75)$$

where, the plant output value, y_p , is directly measured. On the other hand, the plant gradients $\partial \mathbf{y}_p / \partial \mathbf{u}$ need to be estimated. Obtaining reliable gradient estimates with noisy measurements can be a challenging task. In this paper, the methodology proposed by (GAO; WENZEL; ENGELL, 2016), which uses quadratic models to compute plant gradients, is applied. The methodology uses current and past operating points in order to compute local quadratic approximation of plant. Then, the gradients are obtained by differentiating the approximated functions. After applying the modifiers to the model optimization problem, the modified version of the economic optimization problem (Equation (6)) is solved:

$$\begin{aligned} u_{k+1}^* &= \arg \min_{\mathbf{u}} \phi_{profit}(\mathbf{u}, \mathbf{y}_{ad,k}(\mathbf{u})) \\ \text{s.t. } &\mathbf{y}_{ad,k}(\mathbf{u}) = \mathbf{y}(\mathbf{u}) + \boldsymbol{\epsilon}_k^y + (\boldsymbol{\lambda}_k^y)^T(\mathbf{u} - \mathbf{u}_k) \\ &g_i(\mathbf{u}, \mathbf{y}_{ad,k}(\mathbf{u})) \leq 0, \quad i = 1, \dots, n_g \end{aligned} \quad (76)$$

where, g_i are the components of the optimization problem constraints set G and n_g is the total number of elements of this set.

Methodology Outline

The proposed methodology is based on the postulation of several model parts or “blocks”, which represent components of the complete model. Then, by examining the modifiers between successive MA executions, the methodology builds the full model by choosing the block combination that best fits the modifiers (i.e. plant measurements and gradients). Rather than determining a fixed model structure, the methodology allows the model to evolve with the data obtained during the optimization runs. A simplified diagram of the proposed methodology is shown in Figure 30.

Figure 30 – Model structure adaptation algorithm concept (Source: own elaboration).

As the size of the model increases, the number of blocks and the their possible arrangements increase exponentially. Thus, the exhaustive exploration of the block combinations becomes intractable to be solved by a brute force approach as in Figure 30. Alternatively, the combination problem can be rearranged in an optimization structure. The formulation is based on the systematic framework of superstructure optimization proposed by Yeomans and Grossmann (1999). Let the model be represented by the set $Y = B_1 \cup B_2 \cup \dots \cup B_{n_b}$, which is composed by n_b different blocks B . For each one of these blocks, the set $A_b = \{j \in B_b \mid \text{model } j \text{ can be assigned to block } B_b\}$ is also defined. Clearly, only one of the available model options for a given block can be selected at the k th iteration. The optimization problem can be written as:

$$\begin{aligned}
 z_{k+1}^* = \arg \min_{\mathbf{z}} \quad & \psi := (\epsilon_k^y)^T \cdot \epsilon_k^y + \tilde{\lambda}_k^y \cdot (\tilde{\lambda}_k^y)^T \\
 \text{s.t.} \quad & \epsilon_k^y = \mathbf{y}_p(\mathbf{u}_k) - \mathbf{y}(\mathbf{u}_k, h_1(\mathbf{u}_k), \dots, h_{n_b}(\mathbf{u}_k)), \\
 & (\lambda_k^y)^T = \left(\widehat{\frac{\partial \mathbf{y}_p}{\partial \mathbf{u}}}(\mathbf{u}_k) - \frac{\partial \mathbf{y}}{\partial \mathbf{u}}(\mathbf{u}_k, h_1(\mathbf{u}_k), \dots, h_{n_b}(\mathbf{u}_k)) \right) \\
 & \left[\bigvee_{j \in A_1} \begin{pmatrix} z_{1j} \\ h_1 = h_{1j}(\mathbf{u}_k) \end{pmatrix} \right] \wedge \dots \wedge \left[\bigvee_{j \in A_{n_b}} \begin{pmatrix} z_{n_b j} \\ h_{n_b} = h_{n_b j}(\mathbf{u}_k) \end{pmatrix} \right] \\
 & \sum_{A_b} z_{bj} = 1, \quad b = 1, \dots, n_b, \quad z_{bj} \in \{0, 1\}
 \end{aligned} \tag{77}$$

The tilde symbol in the objective function is used to represent $\tilde{\lambda}_k^y \cdot (\tilde{\lambda}_k^y)^T := \lambda_{u_1,k}^y \cdot (\lambda_{u_1,k}^y)^T + \dots + \lambda_{u_{n_u},k}^y \cdot (\lambda_{u_{n_u},k}^y)^T$, where $\lambda_{u_m}^y$ indicates that is the modifier of the gradient in relation to the m th manipulated variable. All the variables in Equation (77) were already defined except z_{bj} and $\mathbf{z}$. z_{bj} are Boolean variables that indicate if the model j can be associated to block B and $\mathbf{z}$ is the vector of all z_{bj} . Once the optimization is formulated as above, it can be solved with any mixed-integer nonlinear programming (MINLP) or by genetic algorithm (GA) solvers. The flowsheet of the whole optimization cycle (RTO cycle + model structure adaptation) is presented in Figure 31.

Figure 31 – Flowsheet of the whole optimization cycle, where the model structure adaptation is nested in the optimization cycle (Source: own elaboration)

Case study - gas lifted oil well network

In offshore production of oil and gas, the reservoir pressure affects the product outflow. In cases that the pressure is not high enough to economically lift the fluids from the reservoir to the surface, artificial lifting methods are required. One of the most commonly applied methods is gas injection (KRISHNAMOORTHY; FOSS; SKOGESTAD, 2016), in which gas is injected at the well bottom, reducing fluid mixture density and decreasing the pressure at the well bottom. As a result, the inflow from the reservoir increases. However, increasing the gas lift flowrate also results on an increase of frictional pressure drop, which decreases the well production. There is a trade-off between gas lift flowrate and well production that is intensified in cases of well networks. Allocation of limited

Figure 32 – Gas lifted oil well network containing two wells. The system is divided in three sections: wells, annulus (a void between the pipeline and the external tubing where gas is injected) and riser (pipeline that transports the gas/oil mixture from the wells to the surface) (Source: adapted from Krishnamoorthy, Foss and Skogestad (2016)).

resources among wells, like lift gas availability, and production capacity constraint affect the optimal well network production. This scenario shows potential for RTO applications (DARBY et al., 2011).

The optimal operation is achieved by maximizing the unit profit, which is composed by the revenues of oil production minus the costs related to compression of the gas for artificial lifting. The optimization problem considers processing capacity constraints and upper bounds on the maximum gas lift flowrates:

$$\begin{aligned}
 \max_{\mathbf{u}=[w_{gl,1}, w_{gl,2}]^T} \quad & \phi_{profit} := w_{oTot}^2 - 0.5 \sum_{i=1}^2 w_{gl,i}^2 \\
 \text{s. t.} \quad & G := \begin{bmatrix} w_{gTot} \\ w_{gl,1} \\ w_{gl,2} \end{bmatrix} \leq \begin{bmatrix} w_{gM} \\ w_{glM} \\ w_{glM} \end{bmatrix}
 \end{aligned} \tag{78}$$

in which, w_{gM} is the maximum gas processing capacity of the system and w_{glM} is the maximum gas lift flowrate. The gas lift mass flowrates are the decision variables ($\mathbf{u} = [w_{gl,1}, w_{gl,2}]^T$). The total oil and gas production of the network is calculated using the input-output mapping, $[w_{oTot}, w_{gTot}]^T = \mathbf{y}(w_{gl,1}, w_{gl,2})$, which represents the steady-state solution of the well network model. The process is modeled as a nonlinear system of

equations, consisting in mass balances for oil and gas and relations for calculating density, flow and pressure. For a detailed description of the model equations, please refer to (KRISHNAMOORTHY; FOSS; SKOGESTAD, 2016). Note that the mapping $\mathbf{y}$ is adapted by the modifier adaptation method, i.e. $\mathbf{y} \mapsto \mathbf{y}_{ad}$ like in Equation (74).

In order to consider the block model approach proposed in the previous section, three different blocks are considered in the model. Note that with only 3 blocks 2^3 different combinations are possible. Thus, the number of available models increase exponentially with the number of blocks, which highlights the importance of elaborating the model structure decision as an optimization problem (Equation (77)).

- Block 1:** Pressure loss due to friction along the wells. Two different models for the head loss are proposed: First, pressure drop is negligible. Second, the Darcy-Weisbach equation is used.
- Block 2:** Pressure loss due to friction along the riser. The models related to this block are the same structure as in *Block 1*.
- Block 3:** reservoir model. Two different values are considered for the productivity index (PI), which relates the oil flowrate leaving the reservoir, w_{ro} , with the pressure difference between the reservoir and well bottom hole ($p_r - p_{bh}$). The reservoir model is represented by: $w_{ro,i} = PI_i(p_r - p_{bh,i})$, $i = 1, 2$, with $PI = \{[5; 5] \vee [7; 7]\}$.

In order to test the ability of the methodology to track changes in the plant, two different plant behaviors were modeled as shown in Figure 33. In Region *A*, pressure loss due to friction is taken into account in both wells and in the riser. In turn, Region *B* does not consider frictional pressure drop (i.e. the pressure difference between the top and bottom is related only with the static pressure exerted by the liquid column).

Case study

The case study shows the application of the proposed methodology in the gas lifted well network. The following assumptions are made:

1. Modifier Adaptation (of the measurements) is used for optimizing the gas lifted oil well network;
2. All blocks proposed in the previous section are used. Thus, the algorithm has 8 different model options to choose based on the modifiers;
3. the measurement of all the pressures in the system are available, however they are affected by noise. The measurements have 1% noise, which is neither correlated in time nor between measurements. The noisy values are generated as in Chapter 3;
4. The plant gradients are estimated by the methodology proposed by Gao, Wenzel and Engell (2016). This method has been tested in a gas lifted well network and

shown acceptable results for predicting plant gradients (MATIAS; LE ROUX; JÄSCHKE, (in press)).

The results of the optimization run are shown below. They are plotted together with the plant contour lines (Figure 33b). In addition to the optimization, the model structure updating algorithm proposed in Equation 77 is also implemented. The model structure selection also takes advantage of the plant probing points, which are used for obtaining the gradients, and calculates the modifiers in these points. Along the optimization path, markers are used for indicating which model is chosen. The correspondence of markers and models is shown in Table 12.

Table 12 – Marker and plant model correspondence. Model 2 is the correct model for plant region *A* and model 8 for region *B*

Model	Symbol	Block 1 - Well frictional pressure drop	Block 2 - Riser frictional pressure drop	Block 3 - Reservoir model parameters (<i>PI</i>)
M_1	$\triangleleft$	without	without	[7; 7]
M_2	$\bigcirc$	without	without	[5; 5]
M_3	$\star$	without	with	[7; 7]
M_4	∇	without	with	[5; 5]
M_5	$\triangle$	with	without	[7; 7]
M_6	$\square$	with	without	[5; 5]
M_7	$\triangleright$	with	with	[7; 7]
M_8	$\diamond$	with	with	[5; 5]

Figure 34a shows that the methodology is able to reach the plant optimum in a scenario with noise and structural plant-model mismatch. Note that in addition to the set

Figure 33 – Gas lifted well network plant surface (Source: own elaboration)

(a) Profit contour lines for the plant. The dashed blue line is the maximum gas production constraint and dashed black lines are the maximum gas lift flow rate for each well. The plant optimum is also indicated.

(b) Models used in each RTO execution and plant probing for gradient estimation. Full markers indicate that the correct model was chosen for the Region. Marker and model structure correspondence is shown in Table 12

Figure 34 – Models structure selection during optimization (Source: own elaboration).

points computed by MA, there are several plant probing points, which are used by the gradient estimation method in order to obtain additional information of the plant surface (20 probing setpoints vs 10 optimization setpoints).

Figure 34b shows which models are chosen along the optimization path. Even with noise, the methodology is able to track the right model for the region. When the plant operational point changes from region *A* to *B*, the algorithm takes one iteration to find the right model (between the 5th and 6th iterations). Next, there are three plant probes in Region *B* due to the gradient estimation method. These points are in the transition between two regions, as shown in Figure 34a. Hence, there is no correct model. Finally, after the 10th iteration, the operation does not leave Region *B*. However, the algorithm chooses the wrong model in 5 iterations. It picks up models M_4 and M_6 , which have the correct reservoir model but the wrong pressure drop model. Note that the methodology never chooses the wrong reservoir model. As the effects of the pressure drop are less acute, noise has a larger effect on the comparison between models. Moreover, as the pressures over the well network are coupled, the pressure drop model effects are more dispersed among the measurements and are more prone to be affected by noise. Nevertheless, the methodology is able to choose the correct models in most iterations.

Conclusion

A systematic approach that combines Modifier Adaptation (MA), an RTO variant, with an online model structure adaptation method has been proposed in this appendix.

Without deriving a fixed model for the system, the novel approach is able to determine online the best model structure (i.e. more consistent with plant data) combining different sub-models while optimizing the process to the actual plant optimum. The methodology is based on the following idea: The system model is divided in several sub-models, so called “blocks”. For each of these blocks, there are two or more model options that describe a given phenomenon happening in the process. Then, using an optimization problem, the approach combines some of the blocks in order to describe the inner workings of the process. The comparison criterion among the model candidates is based on modifiers from Modifier Adaptation techniques, which are calculated based on the deviation of model predictions to plant measurements and also of the deviation of model and plant gradients.

This approach is specially appealing in cases where some aspects of the process are not well known or where there is not enough information to make a priori assumptions for the model structure. Model uncertainties can be mitigated by the proposition of several knowledge-based models for the same phenomenon. While all available knowledge of the process is used to build the blocks, the missing information (representing interconnections of the blocks) is determined fitting the model to process data. The combination of process knowledge with online data can be attractive and rewarding.

The novel approach is implemented in a gas lifted oil well network. The case study contains a plant with two different behaviors and 8 possible models are proposed. Among this population of models, there are models valid for each of the plant behaviors, models that describe the plant behavior partially and models that completely fail to predict it. The simulation results show that the methodology is capable of choosing the right model for the plant region, while maximizing the production of the oil well network.

This novel approach is the basis of a development towards online model structure improvement, unifying process knowledge (represented by the model blocks) and data (by the modifiers). Due to the proposed framework, the model is allowed to evolve over time. New blocks can be added (or deleted) during the operation, changing the paradigm that the model structure is an immutable entity. Moreover, the new method to incorporate plant data in order to adapt the model structure is an interesting contribution and can take profit from the the rapid advances in computational methods.