



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE ENGENHARIA ELÉTRICA**

**LUIS PAULO FAGUNDES**

**SELEÇÃO ALEATÓRIA DA ESTRUTURA DE  
MODELOS COM AUXÍLIO DA TAXA DE REDUÇÃO  
DO ERRO E HERANÇA GENÉTICA**

Uberlândia  
2022

**LUIS PAULO FAGUNDES**

**SELEÇÃO ALEATÓRIA DA ESTRUTURA DE MODELOS COM  
AUXÍLIO DA TAXA DE REDUÇÃO DO ERRO E HERANÇA  
GENÉTICA**

**Tese** apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Uberlândia, como parte dos requisitos para obtenção do título de **DOUTOR EM CIÊNCIAS**.

Área de concentração: Sistemas de Energia Elétrica.

Linha de Pesquisa: Controle e Automação.

Orientador: Prof. Dr. Aniel S. de Moraes

Coorientador: Prof. Dr. Luis Cláudio de Oliveira Lopes

Uberlândia - MG

2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

F156 Fagundes, Luis Paulo, 1984-  
2022 Seleção aleatória da estrutura de modelos com auxílio  
da taxa de redução do erro e herança genética [recurso  
eletrônico] / Luis Paulo Fagundes. - 2022.

Orientador: Aniel Silva de Moraes.

Coorientador: Luis Claudio Oliveira Lopes.

Tese (Doutorado) - Universidade Federal de Uberlândia,  
Pós-graduação em Engenharia Elétrica.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.te.2022.391>

Inclui bibliografia.

1. Engenharia elétrica. I. Moraes, Aniel Silva de,  
1979-, (Orient.). II. Lopes, Luis Claudio Oliveira,  
1964-, (Coorient.). III. Universidade Federal de  
Uberlândia. Pós-graduação em Engenharia Elétrica. IV.  
Título.

CDU: 621.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:  
Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074



## UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Engenharia Elétrica  
 Av. João Naves de Ávila, 2121, Bloco 3N - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
 Telefone: (34) 3239-4707 - www.posgrad.feelt.ufu.br - copel@ufu.br



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Elétrica				
Defesa de:	Tese de Doutorado, 301, PPGEELT				
Data:	Vinte e dois de julho de dois mil e vinte e dois	Hora de início:	14:00	Hora de encerramento:	17:00
Matrícula do Discente:	11713EEL008				
Nome do Discente:	Luis Paulo Fagundes				
Título do Trabalho:	Seleção aleatória da estrutura de modelos com auxílio da taxa de redução do erro e herança genética				
Área de concentração:	Sistemas de energia elétrica				
Linha de pesquisa:	Controle e Automação				
Projeto de Pesquisa de vinculação:	Coordenador do projeto: Aniel Silva de Moraes. Título do projeto: Controle Cooperativo de Enxame de Veículos Aéreos não Tripulados Especializados no Monitoramento e Detecção de Falhas na Indústria 4.0. Agência financiadora: CNPq. Número do processo na agência financiadora: 403054/2021-4. Vigência do projeto: de 17/03/2022 a 31/03/2025.				

Reuniu-se por meio de videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Elétrica, assim composta: Professores Doutores: Daniel Costa Ramos - FEELT/UFU; Renato Santos Carrijo - FEELT/UFU; Maurício Bezerra de Souza Júnior - UFRJ; Henrique José Avelar- CEFET; Aniel Silva de Moraes - FEELT/UFU, orientador(a) do(a) candidato(a).

Iniciando os trabalhos o(a) presidente da mesa, Dr(a). Aniel Silva de Moraes, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado(a).

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **MAURÍCIO BEZERRA DE SOUZA JÚNIOR, Usuário Externo**, em 22/07/2022, às 16:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Aniel Silva de Moraes, Professor(a) do Magistério Superior**, em 22/07/2022, às 16:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Daniel Costa Ramos, Professor(a) do Magistério Superior**, em 22/07/2022, às 16:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Renato Santos Carrijo, Professor(a) do Magistério Superior**, em 22/07/2022, às 16:15, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Henrique José Avelar, Usuário Externo**, em 22/07/2022, às 16:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3768462** e o código CRC **44DB7C78**.

## AGRADECIMENTOS

Aos meus pais Joel e Divina, por me propiciarem toda a base educacional que permitiu que eu caminhasse até aqui;

A minha madrinha que desde sempre foi uma pessoa muito presente e muito especial na minha vida;

Aos meus irmãos Cláudio e Rosângela, e também aos amigos Nivaldo, Amanda, Leonardo, Pablo e Germano, por todos os bons momentos vividos ao longo dessa jornada;

Ao meu primo Frederico que foi meu companheiro de viagens e estudos na Universidade de Uberlândia ao longo de 7 anos;

À Universidade Federal de Uberlândia e à Faculdade de Engenharia Elétrica pela oportunidade de realizar este Curso;

Ao Prof. Dr. Aniel Silva de Moraes pela orientação, oportunidade de aprendizado e inestimável apoio, materializados pelo interesse sempre presentes em nosso trabalho;

Ao Prof. Dr. Luis Cláudio de Oliveira Lopes pela coorientação, oferecendo com extrema generosidade e paciência o seu tempo e conhecimento para realização deste trabalho;

Ao Prof. Dr. Josué Silva de Moraes que foi meu orientador durante os primeiros anos do curso e que sempre se demonstrou muito acessível com boas ideias e ótimas conversas;

A todos que em mim depositaram sua confiança e, mesmo não colaborando diretamente na área técnica, dirigiram suas boas energias, pensamentos, palavras e intenções em favor da concretização deste sonho.

Fagundes, L. P., **SELEÇÃO ALEATÓRIA DA ESTRUTURA DE MODELOS COM AUXÍLIO DA TAXA DE REDUÇÃO DO ERRO E HERANÇA GENÉTICA**. 2022. 132 F. Tese de Doutorado, Universidade Federal de Uberlândia, Uberlândia.

## RESUMO

Os sistemas presentes na indústria e na ciência são comumente não lineares, o que fez com que métodos para a seleção da estrutura deste tipo de sistema fossem amplamente estudadas ao longo dos últimos trinta anos. Existem muitos métodos na literatura para lidar com a etapa de seleção da estrutura de modelos no processo de identificação de sistemas, embora estes métodos tenham seus benefícios específicos, eles enfrentam algumas dificuldades em gerar modelos parcimoniosos. Neste trabalho, dois métodos baseados na Seleção Aleatória da Estrutura de Modelos (Randomized Model Structure Selection - RaMSS) são introduzidos com o objetivo de lidar com o problema da seleção da estrutura. O primeiro, chamado de Seleção Aleatória da Estrutura de Modelos com Taxa de Redução do Erro (*Randomized Model Structure Selection with Error Reduction Ratio - RaMSS-ERR*) usa a taxa de redução do erro como um pré filtro para a análise e seleção dos termos, melhorando a convergência, e o segundo, Seleção Aleatória da Estrutura de Modelos com Herança Genética (Randomized Model Structure Selection with Genetic Inheritance - RaMSS-EGI) que utiliza um processo de herança genética com o objetivo de acelerar a convergência. Os métodos foram aplicados em modelos de referência comumente usados na literatura de uma entrada e uma saída, e também em modelos de múltiplas entradas e múltiplas saídas identificando os com maior parcimônia do que alguns métodos presente na literatura. Foram ainda aplicados na identificação de sistemas com grandes conjuntos de regressores candidatos, na identificação de um reator tanque agitado contínuo e em uma coluna C3/C4. Os resultados mostram que os métodos propostos podem ser utilizados para identificar ambas estruturas, lineares e não lineares com um número reduzido de iterações, tempo de processamento computacional e número de modelos explorados.

*Palavras Chave: Identificação de sistemas, seleção da estrutura aleatória, taxa de redução do erro, seleção da estrutura, NARX, MIMO.*

Fagundes, L. P., **RANDOMIZED MODEL STRUCTURE SELECTION AIDED BY ERROR REDUCTION RATE AND GENETIC INHERITANCE**. 2022. 132 s. Phd Thesis, Universidade Federal de Uberlândia, Uberlândia.

## ABSTRACT

The systems present in industry and science are commonly non-linear, which has made methods for selecting the structure of this type of system widely studied over the last thirty years. There are several methods in literature to deal with the model structure selection in system identification, although these methods have their specific benefits, they face some difficulties in selecting a parsimonious model structure. In this work, two methods based on the Randomized Model Structure Selection (RaMSS), are introduced in order to deal with the model structure selection problem. The first one, named Randomized Model Structure Selection with Error Reduction Ratio (RaMSS-ERR), uses the error reduction rate as a pre-filter in the terms analysis, improving the convergence, the second one, Randomized Model Structure Selection with Genetic Inheritance (RaMSS-EGI) uses a genetic inheritance in order to get faster convergence. The methods were applied to reference models commonly used in the literature of single input and single output, and in models with multiple input and multiple outputs identifying it with more parsimony than some methods present in the literature. They were also applied in the identification of systems with a large candidate regressor set, in the identification of a continuous stirred-tank reactor, and in a C3/C4 column system. The results show that the proposed methods may be used to identify both linear and nonlinear model structures with a reduced number of iterations, computational time, and number of explored models.

*Keywords: System Identification, randomized model structure selection, error reduction rate, Structure Selection, NARX, MIMO.*



## Lista de Figuras

2.1	Fluxograma Geral de um Algoritmo Genético. . . . .	35
3.1	Evolução da probabilidade de inclusão dos regressores (RIP), (a) RaMSS e (b) RaMSS-EGI. . . . .	43
3.2	Algoritmo Geral para RaMSS-ERR e RaMSS-EGI. . . . .	45
3.3	Processo para gerar indivíduo no algoritmo RaMSS-EGI . . . . .	52
5.1	Frequência de modelos extraídos pelos algoritmos RaMSS e RaMSS-ERR . . .	71
5.2	Tempo decorrido para convergir e seleção correta em função do tamanho da população, (a), (c), (e) (g) para RaMSS-EGI, (b), (d), (f) (h) para RaMSS-ERR. . . . .	73
5.3	Comparação dos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI nos sistemas $S_4$ e $S_2$ com sinal de entrada pouco excitativo. . . . .	77
5.4	Sistema CSTR. . . . .	82
5.5	Sinais usados na identificação do sistema CSTR. Entrada $u$ , saídas $C_A$ e $C_B$ , respectivamente em a, b e c. . . . .	84
5.6	Resposta dos modelos identificados usando RaMSS-ERR para a concentração $C_A$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz $C_{AReal} \times C_{AEstimado}$ (d). . . . .	86
5.7	Resposta dos modelos identificados usando RaMSS-ERR para a concentração $C_B$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz $C_{BReal} \times C_{BEstimado}$ (d). . . . .	87

5.8	Resposta dos modelos identificados usando RaMSS-EGI para a concentração $C_A$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz $C_{A_{Real}} \times C_{A_{Estimated}}$ (d). . . . .	88
5.9	Resposta dos modelos identificados usando RaMSS-EGI para a concentração $C_B$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz $C_{B_{Real}} \times C_{B_{Estimated}}$ (d). . . . .	89
5.10	Tempo decorrido até a convergência (a) e (b), e Seleção Correta (c) e (d), todos em função do tamanho da população para o método RaMSS-EGI. . . . .	94
5.11	Sistema de separação C3/C4. . . . .	97
5.12	Sinais usados na identificação do sistema C3/C4. Entradas $u_1$ e $u_2$ , saídas $y_1$ e $y_2$ , respectivamente em (a), (b), (c) e (d). . . . .	99
5.13	Resposta dos modelos identificados usando RaMSS-EGI para a concentração $y_1$ . Teste dinâmico (a), resíduos (b), primeira bissetriz $y_{1_{Real}} \times y_{1_{Estimado}}$ (c). . . . .	101
5.14	Resposta dos modelos identificados usando RaMSS-EGI para a concentração $y_2$ . Teste dinâmico (a), resíduos (b), primeira bissetriz $y_{2_{Real}} \times y_{2_{Estimado}}$ (c). . . . .	102

---

## Lista de Tabelas

5.1	Modelos de Benchmark usados para testar o desempenho dos métodos RaMSS-ERR e RaMSS-EGI. . . . .	68
5.2	Influência do ERR nos algoritmos aleatórios com diferentes valores iniciais de RIP . . . . .	70
5.3	Análise comparativa dos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI. . . . .	75
5.4	Resultados médios para os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI aplicados aos sistemas $\mathcal{S}_6$ e $\mathcal{S}_7$ . . . . .	79
5.5	Resultados médios para os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI aplicados ao sistema $\mathcal{S}_8$ . . . . .	81
5.6	Valores dos parâmetros para o processo no CSTR. . . . .	82
5.7	Métricas obtidas na identificação do CSTR usando ambos os métodos RaMSS-ERR and RaMSS-EGI. . . . .	85
5.8	Modelos de Benchmark MIMO usados para testar o desempenho dos métodos propostos. . . . .	92
5.9	Análise comparativa dos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI para os sistemas MIMO. . . . .	95
5.10	Métricas obtidas na identificação do sistema de separação C3/C4 usando RaMSS-ERR e RaMSS-EGI. . . . .	98



## Lista de Símbolos

$A$	- Fração de simulações que resultaram na seleção correta da estrutura do modelo
$AMS$	- Tamanho médio do modelo ( <i>Average Model Size</i> )
$BIC$	- Critério bayesiano de informação ( <i>Bayesian Information Criteria</i> )
$CR$	- Taxa de convergência ( <i>Convergence Rate</i> )
$d$	- Atraso puro de tempo
$ERR$	- Taxa de redução do erro ( <i>Error Reduction Ratio</i> )
$EM$	- Número de diferentes modelos explorados ( <i>Explored Models</i> )
$ET$	- Tempo transcorrido até a convergência ( <i>Elapsed Time</i> )
$\mathbf{e}(k)$	- Vetor de ruídos presente nos sistemas
$F$	- Relação de não linearidade que pode ser uma função polinomial, racional, uma rede neural ou ainda rede de funções de base radial
$F.AMS$	- Tamanho médio final dos modelos no processo de identificação ( <i>Final Average Model Size</i> )
$FROLS$	- Algoritmo de mínimos quadrados ortogonais com regressão direta ( <i>Forward Orthogonal Least Squares</i> )
$g_i$	- Parâmetros do modelo ortogonalizado
$iOFR$	- Regressão ortogonal direta iterativa ( <i>Iterative Orthogonal Forward Regression</i> )
$IT$	- Número de iterações ( <i>Iteration</i> )
$\mathcal{I}_j^p$	- Índice que mede a importância dos regressores
$\mathcal{J}_{pre}$	- Índice de performance baseado no MSPE
$\mathcal{J}_{sim}$	- Índice de performance baseado no MSSE
$l$	- Grau de não linearidade do modelo
$m$	- Número total de termos candidatos ao modelo
$M$	- População de modelos
$M.AMS$	- Tamanho médio máximo dos modelos durante o processo de identificação ( <i>Maximum Average Model Size</i> )

$MQO$	- Mínimos quadrados ortogonais
$MSPE$	- Média quadrática do erro de predição (Mean square prediction error)
$MSSE$	- Média quadrática do erro de simulação (Mean Square Simulation Error)
$N$	- Número de amostras do sinal de entrada e saída
$n_e$	- Atraso máximo do ruído
$n_u$	- Atraso máximo da saída
$n_y$	- Atraso máximo da entrada
$\mathcal{O}$	- Caminho de ortogonalização
$OFR$	- Algoritmo ortogonal com regressão direta ( <i>Orthogonal Forward Regression</i> )
$\mathcal{P}$	- Distribuição de Probabilidade de Inclusão de Regressores
$\mathcal{P}_\phi$	- Distribuição de probabilidade para um dado modelo $\phi$
$RaMSS$	- Seleção aleatória da estrutura do modelo ( <i>Randomized Model Structure Selection</i> )
$RaMSS - ERR$	- Seleção aleatória da estrutura do modelo com taxa de redução do erro ( <i>Randomized Model Structure Selection with Error Reduction Error</i> )
$RaMSS - EGI$	- Seleção aleatória da estrutura do modelo com taxa de redução do erro e herança genética ( <i>Randomized Model Structure Selection with Error Reduction Error and Genetic Inheritance</i> )
$RIP$	- Probabilidade de inclusão do regressor ( <i>Regressor Inclusion Probability</i> )
$\mathcal{R}_e$	- Subconjunto de termos do dicionário $\mathcal{R}_m$
$\mathcal{R}_m$	- Conjunto de regressores candidatos considerado na seleção da estrutura
$SRR$	- Taxa de redução do erro de simulação (Simulation error reduction error)
$u_i$	- Entrada do sistema, na identificação de sistemas não lineares, usualmente é um sinal com amplo espectro de frequência
$\mathbf{u}(k)$	- Vetor de entradas do sistema MIMO no tempo $k$
$w_i$	- Regressores do modelo ortogonalizado
$WGN(\cdot)$	- Ruído branco gaussiano (White Gaussian Noise)
$WUN(\cdot)$	- Ruído branco uniforme (White Uniform Noise)
$y(k)$	- saída do sistema ou modelo no tempo $k$
$Z_M$	- Número total de indivíduos na população $M$
$\mathbf{y}(k)$	- Vetor de saídas do sistema MIMO no tempo $k$
$\beta$	- Parâmetro que define se o índice de performance será baseado em predição ou simulação
$\delta$	- limiar para seleção de regressores no ERR
$\gamma$	- Parâmetro de sintonia da atualização da RIP
$\theta_i$	- Vetor de parâmetros do modelo
$\Theta$	- Vetor de parâmetros do modelo na equação em forma matricial
$\mu_j$	- Probabilidade de selecionar um regressor $j$
$\xi$	- Ruído presente no modelo, usualmente considerado ruído branco
$\kappa$	- Limiar que serve como critério de parada no algoritmo ERR
$\rho_j$	- Variável de Bernoulli
$\hat{\sigma}_j$	- Variância do $j$ -ésimo regressor
$\hat{\sigma}_e$	- Variância do erro
$\tau$	- Número de regressores no modelo
$\phi$	- Variável discreta que assume valores em $\mathcal{F}$
$\psi_i$	- Regressores do modelo
$\Psi$	- Matriz de regressores

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	Motivação . . . . .	3
1.2	Justificativas . . . . .	5
1.2.1	O Problema da Seleção da Estrutura . . . . .	5
1.2.2	Formulação do Problema . . . . .	6
1.3	Objetivos Gerais e Específicos . . . . .	8
1.4	Organização do Trabalho . . . . .	9
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>11</b>
2.1	Estado da Arte . . . . .	11
2.1.1	Sistemas SISO . . . . .	12
2.1.2	Sistemas MIMO . . . . .	14
2.2	Modelos NARX para o caso SISO . . . . .	16
2.3	Modelos NARX para o caso MIMO . . . . .	17
2.4	Estimador de Mínimos Quadrados Ortogonais e Taxa de Redução do Erro (ERR)	18
2.4.1	Estimador de Mínimos Quadrados Ortogonais . . . . .	19
2.4.2	Taxa de Redução do Erro . . . . .	21
2.5	Algoritmo ortogonal com regressão direta (OFR) . . . . .	23
2.6	Teste-t de Student . . . . .	29
2.7	Desempenho do Modelo . . . . .	31
2.7.1	Erro Quadrático Médio . . . . .	32
2.7.2	CrITÉrio da Informação Bayesiana . . . . .	33
2.8	Herança Genética . . . . .	34

2.9	Considerações Finais . . . . .	36
<b>3</b>	<b>IDENTIFICAÇÃO DE SISTEMAS SISO</b>	<b>39</b>
3.1	Formulação do Problema e a Seleção Aleatória da Estrutura de Modelos SISO	39
3.1.1	Formulação do Problema de Seleção da Estrutura de Modelos SISO . .	39
3.1.2	Seleção Aleatória da Estrutura de Modelos . . . . .	40
3.2	Métodos propostos para o caso SISO . . . . .	43
3.3	Algoritmo Geral para o caso SISO . . . . .	44
3.3.1	Gerar População Inicial . . . . .	47
3.3.2	Estimação dos parâmetros e teste-t de Student . . . . .	48
3.3.3	Atualização da probabilidade de inclusão de regressores (RIP) . . . . .	50
3.3.4	RaMSS-ERR e RaMSS-EGI . . . . .	51
3.4	Considerações Finais . . . . .	53
<b>4</b>	<b>IDENTIFICAÇÃO DE SISTEMAS MISO E MIMO</b>	<b>55</b>
4.1	Formulação do Problema e a Seleção Aleatória da Estrutura de Modelos MIMO	55
4.1.1	Formulação do Problema de Seleção da Estrutura de Modelos MIMO .	55
4.1.2	Seleção Aleatória da Estrutura de Modelos MIMO . . . . .	57
4.2	Métodos propostos para o caso MIMO . . . . .	59
4.3	Algoritmo Geral para o caso MIMO . . . . .	59
4.3.1	Gerar população inicial no caso MIMO . . . . .	61
4.3.2	Estimação dos parâmetros e teste-t de Student no caso MIMO . . . .	62
4.3.3	Atualização da probabilidade de inclusão do regressor (RIP) no caso MIMO . . . . .	63
4.3.4	Gerar população para o método RaMSS-EGI no caso MIMO . . . . .	64
4.4	Considerações Finais . . . . .	65
<b>5</b>	<b>AVALIAÇÃO DOS ALGORITMOS DE IDENTIFICAÇÃO PROPOSTOS</b>	<b>67</b>
5.1	Sistemas SISO . . . . .	67
5.1.1	Modelos SISO utilizados como referência para testes . . . . .	68
5.1.2	Parâmetros para o caso SISO . . . . .	69

5.1.3	Sensibilidade dos algoritmos aleatórios ao uso da taxa de redução do erro (ERR) . . . . .	69
5.1.4	Tamanho da População . . . . .	71
5.1.5	Análise e testes comparativos para o caso SISO . . . . .	72
5.1.6	Sistemas com entrada pouco excitativa . . . . .	76
5.1.7	Identificação de sistemas com alto grau de não-linearidade . . . . .	77
5.1.7.1	Sistemas com Resposta Inversa . . . . .	78
5.1.7.2	Modelo Artificial com alto grau de não linearidade . . . . .	80
5.1.8	Identificação do Reator Tanque Agitado . . . . .	81
5.1.8.1	Identificação usando RaMSS-ERR . . . . .	83
5.1.8.2	Identificação usando RaMSS-EGI . . . . .	85
5.1.9	Discussão sistemas SISO . . . . .	85
5.2	Sistemas MISO e MIMO . . . . .	91
5.2.1	Modelos MISO e MIMO utilizados como referência para testes . . . . .	91
5.2.2	Parâmetros para o caso MIMO . . . . .	92
5.2.3	Tamanho da População e Probabilidade de Inclusão dos Regressores . . . . .	92
5.2.4	Análise e testes comparativos para o caso MIMO . . . . .	95
5.2.5	Identificação da Coluna de Separação C3/C4 . . . . .	96
5.2.6	Discussão sistemas MISO e MIMO . . . . .	103
<b>6</b>	<b>CONCLUSÕES E SUGESTÕES</b>	<b>105</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>109</b>



# CAPÍTULO I

## INTRODUÇÃO

Modelos matemáticos de sistemas podem ser úteis em todas as fases da engenharia, desde a sua concepção até a sua operação. A modelagem matemática tem grande importância nas ciências e engenharia, pois permite um melhor entendimento dos sistemas estudados, desde a pesquisa e desenvolvimento, até o treinamento de técnicos, a realização de projetos para sistemas complexos de diversas áreas, entre outros.

Tais modelos consistem em uma equação ou conjunto de equações capazes de caracterizar a dinâmica de um sistema, permitindo assim, que seja inferida informações a respeito do sistema.

Neste contexto a identificação de sistema surge como um ramo da modelagem capaz de fornecer modelos que descrevem o comportamento de um sistema por meio do processamento de suas entradas e saídas (Billings, 2013).

Em termos práticos, a complexidade dos sistemas reais e a existência de dados não observáveis impedem a concepção de um modelo que seja uma descrição exata dos sistemas (Keesman, 2011). Contudo, mesmo que houvesse dados disponíveis para a concepção de um modelo que representasse o sistema em questão nos mínimos detalhes, tal modelo seria tão complexo que inviabilizaria seu uso prático.

Na prática, o que se tem são modelos capazes de representar apenas satisfatoriamente os sistemas, a depender da aplicação, isto é, nem sempre um modelo bom para predição

será também um bom modelo para simulação.

Uma classe de modelos muito comum e que será objeto de estudo no presente trabalho é aquela que representa os modelos como uma equação com regressores, isto é, referências às entradas e saídas passadas que são utilizadas para descrever o modelo.

Usualmente a identificação de sistemas com base em modelos que utilizam regressores, consiste em cinco passos (Boynton et al., 2018), sendo eles:

- **Determinação da estrutura:** Nesse passo o foco está em definir se será adotado uma estrutura linear ou não linear e identificar quais regressores devem fazer parte do modelo.
- **Estimação de parâmetros:** Uma vez identificada a estrutura do modelo, deve-se estimar os parâmetros. Geralmente tal procedimento é feito usando métodos baseado na minimização do erro de predição, tais como mínimos quadrados ortogonais ou mínimos quadrados estendidos.
- **Validação do modelo:** Com o modelo em mãos é feito uma validação deste, a qual pode ser estatística, utilizando testes de correlação, ou dinâmica, usando simulações livres, comparando a capacidade do modelo obtido em representar as dinâmicas contidas nos dados de identificação.
- **Predição:** Nessa etapa o sistema é simulado em n-passos à frente para se determinar sua capacidade de predição, quando tem-se por objetivo a aplicação do modelo em técnicas de controle. O sistema pode ainda ser utilizado em simulação livre, ou infinitos passos à frente, quando tem-se por objetivo a aplicação do modelo em otimização de sistemas ou treinamento de operadores.
- **Análise:** Uma vez que o modelo obtido foi avaliado e considerado adequado, o mesmo pode ser utilizado para se obter informações a respeito da dinâmica do sistema, para fins de controle, simulação, detecção de falhas ou mesmo treinamento de operadores em processos.

Dentre esses passos listados, ocupa posição de destaque a etapa de determinação da estrutura, muitas vezes denominada também de seleção da estrutura. Nesta etapa, o grau

de não linearidade e os atrasos das variáveis de entrada e saída do modelo tomam um papel central na complexidade do problema.

## 1.1 Motivação

Os modelos dinâmicos obtidos a partir do processo de identificação de sistemas podem ser aplicados de diversas maneiras, dentre elas destacam-se predição, simulação, otimização, análise, controle e detecção de falhas (Nelles, 2001).

- Predição: são utilizados dados de entradas e saídas passadas aplicadas ao modelo para prever o comportamento futuro, muito utilizado em controle preditivo, na previsão do tempo e mercado de ações.
- Simulação: é utilizada na otimização, controle e detecção de falhas. Difere da predição por utilizar apenas a entrada. É também chamada de simulação livre ou predição de infinitos passos à frente.
- Otimização: busca-se um ponto de operação ótimo ou um perfil de entradas que melhore o desempenho do sistema considerado. A vantagem da otimização utilizando modelo está no fato de que muitas vezes a otimização utilizando sistemas reais demanda muito tempo, pode levar o sistema a condições indesejadas de operação, ou ainda a um desperdício de matéria prima, problemas esses facilmente contornáveis quando se faz a otimização utilizando modelos.
- Análise: alguns conhecimentos intuitivos podem ser adquiridos simplesmente analisando como o modelo responde a determinadas entradas. É possível por exemplo inferir regras de controle *fuzzy*.
- Controle: muitas técnicas de controle avançado são baseadas no modelo. Nesse caso o modelo obtido é a base para a criação de estratégias de controle. Um exemplo é o controle preditivo baseado em modelo.
- Detecção de falhas: comparando-se a saída do sistema com a saída do modelo e de posse da entrada aplicada no sistema, é possível observar falhas no sistema em questão,

evitando situações perigosas, danos ao meio ambiente ou avarias em ativos.

As vantagens de se ter um modelo dinâmico capaz de descrever o comportamento de um sistema são muitas. Por meio do modelo, computadores podem ser programados para se comportar como os sistemas em estudo e, assim, diversos experimentos podem ser feitos, tais como: diferentes projetos podem ser testados sem a necessidade de gastar recursos com sua construção física, experimentos que seriam perigosos no mundo real podem ser executados com total segurança, respostas de como o sistema se comporta frente a determinadas entradas podem ser obtidas sem gasto de insumos ou excesso de tempo, dentre outras possibilidades (Billings, 2013).

Todas as vantagens e aplicações citadas anteriormente, evidenciam a necessidade de se obter modelos capazes de descrever sistemas dinâmicos. Nesse contexto, a identificação de sistemas tem importância fundamental por permitir a obtenção de modelos capazes de atender as especificações de projeto, poder ser aplicada sem o dispêndio de grandes aportes financeiros, não necessitar de mão de obra altamente especializada para cada etapa de modelagem do sistema, além de permitir que uma mesma técnica de identificação seja aplicada nos mais variados tipos de sistemas.

Neste contexto, a utilização de algoritmos de seleção aleatória da estrutura dos modelos tem ganhado destaque por ser capaz de fornecer modelos eficientes e simples, evitando o aparecimento de dinâmicas espúrias, mais eficiente do ponto de vista computacional quando comparado com outros métodos de seleção da estrutura, além disso, os algoritmos de seleção aleatória da estrutura dispensam o conhecimento prévio sobre a estrutura dos modelos, tendo também um ótimo desempenho quando o espaço de busca se torna muito grande (Falsone *et al*, 2015).

Este trabalho apresenta um processo para seleção da estrutura de modelos em um cenário probabilístico combinando a taxa de redução do erro (ERR - *Error Reduction Ratio*) e herança genética. O uso do ERR como um pré-filtro na seleção da estrutura elimina os regressores que contribuem pouco para a diminuição da variância da saída. Já o processo de herança genética garante a transmissão de bons regressores entre gerações, além de manter na população os melhores indivíduos já encontrados. Essa combinação de ERR e herança

genética garante uma convergência em menor tempo e número de iterações, fazendo com que o algoritmo tenha uma performance melhor na seleção da estrutura de modelos quando o conjunto de regressores se torna grande (acima de 10.000 regressores).

## 1.2 Justificativas

### 1.2.1 O Problema da Seleção da Estrutura

Os modelos denominados NARX (Nonlinear AutoRegressive with Exogenous Input) foram introduzidos por Billings e Leontaritis (1985) como uma nova proposta de representação de sistemas não lineares sujeitos a entradas exógenas e são o foco do presente trabalho.

A seleção da estrutura para modelos NARX consiste em um problema de alta complexidade devido ao grande espaço de busca, o qual aumenta consideravelmente com o grau de não linearidade e com os atrasos.

A diminuição do espaço de busca pode ser feita conhecendo-se características prévias do sistema a ser identificado, ou com base em algoritmos de pré-processamento de dados. Assim a diminuição do número total de regressores candidatos reduz o espaço de busca e pode ser obtido conhecendo-se a ordem do sistema, isto é, os atrasos. Neste contexto, destacam-se os trabalhos de Bomberger e Seborg (1998), Kukreja *et al.* (2000), Feil *et al.* (2004) e Wei *et al.* (2004) onde métodos para seleção da ordem são propostos. Embora estes métodos diminuam o espaço de busca, não resolvem o problema de seleção da estrutura, visto que mesmo aplicando estes métodos o espaço de busca pode ainda ser grande.

Dentre vários métodos propostos para seleção da estrutura, o algoritmo ortogonal com regressão direta (*Orthogonal forward regression with error reduction rate - OFR-ERR*) proposto em Korenberg *et al.* (1988), é, indiscutivelmente, o mais popular. Tal algoritmo tem como base a ideia de desacoplar a cada passo, os regressores em um espaço ortogonal, de forma que a contribuição de cada regressor possa ser quantificada. Dessa forma, os regressores são incluídos no modelo um por vez, considerando sempre a contribuição deste para explicar a variância da saída, essa métrica é denominada taxa de redução do erro (*Error Reduction Ratio - ERR*) e é amplamente utilizada na literatura.

Mao e Billings (1997) mostraram que a principal desvantagem do método OFR-

ERR é que este executa uma busca local, isto é, o ERR depende diretamente do caminho de ortogonalização e é altamente dependente da ordem na qual os regressores são ortogonalizados, o que significa que nem sempre o OFR-ERR é capaz de encontrar um modelo adequado.

Outro problema relacionado ao ERR se refere ao sinal de entrada. Para propósitos de identificação é sempre desejável que o sinal de entrada seja um ruído branco (Aguirre, 2015), o que nem sempre é possível em sistemas reais, que geralmente apresentam características em baixa frequência.

Piroddi e Spinelli (2003) mostraram que o algoritmo OFR-ERR produz modelos ruins quando o sinal de entrada é pouco excitativo ou filtrados, situações estas que não são as ideais para a identificação de um sistema, mas que servem para avaliar a robustez dos métodos. No mesmo artigo, Piroddi e Spinelli apresentam a taxa de redução do erro de simulação (*Simulation Error Rate - SRR*), que apresenta desempenho superior ao ERR, contudo é altamente dispendiosa no sentido computacional.

Muitas técnicas para selecionar a melhor estrutura de modelos não lineares têm sido sugeridas na literatura (Billings *et al.*, 1988; Aguirre e Billings, 1995; Mao e Billings, 1997; Baldacchino *et al.*, 2012; Billings *et al.* 2015; Solares e Wei, 2015; Avelina *et al.*, 2017; Gu e Wei, 2018; Retes e Aguirre, 2019; Hafiz *et al.*, 2019a), contudo muitas vezes estas técnicas dependem de fatores tais como o período de amostragem, ruído de medição ou ainda de considerações muito restritivas sobre os ruídos. Além disso, nenhuma das técnicas é significativamente melhor que as demais em todas as ocasiões.

### 1.2.2 Formulação do Problema

A seleção da estrutura dos modelos é a etapa mais desafiadora do processo de identificação de sistemas. Esta etapa consiste em selecionar um subconjunto  $\mathcal{R}_e$  a partir de um dado conjunto de regressores candidatos  $\mathcal{R}_m$ , com  $\mathcal{R}_e \subset \mathcal{R}_m$ , pela otimização de uma função critério adequada, da seguinte forma:

$$\mathcal{R}_e = \arg \min_{\mathcal{R}_e \subset \mathcal{R}_m} J(\mathcal{R}_m) \quad (1.1)$$

Ao contrário do que é feito muitas vezes na identificação de sistemas lineares, no

contexto de sistemas não lineares, esta tarefa dificilmente pode ser executada usando técnicas baseadas em critérios da informação, como o critério da informação de Akaike (*Akaike Information Criterion - AIC*) e o critério de informação Bayesiano (*Bayesian Information Criterion - BIC*) (Schwarz, 1978), devido ao fato de que estes métodos contrapõem a acurácia do modelo e a quantidade de termos, e muitos modelos com o mesmo tamanho e diferentes desempenhos podem ser construídos com diferentes regressores (Falsone *et al.*, 2015).

Como mencionado previamente, o grau de não linearidade e os atrasos possuem um papel central na determinação da complexidade do processo de seleção da estrutura. Isto se deve ao fato de que esta etapa sofre de um problema muito conhecido em problemas de otimização chamado *maldição da dimensionalidade*. Esse termo se refere ao fato do espaço de busca crescer muito com o aumento dos atrasos e grau de não linearidade. Por exemplo, para o modelo NARX, a quantidade de termos candidatos é dada pelas Equações (1.2) e (1.3).

$$m = \binom{n + l - 1}{l} \quad (1.2)$$

$$n = n_y + (n_u - d + 1) + n_e + 1 \quad (1.3)$$

em que  $l$  é o grau de não linearidade do modelo,  $n_y$ ,  $n_e$  e  $n_u$  são respectivamente os atrasos da saída, os atrasos dos erros (quando inseridos no modelo) e atrasos da entrada, e  $d$  é um atraso puro de tempo, usualmente considerado como  $d = 1$ .

Considere como exemplo um sistema com grau de não linearidade 2 e atrasos máximos 2, para tal sistema tem-se 28 regressores candidatos, aumentando o grau de não linearidade para 5 e os atrasos máximos para 6, números estes ainda pequenos para representar sistemas reais, tem-se 33.649 regressores candidatos. Em Peng *et al.* (2011) é apresentado um modelo com  $n_y = 96$ ,  $n_u = 93$  e  $n_e = 0$ , grau de não linearidade  $l = 3$ , resultando em um dicionário de termos de 1.161.280 regressores candidatos. Caso fosse necessário identificar um modelo com cinco regressores usando esse dicionário, seria necessário validar  $2,11 \times 10^{30}$  modelos, o que seria altamente oneroso do ponto de vista computacional.

Estabelecer valores muito altos para os atrasos e para o grau de não linearidade pode fazer com que o processo de identificação se torne demasiadamente demorado, e o oposto,

isto é, estabelecer valores baixos, pode fazer com que o conjunto de regressores candidatos não contemple os melhores regressores capazes de representar o sistema com o menor número de termos, o que, normalmente, é compensado acrescentando-se ao modelo vários outros regressores, além de aumentar o custo computacional, pois mais regressores são analisados, resulta também em modelos com regimes dinâmicos espúrios (Aguirre e Billings, 1995).

No presente trabalho é tratada também a identificação de sistemas com Múltiplas Entradas e Múltiplas Saídas (*Multiple input and Multiple Output - MIMO*), os quais possuem um maior número de entradas e saídas, e conseqüentemente um maior número de atrasos e regressores candidatos, tornam o processo de seleção da estrutura ainda mais complicado devido ao aumento de regressores candidatos. Considerando um sistema MIMO com  $q$  entradas e  $s$  saídas, a Equação (1.3) se torna:

$$n = n_{y_1} + n_{y_2} + \dots + n_{y_s} + (n_{u_1} + n_{u_2} + \dots + n_{u_q} - d + 1) + 1 \quad (1.4)$$

Para um sistema MIMO simples, com duas entradas e duas saídas apenas, considerando atrasos de  $n_{y_1} = n_{y_2} = 96$ ,  $n_{u_1} = n_{u_2} = 93$  e  $n_e = 0$ , não linearidade  $l = 3$ , o dicionário de termos teria 9.145.270 regressores candidatos. E conseqüentemente, para um modelo com cinco regressores usando esse dicionário de termos, teria-se  $6,39 \times 10^{34}$  modelos a serem validados. Assim nota-se que a identificação da estrutura de modelos MIMO é ainda mais complexa do que para modelos SISO.

### 1.3 Objetivos Gerais e Específicos

A literatura recente tem focado em soluções que melhoram a estratégia de busca, como algoritmos evolutivos e neuronais, e em algoritmos que fazem uso da maior disponibilidade computacional da atualidade, como algoritmos de seleção aleatória e de computação distribuída (Falsone *et al*, 2015; Avelina *et al*, 2017; Hafiz *et al*, 2019a).

Ainda que os métodos citados anteriormente apresentem suas contribuições, o problema da seleção da estrutura ainda permanece em aberto. Destarte, o presente trabalho tem como objetivo geral o desenvolvimento de um método para seleção da estrutura de modelos SISO e MIMO, fazendo uso de algoritmos de seleção aleatórios auxiliados pela taxa de redução

do erro e por herança genética, visando acelerar a convergência da função de probabilidade de inclusão dos regressores.

Os objetivos específicos deste trabalho são:

- Desenvolver um algoritmo capaz de selecionar a estrutura de modelos SISO e MIMO fazendo uso de algoritmos de seleção aleatórios auxiliados pela taxa de redução do erro e por herança genética;
- Avaliar o desempenho do algoritmo proposto na identificação de sistemas SISO e MIMO com diferentes tipos de não linearidades, bem como identificar os parâmetros de influência na convergência;
- Avaliar o desempenho do algoritmo proposto na identificação de sistemas com entrada pouco excitativa;
- Avaliar o desempenho do algoritmo na identificação de sistemas com alto grau de não linearidade e longos atrasos;
- Aplicar o algoritmo proposto na identificação de um sistema reator tanque agitado contínuo e em uma coluna de separação C3/C4.

#### **1.4 Organização do Trabalho**

A presente tese está dividida da seguinte forma: o Capítulo 1 aborda uma introdução ao tema, no Capítulo 2 é traz a fundamentação teórica para desenvolvimento do método proposto, o Capítulo 3 apresenta a proposta de trabalho para sistemas de uma entrada e uma saída (SISO) e o Capítulo 4 apresenta o algoritmo proposto para o caso de múltiplas entradas e múltiplas saídas (MIMO), no Capítulo 5 os métodos propostos são aplicados em diversas situações, tanto para sistemas SISO quanto para sistemas MIMO. E por fim, o trabalho é concluído no Capítulo 6 com algumas considerações sobre os resultados e sugestões de trabalhos futuros.



# CAPÍTULO II

## REVISÃO DA LITERATURA

O presente capítulo tem por objetivo apresentar o estado da arte concernente à seleção da estrutura de modelos, bem como solidificar o conhecimento a respeito de aspectos necessários para a implementação da proposta. Na discussão que se segue é apresentado o estado da arte desenvolvido para os sistemas SISO e posteriormente para os sistemas MIMO. Em seguida são revisados conceitos sobre os modelos NARX para sistemas SISO e posteriormente para sistemas MIMO. Então passa-se a uma apresentação de algumas ferramentas essenciais para a proposição do método, tais como algoritmos de mínimos quadrados ortogonais, taxa de redução do erro, algoritmo ortogonal com regressão direta, teste-t de Student, índices de desempenho dos modelos e herança genética.

### 2.1 Estado da Arte

Visando resolver os problemas inerentes à seleção da estrutura, tem surgido na literatura diversas vertentes de trabalhos, alguns que visam à diminuição do espaço de busca, outros focados em características métricas, e outros em como a busca pelos melhores regressores é feita. O presente trabalho tem por objetivo tratar o problema da seleção da estrutura de modelos sob duas óticas, dos sistemas de uma entrada e uma saída (sistemas SISO) e dos sistemas com múltiplas entradas e múltiplas saídas (sistemas MIMO). Assim a revisão da

literatura foi dividida sob o aspecto desses dois tipos de sistemas.

### 2.1.1 *Sistemas SISO*

Na literatura corrente, o problema da seleção da estrutura tem sido atacado em duas vertentes distintas. Tem-se de um lado soluções focadas na mudança do índice métrico e de outro soluções focadas na mudança da estratégia de busca.

As soluções focadas no índice métrico tentam melhorar ou substituir a taxa de redução do erro por outra métrica, na literatura destacam-se os seguintes métodos.

- *Simulation Error*: Proposto em Piroddi e Spinelli (2003), este método substitui o ERR por uma taxa de redução do erro de simulação. Os regressores candidatos são testados um por vez e inseridos no modelo com base na diminuição do erro de simulação. Apresenta como principal vantagem uma maior robustez na seleção dos regressores quando os sinais de entradas não são adequadamente projetados, e como desvantagem o alto custo computacional, uma vez que os regressores candidatos podem ser testados várias vezes, a depender do tamanho do modelo.
- *Mutual Information*: Apresentado em Wei e Billings (2008), o ERR é substituído por um coeficiente de informação mútua. Apesar da melhora, o algoritmo ainda é suscetível a seleção de termos espúrios e requer o uso de um teste-t de Student para eliminar tais regressores.
- *Distância de Correlação*: Em Solares e Wei (2015) é proposto um algoritmo usando distancia de correlação. O índice métrico é capaz de quantificar todos os tipos de não linearidades e dependências não monótonas entre vetores. Dentre as vantagens pode-se citar a melhora na identificação de modelos quando comparado com outros índices métricos, e como desvantagem apresenta alto custo computacional.
- *Ultra Orthogonal Least Square*: Apresentado em Guo *et al.* (2016) essa métrica considera não apenas os dados, mas também a relação entre eles por meio da estimação das derivadas entre os pontos. Apesar de apresentar resultados melhores do que o ERR, esse

método é mais sensível ao ruído e apresenta um aumento da complexidade do algoritmo, visto que é necessário o cálculo das derivadas fracas.

Já dentre as soluções focadas na estratégia de busca, que tentam melhorar a forma como os algoritmos percorrem o espaço factível, pode-se citar:

- Algoritmos Genéticos: Mao e Billings (1997) utilizaram um algoritmo genético para tentar encontrar o melhor caminho de ortogonalização identificando um subconjunto usando o OFR clássico, e então o refinando por meio do algoritmo genético. Embora essa técnica demande grande esforço computacional, este é bem menor do que uma busca exaustiva, mas não garante ainda a otimalidade da solução.
- Regressão direta ortogonal iterativa: Proposta em Guo *et al.* (2014), o iOFR (*Iterative Orthogonal Forward Regression*) consiste em aplicar iterativamente o algoritmo OFR de forma que a próxima busca seja baseada no subconjunto de regressores obtidos no passo anterior, de modo que cada regressor no subconjunto dos regressores iniciais seja um pivô para identificar novos subconjuntos na busca secundária. Esse método também está sujeito ao problema do caminho de ortogonalização e apesar de ser uma clara melhora do clássico OFR-ERR, ainda não garante modelos parcimoniosos.
- Algoritmos de Seleção Aleatória: Falsone *et al.* (2015) apresentaram um algoritmo randômico para a seleção da estrutura de modelos, baseado em uma função de probabilidade que é progressivamente atualizado durante o processo de busca. Apesar de ser capaz de identificar uma variedade de sistemas, com o aumento da ordem dos sistemas e na identificação de modelos com grandes atrasos, os algoritmos aleatórios experimentam problemas de processamento computacional de precisão na seleção da estrutura dos modelos. Em um sistema com alto grau de não linearidade a precisão em selecionar o sistema corretamente foi de apenas 37%, levando aproximadamente 30 minutos para atingir a convergência.
- Algoritmo Distribuído de Seleção Aleatória: Avellina *et al.* (2017) apresentaram uma extensão do algoritmo apresentado em Falsone *et al.* (2015) usando computação distribuída. Basicamente os grupos de regressores candidatos são distribuídos entre unidades

de processamento e, então, informações sobre os modelos selecionados são trocadas entre as unidades. O algoritmo apresenta uma melhora na precisão da identificação de modelos ao custo de uma maior complexidade do processo de identificação.

- *Orthogonal Floating Search*: Hafiz *et al.* (2019a) apresentaram uma adaptação do algoritmo de busca flutuante, já existente no contexto de seleção de características em redes neurais, aplicado a seleção da estrutura de modelos não lineares. O método elimina o efeito de aninhamento devido ao caminho de ortogonalização contudo pode se tornar oneroso computacionalmente frente a identificação de modelos de alta ordem.
- *Two-Dimensional particle swarms*: Hafiz *et al.* (2019b) propuseram um algoritmo baseado em enxames de partículas para selecionar a estrutura de modelos não lineares, o qual incorpora diretamente o número de termos no processo de busca, o que diminui significativamente o número de termos espúrios no modelo final.

### 2.1.2 *Sistemas MIMO*

Os sistemas multivariáveis são muito comuns nas indústrias de óleo e gás (Al-Naumani e Rossiter, 2015) ou em equipamentos como boilers (Rojas *et al.*, 2012), resfriadores centrífugos (Tudoroiu *et al.*, 2020) e até mesmo em helicópteros (Raj *et al.*, 2020). Tais sistemas usualmente requerem um modelo do sistema para que sejam aplicadas técnicas de controle avançado.

Como dito anteriormente, a seleção da estrutura de modelos MIMO é um problema mais complexo e desafiador do que o caso SISO, por considerar um maior número de entradas e saídas, e conseqüentemente um maior número de regressores candidatos. Neste sentido, citam-se algumas abordagens presentes na literatura:

- *Regressão direta ortogonal para sistemas MIMO*: Billings *et al.*, (1989) propuseram uma versão do algoritmo de regressão direta ortogonal para sistemas MIMO. Assim como no caso SISO, para o caso MIMO o algoritmo busca inserir os regressores um por vez, considerando a contribuição deste para explicar a variância da saída. Apesar de amplamente utilizada, assim como no caso monovariável, esta técnica executa uma

busca local e é altamente influenciada pelo caminho de ortogonalização.

- Identificação iterativa: Para identificar sistemas NARX com múltiplas entradas, Hirsch e del Re (2010) propuseram uma combinação de identificação iterativa com o projeto online das entradas de excitação do sistema. O método proposto é baseado em obter uma nova medição projetada para otimizar o determinante da matriz de informação correspondente ao modelo candidato. Embora o método possa ser usado para identificar modelos complexos, a complexidade computacional para se projetar a entrada para resolver a otimização é um problema.
- Identificação usando rede Neuronal inversa: Ho e Nguyen (2013) propuseram a identificação de um sistema MIMO usando uma rede neuronal inversa, baseada na combinação de uma Rede Neuronal multicamadas Perceptron e a estrutura de um modelo ARX. A principal desvantagem neste método é que os parâmetros da rede não têm nenhuma correspondência com os parâmetros físicos do sistema. Além disso, o modelo da rede neuronal é obtido por tentativas.
- Matriz de informação + banco de modelos SISO: Westwick *et al.*, (2018) propuseram um algoritmo de desacoplamento para sistemas MIMO usando uma matriz de transformação seguida por um banco de modelos SISO polinomiais, simplificando a complexidade do modelo. Entretanto, a desvantagem deste método é que o modelo é não linear nos parâmetros, o que faz com que seja necessária uma otimização não convexa para se estimar os parâmetros.
- Matriz de transformação + banco de polinômios univariados: Karami *et al.*, (2021), seguindo o trabalho de Westwick (2018), propuseram um algoritmo de desacoplamento para sistemas MIMO que substitui os polinômios multivariados por uma matriz de transformação seguida de um banco de polinômios univariados, de forma a diminuir o número de parâmetros do modelo substancialmente. Como no trabalho de Westwick, a principal desvantagem é a necessidade de uma otimização não convexa para se estimar os parâmetros, dado que os modelos são não lineares nos parâmetros.

## 2.2 Modelos NARX para o caso SISO

Dentre várias representações de sistemas não lineares, duas apresentam maior destaque, as séries de Volterra e a representação NARX, esta última apresenta como grande vantagem sobre a primeira, uma redução substancial do número de termos no modelo, por exemplo, sistemas que incorporam uma não linearidade na saída quando representados pela série de Volterra requerem centenas de termos, já com a representação NARX tais sistemas podem ser descritos de forma compacta (Chen e Billings, 1989).

Um modelo NARX é essencialmente a descrição da saída usando entradas e saídas passadas. Uma ampla variedade de sistemas não lineares pode ser descrita pela representação NARX (Leontaritis e Billings, 1985; Billings, 2013) que relaciona a saída presente  $y(k)$  com as entradas  $u(\cdot)$  e saídas passadas de uma forma recursiva

$$y(k) = F[y(k-1), \dots, y(k-n_y), u(k-d), \dots, u(k-d-n_u)] + \xi(k) \quad (2.1)$$

em que  $n_y$  e  $n_u$  são os atrasos de saída e entrada respectivamente,  $d \geq 1$  é o atraso puro e  $\xi(k)$  é um ruído. No que se segue, os termos  $y(k-1)$ ,  $y(k-n_y)$ ,  $u(k-1)$ ,  $u(k-n_u)$  são referidos como regressores.

A função base  $F$  na Equação (2.1) pode assumir formas paramétricas e não-paramétricas tais como polinomial, redes neurais, modelos baseados em lógica fuzzy, expansões wavelets, funções de base radial e outras expansões em funções de base (Wei *et al.*, 2004).

Neste trabalho, a função base nos modelos NARX foram definidas como expansões polinomiais, as quais provêm muitas características atrativas tais como a habilidade de representar uma vasta gama de sistemas não lineares utilizando um pequeno número de parâmetros, uma interpretação do modelo simplificada, associando termos do modelo com aspectos físicos do sistema. Além disso, permite uma análise em frequência utilizando funções de resposta em frequência generalizada (Billings, 2013).

Expandindo a função  $F$  como uma função polinomial resulta na representação

NARX como na Equação (2.2),

$$y(k) = \sum_{i=1}^m \psi_i(k)\theta_i + \xi(k) \quad (2.2)$$

em que  $\psi_i(k)$  é a  $i$ -ésima combinação, linear ou não-linear, de  $y$  e  $u$  que são vetores de dimensão  $N \times 1$ , com grau até  $l$ ;  $m$  é o número de monômios distintos no modelo;  $\xi$  é um vetor de ruído com a mesma dimensão de  $y$  e  $\theta_i$  é o  $i$ -ésimo parâmetro a ser estimado.

As representações polinomiais são simplesmente a expansão da função  $F[\cdot]$  em um determinado grau de não linearidade  $l$  a qual determinará o grau de não linearidade dos regressores. Sua principal vantagem está em formar modelos lineares nos parâmetros que podem ser estimados utilizando técnicas de estimação por mínimos quadrados ortogonais.

Desde a introdução dos modelos NARX por Leontaritis e Billings (Leontaritis e Billings, 1985), este passou de um simples tipo de modelo para uma filosofia de identificação de sistemas, sendo usualmente composto pelas seguintes etapas: escolha dos dados de entrada e saída, seleção da estrutura, estimação dos parâmetros e validação.

### 2.3 Modelos NARX para o caso MIMO

A representação NARX MIMO para um sistema com  $q$  entradas e  $s$  saídas ( $q \times s$ ) é definida pela Equação (2.3)

$$\begin{aligned} \mathbf{y}(k) = & \mathbf{F}[\mathbf{y}(k-1), \mathbf{y}(k-2), \dots, \mathbf{y}(k-n_y), \\ & \mathbf{u}(k-d), \mathbf{u}(k-d), \dots, \mathbf{u}(k-d-n_u)] + \mathbf{e}(k) \end{aligned} \quad (2.3)$$

em que

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ \vdots \\ y_s(k) \end{bmatrix}, \mathbf{u}(k) = \begin{bmatrix} u_1(k) \\ \vdots \\ u_q(k) \end{bmatrix}, \mathbf{e}(k) = \begin{bmatrix} \xi_1(k) \\ \vdots \\ \xi_s(k) \end{bmatrix} \quad (2.4)$$

$\mathbf{F} = [F_1, \dots, F_s]^T$  são funções não lineares,  $n_y$  e  $n_u$  são os atrasos máximos da saída e entrada respectivamente,  $d \geq 1$  é o atraso puro e  $\xi_t(k)$  é um vetor de ruído com a mesma dimensão de  $y_t$  associado com a saída  $t$ .

A estrutura MIMO pode ser dividida em  $s$  modelos, cada um representando uma saída. Neste trabalho, assim como no caso SISO, a função  $\mathbf{F}$  é uma função polinomial linear nos parâmetros e a estrutura MIMO pode ser representada pela Equação (2.5),

$$\begin{bmatrix} y_1(k) \\ \vdots \\ y_s(k) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m_1} \psi_{1,i}(k)\theta_{1,i} + \xi_1(k) \\ \vdots \\ \sum_{i=1}^{m_s} \psi_{s,i}(k)\theta_{s,i} + \xi_s(k) \end{bmatrix} \quad (2.5)$$

em que  $\psi_{s,i}(k)$  é a  $s,i$ -ésima combinação linear ou não linear de todas as saídas  $[y_1, \dots, y_s]$  e entradas  $[u_1, \dots, u_q]$  com grau de não linearidade até  $l$ ;  $[m_1, \dots, m_s]$  são os números de monômios distintos nos respectivos modelos;  $[\xi_1, \dots, \xi_s]$  são sequências de ruído e  $[\theta_{1,i}, \dots, \theta_{s,i}]$  são o  $i$ -ésimo parâmetro a ser estimado no  $s$ -ésimo modelo.

A próxima seção apresenta o Estimador de Mínimos Quadrados Ortogonais e a Taxa de Redução do Erro (ERR), introduzidos por Korenberg *et al.*, (1988), no primeiro momento para o caso SISO e posteriormente é apresentada sua extensão para o caso MIMO. Tais algoritmos têm sido empregados em diversas aplicações e consistem no núcleo da identificação de sistemas NARX. Justifica ainda a apresentação de tais conceitos o fato de que a proposta utiliza os estimadores ortogonais para determinar os parâmetros.

## 2.4 Estimador de Mínimos Quadrados Ortogonais e Taxa de Redução do Erro (ERR)

O estimador de mínimos quadrados ortogonais (MQO) e a taxa de redução do erro (ERR) foram desenvolvidos para atuarem conjuntamente na obtenção de modelos NARX/NARMAX (KORENBERG *et al.*, 1988). Por meio do MQO é possível estimar modelos lineares nos parâmetros, como a representação NARX polinomial.

O MQO compõe o núcleo de um algoritmo de estimação ortogonal muito eficiente para estimação de parâmetros em modelos não lineares.

Na estimação de parâmetros em que a matriz de regressores possui muitas colunas, pode acontecer de a matriz pseudoinversa se tornar mal condicionada, ocasionando em problemas numéricos. O estimador MQO resolve esse problema ortogonalizando as colunas sobre os

dados e estimando os parâmetros. A ideia básica nesse algoritmo é definir um modelo auxiliar em que os termos são ortogonais em relação aos dados de identificação, o que permite que cada coeficiente do modelo seja estimado independentemente de outros termos, mesmo na presença de ruídos correlacionados. Uma vez estimado o modelo auxiliar, este é mapeado de volta aos dados originais para formar o modelo original.

De forma complementar ao MQO, tem-se a taxa de redução do erro que consiste em um método que permite quantificar a importância de cada regressor no modelo. A ERR leva em consideração a variância da saída e o modelo proposto para estabelecer um coeficiente que quantifica o quanto da variância pode ser explicada pelo atual modelo. Assim o algoritmo inicia quantificando quanto cada regressor contribui para explicar a variância da saída, e vai incluindo a cada passo aqueles que mais contribuem, até que um valor previamente estabelecido pelo usuário seja satisfeito. As próximas subseções detalham melhor estas duas ferramentas.

#### 2.4.1 Estimador de Mínimos Quadrados Ortogonais

Considere o seguinte modelo linear nos parâmetros

$$y(k) = \sum_{i=1}^m \psi_i(k)\theta_i + \xi(k) \quad (2.6)$$

em que  $y(k)$  são os valores da saída do sistema nos instantes  $k = 1, 2, 3, \dots, N$ ,  $\psi_i(k)$  são os regressores do modelo, isto é, uma combinação predeterminada das variáveis de entrada e saída do sistema a ser modelado, tais como  $y(k-1), y(k-2), u(k-1), u^2(k-2), \dots, etc$ ,  $\theta_i$  com  $i = 1, 2, 3, \dots, m$  são os parâmetros do modelo a ser estimado, e finalmente  $\xi(k)$  é uma sequência de ruído ou erro do sistema.

Para estimar os parâmetros  $\theta_i$  é proposto o seguinte modelo auxiliar

$$y(k) = \sum_{i=1}^m \hat{g}_i w_i(k) + \xi(k) \quad (2.7)$$

em que  $w_i(k), i = 1, 2, \dots, m$  são vetores construídos ortogonalmente sobre os dados

e  $N$  é o número de amostras, tal que

$$\sum_{k=1}^N w_i(k)w_j(k) = \begin{cases} \sum_{k=1}^N w_i^2(k) \neq 0, & i = j \\ 0, & i \neq j \end{cases} \quad (2.8)$$

O processo de ortogonalização pode ser sumarizado da seguinte forma

$$\begin{cases} w_1(k) = \psi_1(k) \\ w_2(k) = \psi_2(k) - a_{1,2}w_1(k) \\ w_3(k) = \psi_3(k) - a_{1,3}w_1(k) - a_{2,3}w_2(k) \\ \vdots \\ w_h(k) = \psi_h(k) - \sum_{r=1}^{h-1} a_{r,h}w_r, h = 2, 3, \dots, m \end{cases} \quad (2.9)$$

em que

$$a_{r,h} = \frac{\sum_{k=1}^N \psi_h(k)w_r(k)}{\sum_{k=1}^N w_r^2(k)}, 1 \leq r \leq h-1 \quad (2.10)$$

Outros processos para ortogonalização podem ser utilizados de forma equivalente às Equações (2.8) a (2.10) como o método clássico de Gram-Schmidt, método modificado de Gram-Schmidt ou método de Golub-Householder (AGUIRRE, 2015).

De acordo com Korenberg e Billings (1988) o valor de  $\hat{g}_i$  é obtido considerando a Equação (2.7) e a condição de ortogonalidade dos vetores  $w_i$ , pela seguinte Equação

$$\hat{g}_i = \frac{\sum_{k=1}^N y(k)w_i(k)}{\sum_{k=1}^N w_i^2(k)}, i = 1, 2, \dots, m \quad (2.11)$$

De acordo com Billings (2013), os parâmetros  $\theta_1, \theta_2, \dots, \theta_m$  podem ser relacionados

aos parâmetros  $\hat{g}_1, \hat{g}_2, \dots, \hat{g}_m$  como

$$\begin{cases} \theta_m &= \hat{g}_m \\ \theta_{m-1} &= \hat{g}_{m-1} - a_{m-1,m}\theta_m \\ \theta_{m-2} &= \hat{g}_{m-2} - a_{m-2,m-1}\theta_{m-1} - a_{m-2,m}\theta_m \\ &\vdots \\ \theta_h &= \hat{g}_h - \sum_{j=h+1}^m a_{h,j}\theta_j, \quad h = m-1, m-2, \dots, 1 \end{cases} \quad (2.12)$$

A Equação (2.11) é também utilizada durante a determinação da taxa de redução de erro, a qual é tratada a seguir.

#### 2.4.2 Taxa de Redução do Erro

A taxa de redução de erro constitui em um índice que permite quantificar o quanto da variância da saída de um sistema pode ser explicada por um determinado regressor, permitindo assim que se possa determinar quais regressores devem ser utilizados na estrutura do modelo (AGUIRRE, 2015).

Considere o modelo auxiliar da Equação (2.7), que elevado ao quadrado tem-se

$$y^2(k) = \left( \sum_{i=1}^m \hat{g}_i w_i(k) + \xi(k) \right) \times \left( \sum_{i=1}^m \hat{g}_i w_i(k) + \xi(k) \right) \quad (2.13)$$

Desenvolvendo esse produto tem-se,

$$\begin{aligned} y^2(k) = \langle y, y \rangle &= \sum_{i=1}^m \hat{g}_i^2 \langle w_i, w_i \rangle + \sum_{i=1, j=1, \forall i \neq j}^m 2\hat{g}_i \hat{g}_j \langle w_i, w_j \rangle + \\ &2 \sum_{i=1}^m \hat{g}_i^2 \langle w_i, \xi \rangle + \langle \xi, \xi \rangle \end{aligned} \quad (2.14)$$

Como os vetores  $w_i$  são ortogonais entre si, o produto interno ( $\langle \cdot, \cdot \rangle$ ) de  $w_i$  e  $w_j$  é igual a zero. Assumindo que  $\xi$  são os resíduos resultantes da diferença entre a saída real  $y$  e a saída estimada pelo método dos mínimos quadrados ortogonais (MQO)  $\hat{y}$ , tem-se que os regressores  $w_i$  não estarão correlacionados com os resíduos  $\xi$ . Tais considerações fazem com que o segundo e o terceiro termo no lado direito da Equação (2.14) sejam iguais a zero,

resultando em

$$\langle y, y \rangle = \sum_{i=1}^m \hat{g}_i^2 \langle w_i, w_i \rangle + \langle \xi, \xi \rangle \quad (2.15)$$

Dividindo os dois lados da Equação (2.15) por  $\langle y, y \rangle$  tem se

$$1 = \frac{\sum_{i=1}^m \hat{g}_i^2 \langle w_i, w_i \rangle}{\langle y, y \rangle} + \frac{\langle \xi, \xi \rangle}{\langle y, y \rangle} \quad (2.16)$$

Na Equação (2.16) a primeira parcela do lado direito é definida como a taxa de redução do erro (ERR) e a segunda parcela do lado direito é definido como a relação sinal-ruído (ESR). De forma prática, não há nenhum interesse em quantificar o quanto o ruído pode explicar da variância da saída, sendo a equação mais útil dada apenas por

$$ERR_i = \frac{\sum_{i=1}^m \hat{g}_i^2 \langle w_i, w_i \rangle}{\langle y, y \rangle} \quad (2.17)$$

Observe que a Equação (2.17) requer que os parâmetros  $\hat{g}_i$  do regressor em consideração sejam calculados, isto é feito conjuntamente com a Equação (2.11), assim tem-se a estimação dos parâmetros e a seleção dos regressores feita ao mesmo tempo.

A Equação (2.2) quando expandida gera um modelo com muitos regressores, o que resulta em alguns problemas, por exemplo, a lei universal não se torna aparente devido ao modelo ser muito complicado, o desempenho preditivo do modelo é deteriorado quando o modelo é aumentado além do tamanho verdadeiro do sistema, o que pode fazer com que regimes espúrios sejam induzidos (Billings, 2013).

Para estabelecer modelos parcimoniosos, pode-se utilizar o algoritmo de mínimos quadrados ortogonais com regressão direta (*Forward Orthogonal Least Square - FROLS*) (Billings *et al.*, 1988). Este algoritmo é baseado no método OLS-ERR, o qual basicamente emprega uma técnica de ortogonalização para desacoplar a estimação de novos parâmetros dos já incluídos no modelo. Em cada passo, o algoritmo calcula o ERR para cada termo candidato, aquele com maior ERR é incluído no modelo. No próximo passo, todos os termos são ortogonalizados com base no termo escolhido no passo anterior, o ERR é calculado para cada um, e o maior

é inserido no modelo. Estes passos são repetidos até que se atinja um limiar previamente estabelecido pelo algoritmo. Na próxima seção o algoritmo FROLS é detalhado.

## 2.5 Algoritmo ortogonal com regressão direta (OFR)

O algoritmo ortogonal com regressão direta OFR (do inglês *Orthogonal Forward Regression*) é também conhecido como FROLS (*Forward Orthogonal Least Square*) introduzido por Billings *et al.* (1989) é uma forma bastante efetiva de lidar com o problema da seleção da estrutura na identificação de sistemas não lineares. Este algoritmo é capaz de considerar cada regressor em potencial um por vez, e apenas incluir no modelo aqueles mais significativos.

Para entender o algoritmo FROLS, considere o seguinte modelo linear nos parâmetros da Equação (2.6) com  $y = [y(1), \dots, y(N)]^T$ , em que  $N$  é o número de amostras da saída,  $\psi_h = [\psi_h(1), \psi_h(2), \dots, \psi_h(N)]^T$ , em que o índice  $h$  indica o  $h$ -ésimo regressor candidato com  $h = 1, 2, \dots, m$  e  $m$  é o número de termos candidatos ao modelo.

O valor de  $m$  pode se tornar muito grande, resultando em modelos sobreparametrizados, o que além de dificultar o processo de estimação dos parâmetros, geralmente resulta em dinâmicas fantasmas no modelo final (Aguirre e Billings, 1995). O algoritmo FROLS evita esse problema, selecionando apenas os regressores mais significativos.

Para formulação do algoritmo, considere  $\mathcal{R}_m = \{\psi_1, \psi_2, \dots, \psi_m\}$  um dicionário de  $m$  termos candidatos ao modelo, o problema de seleção dos regressores consiste em encontrar um subconjunto  $\mathcal{R}_{M_0} = \{\alpha_1, \alpha_2, \dots, \alpha_{M_0}\} = \{\psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_{M_0}}\}$ , sendo  $M_0$  o número de regressores do modelo com ( $M_0 \leq m$ ), de forma que  $y$  pode ser aproximado pela seguinte combinação linear de  $\alpha_1, \alpha_2, \dots, \alpha_{M_0}$ :

$$y = \theta_1 \alpha_1 + \theta_2 \alpha_2 + \dots + \theta_{M_0} \alpha_{M_0} + \xi \quad (2.18)$$

ou em forma matricial,

$$y = A\theta + \xi \quad (2.19)$$

em que  $A = [\alpha_1, \dots, \alpha_{M_0}]$  rank coluna pleno,  $\theta = [\theta_1, \dots, \theta_{M_0}]^T$  é o vetor de parâmetros a ser estimado e  $\xi$  é o erro de aproximação.

O algoritmo inicia estimando o valor de  $\hat{g}_h^{(l)}$ , em que  $l$  indica a iteração, para todos os regressores candidatos e em seguida calculando o ERR, com  $w_h = \psi_h$ , de acordo com as Equações (2.20) e (2.21)

$$\hat{g}_h^{(1)} = \frac{y^T w_h}{w_h^T w_h} \quad (2.20)$$

$$ERR^{(1)}[h] = (\hat{g}_h^{(1)})^2 \frac{w_h^T w_h}{y^T y} \quad (2.21)$$

Uma vez calculada a taxa de redução do erro de todos os regressores, seleciona-se aquele de maior valor para ser incluído no modelo, isto é,

$$l_1 = \arg \max_{1 \leq h \leq m} \{ERR^{(1)}[h]\}$$

o que equivale a dizer

$$ERR[l_1] = \max \{ERR^{(1)}[h] : 1 \leq h \leq m\}$$

Identificado o primeiro regressor mais expressivo, tem-se  $\alpha_1 = \psi_{l_1}$ ,  $w_1 = \psi_{l_1}$  e  $\mathcal{R}_1 = [\psi_{l_1}]$ .

No próximo passo, o regressor  $\psi_{l_1}$  é removido do dicionário, e os outros regressores restantes são ortogonalizados em função de  $w_1$  de acordo com a Equação (2.22).

$$w_h^{(2)} = \psi_h - \sum_{r=1}^1 \frac{\psi_h^T w_r}{w_r^T w_r} w_r, \quad \psi_h \in \mathcal{R} - \mathcal{R}_1 \quad (2.22)$$

Posteriormente, procede-se estimando os parâmetros dos regressores e calculando suas respectivas taxas de redução do erro com as Equações (2.20) e (2.21), em seguida é selecionado o regressor de maior valor em ERR e definido em  $l_2$ , e faz-se  $\alpha_2 = \psi_{l_2}$ ,  $w_2 = \psi_{l_2}$  e  $\mathcal{R}_2 = [\mathcal{R}_1 \ \psi_{l_2}]$ .

Supondo que um modelo  $\mathcal{R}_{s-1}$  com  $(s-1)$  regressores,  $\alpha_1, \alpha_2, \dots, \alpha_{s-1}$ , foram escolhidos no passo  $(s-1)$ , e os regressores restantes foram ortogonalizados em uma base  $w_1, w_2, \dots, w_{s-1}$ . No passo  $s$ , com  $h \neq l_1, h \neq l_2, \dots, h \neq l_{s-1}$ . Para  $h = 1, 2, \dots, m$ , os

próximos passos são dados pelas Equações (2.23), (2.24) e (2.25).

$$w_h^{(s)} = \psi_h - \sum_{r=1}^{s-1} \frac{\psi_h^T w_r}{w_r^T w_r} w_r, \psi_h \in \mathcal{R} - \mathcal{R}_{s-1} \quad (2.23)$$

$$\hat{g}_h^{(s)} = \frac{y^T w_h^{(s)}}{(w_h^{(s)})^T w_h^{(s)}} \quad (2.24)$$

$$ERR^{(s)}[h] = (\hat{g}_h^{(s)})^2 \left[ \frac{(w_h^{(s)})^T (w_h^{(s)})}{y^T y} \right] \quad (2.25)$$

Seleciona-se então o  $s$ -ésimo regressor com maior taxa de redução do erro como,

$$l_s = \arg \max_{1 \leq h \leq m} \{ERR^{(s)}[h]\}$$

Por fim o regressor  $\alpha_s$  é escolhido como  $\psi_{l_s}$ , e  $w_s = w_{l_s}^{(s)}$ . Prosseguindo com o algoritmo passo a passo, é possível então escolher sempre os regressores com maior taxa de redução do erro, até um limiar estabelecido pelo usuário, em que a taxa de redução do erro não apresenta uma melhora significativa do modelo. A definição desse limiar é uma etapa empírica e um tanto quanto subjetiva, dependendo muito da experiência do usuário e tem sido ponto de estudo na identificação de sistemas não lineares (Guo et al., 2014).

O algoritmo FROLS pode ser sumarizado da seguinte forma,

---

**Algoritmo 1:** Algoritmo ortogonal com regressão direta (FROLS)
 

---

1 Passo 1:  $\mathcal{R}_m = [\psi_1, \psi_2, \dots, \psi_m]$ . Para  $h = 1, 2, \dots, m$ , calcule

$$g_h^{(1)} = \frac{y^T w_h}{w_h^T w_h}$$

$$ERR^{(1)}[h] = (g_h^{(1)})^2 \frac{w_h^T w_h}{y^T y}$$

$$l_1 = \arg \max_{1 \leq h \leq m} \{ERR^{(1)}[h]\}$$

2 Faça

$$a_{11} = 1$$

$$w_1 = \psi_{l_1}$$

$$g_1 = g_{l_1}^{(1)}$$

$$err[1] = ERR^{(1)}[l_1]$$

3 Passo  $s$  ( $s \geq 2$ ): Faça  $h \neq l_1, h \neq l_2, \dots, h \neq l_{s-1}$ . Para  $h = 1, 2, \dots, m$ , calcule

$$w_h^{(s)} = \psi_h - \sum_{r=1}^{s-1} \frac{\psi_h^T w_r}{w_r^T w_r} w_r, \psi_j \in \mathcal{R} - \mathcal{R}_{h-1}$$


---

4

$$g_h^{(s)} = \frac{y^T w_h^{(s)}}{(w_h^{(s)})^T w_h^{(s)}}$$

$$ERR^{(s)}[h] = (g_h^{(s)})^2 \frac{(w_h^{(s)})^T w_h^{(s)}}{y^T y}$$

$$l_s = \arg \max_{1 \leq h \leq m} \{ERR^{(s)}[h]\}$$

Faça

$$w_s = w_{l_s}^{(s)}$$

$$g_s = g_{l_s}^{(s)}$$

$$a_{r,s} = \frac{w_r^T \psi_{l_s}}{w_r^T w_r}, r = 1, 2, \dots, s-1$$

$$a_{ss} = 1$$

$$err[s] = ERR^{(s)}[l_s]$$

O algoritmo é finalizado no passo  $M_0$ , quando uma determinada precisão é alcançada, geralmente dada por,

$$1 - \sum_{s=1}^{M_0} err(s) \leq \delta$$

em que  $\delta$  é um valor bem pequeno (e.g.  $\delta \leq 10^{-2}$ ).

O modelo final é então uma combinação linear dos  $M_0$  regressores selecionados como mais expressivos no algoritmo 1, de forma que

$$y(k) = \sum_{i=1}^{M_0} g_i w_i(k) + e(k) \quad (2.26)$$

na forma ortogonalizada, ou equivalentemente na base original como

$$y(k) = \sum_{h=1}^{M_0} \beta_{l_h} \psi_{l_h}(k) + e(k) \quad (2.27)$$

Os valores de  $\beta = [\beta_{l_1}, \beta_{l_2}, \dots, \beta_{l_{M_0}}]^T$  podem ser calculado como  $A\beta = g$ , em que

$g = [g_1, g_2, \dots, g_{l_{M_0}}]^T$  e

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1M_0} \\ 0 & 1 & \cdots & a_{2M_0} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & a_{M_0-1, M_0} \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (2.28)$$

Para selecionar os monômios que farão parte da estrutura dos modelos no caso MIMO, é utilizada a taxa de redução do erro em sua forma MIMO, juntamente com o algoritmo FROLS considerando cada saída  $s$  independentemente como já descrito previamente, assim o algoritmo (1) é repetida  $s$  vezes, um para cada saída.

O ERR para o  $t$ -ésimo modelo do caso MIMO é definido como,

$$ERR_{t,i} = \frac{g_{t,i}^2 \langle w_{t,i}, w_{t,i} \rangle}{\langle Y_t, Y_t \rangle} \times 100 \quad (2.29)$$

em que  $\langle \cdot, \cdot \rangle$  é o produto interno,  $w_{t,i}$  é obtido a partir da ortogonalização de  $\psi_{t,i}$  usando o método de ortogonalização de Gram-Schmidt (Ghaboussi e Wu, 2016),  $Y_t$  é um vetor da  $t$ -ésima saída e  $g_{t,i}$  é o  $t,i$ -ésimo parâmetro auxiliar que pode ser calculado diretamente a partir de  $w_{t,i}$  e  $Y_t$  usando

$$\hat{g}_{t,i} = \frac{\langle w_{t,i}^T, y_t \rangle}{\langle w_{t,i}^T, w_{t,i} \rangle}, i \in [1, m_t] \quad (2.30)$$

em que  $m_t$  é o número total de possíveis termos.

Nas Equações (2.29) e (2.30),  $t$  é o  $t$ -ésimo modelo associado à saída  $t$  e pertence ao intervalo  $[1, \dots, s]$ .

Embora o ERR seja utilizado para estabelecer quais regressores são importantes para o modelo, algumas vezes regressores que não são significativos podem ser selecionados, nestes casos o teste-t de Student pode ser utilizado para verificar se o parâmetro estimado relacionado com determinado regressor é estatisticamente significativo (Avellina *et al.*, 2017; Montgomery, 2013; Wei e Billings, 2008). A próxima seção traz detalhes do teste-t de Student.

## 2.6 Teste-t de Student

Ao se identificar um sistema usando o algoritmo FROLS, pode ocorrer de alguns regressores serem incluídos no modelo, mas apresentarem um parâmetro não significativo, isto é, muito próximo de zero, nestes casos o teste-t de Student é usado para verificar se tais parâmetros são estatisticamente significativos.

Sendo a Equação (2.2) linear nos parâmetros, a mesma pode ser escrita como,

$$Y = \Psi^T \Theta + \Xi \quad (2.31)$$

em que  $Y$  é o vetor de saída de dimensão  $N \times 1$ ,  $\Psi$  é a matriz de regressores de dimensão  $\tau \times N$ ,  $\Theta$  é o vetor de parâmetros  $\tau \times 1$  e  $\Xi$  é um vetor de ruído de dimensão  $N \times 1$ .

O vetor de parâmetros  $\Theta$  pode ser estimado utilizando um algoritmo de mínimos quadrados, resultando em

$$\hat{\Theta} = (\Psi \Psi^T)^{-1} \Psi Y \quad (2.32)$$

A variância do  $j$ -ésimo regressor pode ser calculada como (Avellina *et al.*, 2017),

$$\hat{\sigma}_j^2 = \hat{\sigma}_e^2 V_{jj} \quad (2.33)$$

em que  $V_{jj}$  é o  $j$ -ésimo elemento da diagonal de  $V = (\Psi \Psi^T)^{-1}$  e  $\hat{\sigma}_e^2$  é dado por

$$\hat{\sigma}_e^2 = \frac{1}{N - \tau} \|Y - \Psi \hat{\Theta}\|^2 \quad (2.34)$$

Considerando  $t_{\alpha, N-\tau}$  como o  $100(1 - \alpha)$ th percentil da distribuição t de Student com  $N - \tau$  graus de liberdade, o seguinte intervalo de confiança é utilizado para determinar a relevância estatística de cada regressor.

$$\left[ \hat{\Theta}_j - \hat{\sigma}_j t_{\alpha, N-\tau}, \hat{\Theta}_j + \hat{\sigma}_j t_{\alpha, N-\tau} \right] \quad (2.35)$$

Se zero pertence ao intervalo em (2.35), então  $\hat{\Theta}_j$  não é estatisticamente diferente de zero, e o regressor pode ser descartado.

Para o caso MIMO, o t-ésimo modelo na equação linear nos parâmetros pode ser escrita como,

$$Y_t = \Psi_t^T \Theta_t + \Xi_t \quad (2.36)$$

em que

$$Y_t = \begin{bmatrix} y_t(1) \\ \vdots \\ y_t(N) \end{bmatrix}, \Psi_t = \begin{bmatrix} \psi_{t,1}(1) & \dots & \psi_{t,1}(N) \\ \vdots & \ddots & \vdots \\ \psi_{t,\tau_t}(1) & \dots & \psi_{t,\tau_t}(N) \end{bmatrix} \quad (2.37)$$

$$\Theta_t = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{\tau_t} \end{bmatrix}, \Xi_t = \begin{bmatrix} \xi_t(1) \\ \vdots \\ \xi_t(N) \end{bmatrix}$$

e  $Y_t$  é o t-ésimo vetor de saída de dimensão  $N \times 1$ ,  $\Psi_t$  é a t-ésima matriz de regressores de dimensão  $\tau_t \times N$ ,  $\Theta_t$  é o t-ésimo vetor de parâmetros de dimensão  $\tau_t \times 1$  e  $\Xi_t$  é o t-ésimo vetor de ruído de dimensão  $N \times 1$ .

O parâmetro  $\Theta_t$  pode ser estimado usando um algoritmo de mínimos quadrados de acordo com a Equação (2.38).

$$\hat{\Theta}_t = (\Psi_t \Psi_t^T)^{-1} \Psi_t Y_t \quad (2.38)$$

A variância do t,j-ésimo regressor do vetor de parâmetros  $\hat{\Theta}_t$  pode ser calculada como,

$$\hat{\sigma}_{tj}^2 = \hat{\sigma}_{et}^2 V_{t,jj} \quad (2.39)$$

em que  $V_{t,jj}$  é o j-ésimo elemento da diagonal de  $V_t = (\Psi_t \Psi_t^T)^{-1}$  e  $\hat{\sigma}_{et}^2$  é dado por

$$\hat{\sigma}_{et}^2 = \frac{1}{N - \tau_t} \|Y_t - \Psi_t \hat{\Theta}_t\|^2 \quad (2.40)$$

em que  $N$  é o número de amostras e  $\tau_t$  é o número de regressores no vetor  $\Psi_t$ .

Para cada regressor, o  $T_{\alpha, N - \tau_t}$  é o  $100(1 - \alpha)$ th percentil da distribuição t de Student com  $N - \tau_t$  graus de liberdade, e o intervalo de confiança é dado por,

$$\left[ \hat{\Theta}_{t,j} - \hat{\sigma}_{t,j} T_{\alpha, N - \tau_t}, \hat{\Theta}_{t,j} + \hat{\sigma}_{t,j} T_{\alpha, N - \tau_t} \right] \quad (2.41)$$

Se zero pertence ao intervalo,  $\hat{\Theta}_{t,j}$  não é significativo, e o regressor é rejeitado.

O teste t de Student é um passo importante para a implementação dos algoritmos randomizados (Falsone *et al.*, 2015) que será tratado no capítulo 3.

Nas discussões que se seguem, os modelos MIMO são divididos em  $s$  modelos de múltiplas entradas e uma saída (*Multiple Input Single Output - MISO*) na forma da Equação (2.42).

$$y_t(k) = \sum_{i=1}^{m_t} \psi_{t,i}(k)\theta_{t,i} + \xi_t(k) \quad (2.42)$$

Note que na Equação (2.42) todas as saídas  $[y_1, \dots, y_s]$  e todas as entradas  $[u_1, \dots, u_q]$  são parte do conjunto de regressores candidatos  $\mathcal{R}_m$ .

## 2.7 Desempenho do Modelo

O Erro Quadrático Médio de Predição (*Mean Squared Prediction Error - MSPE*) e o Critério de informação Bayesiano (BIC) são métodos mais populares para medir o desempenho de um modelo identificado.

O método MSPE é usado para atualizar a probabilidade de inclusão dos regressores. Isto é feito calculando o MSPE para cada modelo na população atual, então um índice é calculado para cada regressor, este índice é usado para aumentar ou diminuir a probabilidade de inclusão do respectivo regressor com base na sua presença ou ausência no modelo. Este processo é detalhado na sessão 3.1.2.

A seleção da estrutura é um problema multiobjetivo onde busca-se encontrar um modelo parcimonioso sem termos espúrios a fim de prevenir uma sobreparametrização e garantir o menor erro de adequação do modelo aos dados (Hafiz *et al.*, 2019b). O critério BIC contrapõe a precisão do modelo com o seu tamanho, neste trabalho tal critério é utilizado para medir o desempenho dos melhores modelos dentro da população atual. Este critério é utilizado nos algoritmos propostos para selecionar quais modelos participarão dos processos de herança genética.

### 2.7.1 Erro Quadrático Médio

O índice MSPE para o caso SISO é dado por

$$MSPE = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}(k))^2 \quad (2.43)$$

em que  $\hat{y}(k)$  denota a predição de um passo a frente de  $y(k)$ .

Para predições com um amplo horizonte de tempo, o Erro Quadrático Médio de Simulação (*Mean Squared Simulation Error - MSSE*) é mais apropriado do que o MSPE e é dado por

$$MSSE = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}_{sim}(k))^2 \quad (2.44)$$

em que  $\hat{y}_{sim}(k)$  denota a saída simulada do modelo. Embora a Equação (2.44) seja mais confiável e robusta para alguns problemas típicos de identificação de sistema como dados obtidos com entradas de baixa frequência ou dados sobre-amostrados (Piroddi e Spinelli, 2003), o cálculo desse índice demanda um alto esforço computacional.

Em identificação de sistemas, o sinal de entrada tem um papel muito importante, já que este dita as frequências e amplitudes presente no sinal de saída (Ljung e Soderstrom, 1983).

Algumas vezes na literatura atual um sinal pouco excitativo é usado na identificação de sistemas com o objetivo de medir o desempenho do método de identificação (Piroddi e Spinelli, 2003; Wei e Billings, 2008; Aguirre, 2010). No entanto, como enfatizado em Hafiz *et al.* (2019a), tais sinais não devem ser usados na identificação de sistemas e são usados apenas como uma representação do pior cenário. Na prática, sinais de entrada para identificação de sistemas devem ser persistentemente excitantes.

De forma semelhante ao caso SISO, o índice MSPE para o caso MIMO pode ser calculado pelas Equações (2.45) e (2.46).

$$MSPE_t = \frac{1}{N} \sum_{k=1}^N (y_t(k) - \hat{y}_t(k))^2 \quad (2.45)$$

$$MSSE_t = \frac{1}{N} \sum_{k=1}^N (y_t(k) - \hat{y}_{tsim}(k))^2 \quad (2.46)$$

em que  $\hat{y}_t(k)$  denota a predição de um passo a frente de  $y_t(k)$  e  $\hat{y}_{tsim}(k)$  denota a saída simulada, todas relacionadas ao t-ésimo modelo.

Uma variante das Equações (2.45) e (2.46) são as versões exponenciais, as quais são usadas para amplificar a diferença entre modelos com desempenho similares. As variantes são dadas pelas Equações (2.47) e (2.48).

$$\mathcal{J}_{tpre} = e^{-K \cdot MSPE_t}, \quad (2.47)$$

$$\mathcal{J}_{tsim} = e^{-K \cdot MSSE_t}, \quad (2.48)$$

em que  $K$  é um parâmetro de sintonia.

As Equações (2.47) e (2.48) podem ser combinadas em um único índice (Aguirre *et al.*, 2010), dado por,

$$\mathcal{J} = \beta \mathcal{J}_{tsim} + (1 - \beta) \mathcal{J}_{tpre} \quad (2.49)$$

em que  $\beta \in [0, 1]$ .

A versão exponencial do erro quadrático médio de predição é usada neste trabalho para determinar a importância dos regressores em detrimento do erro quadrático médio de simulação por exigir um menor custo computacional. A importância dos regressores é discutida na seção 3.1.2.

### 2.7.2 Critério da Informação Bayesiana

Um critério de seleção de modelos para sistemas não lineares é dificilmente aplicável, já que nenhuma relação simples entre tamanho do modelo e precisão podem ser estabelecidos (Falsone *et al.*, 2015). Alguns trabalhos utilizam BIC como critério para determinar o tamanho de modelos ou uma função critério para selecionar o melhor modelo (Akinola *et al.*, 2019; Wei e Billings *et al.*, 2008; Hafiz *et al.*, 2019a; 2019b).

A função BIC pode ser usada para selecionar o melhor modelo da população atual,

e utilizá-lo em um estágio posterior. O tamanho do modelo é determinado pela convergência dos modelos na população, os quais ao longo das iterações, convergem para um tamanho comum.

O critério BIC para os sistemas SISO é dado pela Equação (2.50).

$$BIC = N \times \ln \left( \frac{1}{N} \times MSPE \right) + \tau \times \ln(N) \quad (2.50)$$

em que  $N$  é o número de amostras nos dados e  $\tau$  é o número de regressores no modelo considerado.

Já para o caso MIMO, o BIC para o t-ésimo modelo é dado pela Equação (2.51).

$$BIC_t = N \times \ln \left( \frac{1}{N} \times MSPE_t \right) + \tau_t \times \ln(N) \quad (2.51)$$

em que  $N$  é o número de amostras nos dados e  $\tau_t$  é o número de regressores no t-ésimo modelo.

## 2.8 Herança Genética

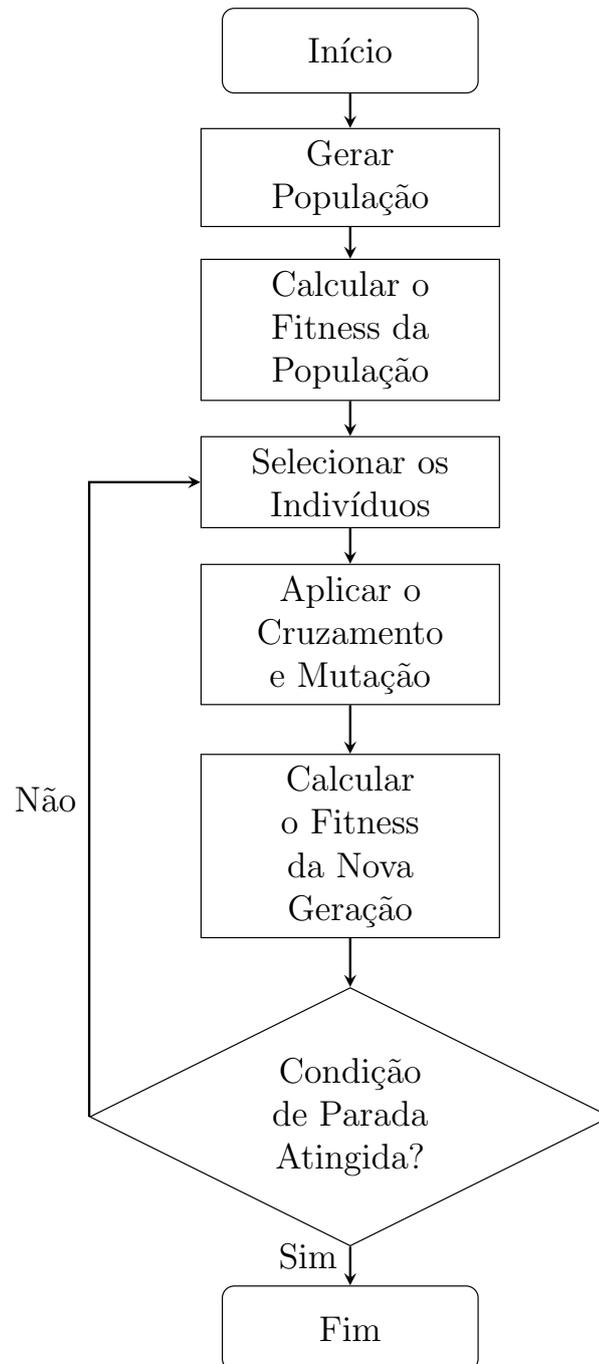
O termo herança genética utilizado no presente trabalho teve inspiração e se assemelha muito aos algoritmos genéticos que foram introduzidos por David E. Goldberg e John H. Holland em seu trabalho intitulado *Genetic Algorithms and Machine Learning*, (Goldberg e Holland, 1988). Desde então os algoritmos genéticos foram aplicados nas mais diversas áreas.

Um algoritmo genético é um método que busca soluções aproximadas dentro de um espaço factível. Pode ser aplicado tanto em problemas combinacionais ou de otimização.

Basicamente os algoritmos genéticos possuem as seguintes etapas: Geração da população inicial, cálculo do fitness de cada indivíduo, seleção dos indivíduos, cruzamento e mutação. O fluxograma na Figura (2.1) resume esse processo.

Usualmente os algoritmos genéticos iniciam gerando uma população de soluções candidatas respeitando uma distribuição normal. Essa população é criada com base no espaço factível de forma aleatória. A fim de avaliar os indivíduos, uma função quantifica o quão satisfatória essa solução é para o problema proposto, tal função é usualmente denominada

Figura 2.1: Fluxograma Geral de um Algoritmo Genético.



*Fitness.*

De posse do Fitness de cada indivíduo, o algoritmo realiza uma seleção de quais indivíduos farão parte do processo de cruzamento, no presente trabalho não é realizado o cruzamento, mas sim uma herança genética, isto é, parte do DNA dos pais é transmitido diretamente aos filhos.

Na literatura existem várias formas de selecionar os indivíduos tais como torneio, roleta, seleção por truncamento, seleção por ranqueamento linear e seleção por ranqueamento exponencial. A mais simples e comum é o método por roleta.

O processo de herança genética é o responsável por produzir novos indivíduos a partir dos previamente selecionados na etapa da roleta. Basicamente essa etapa parte do pressuposto que características dos pais que apresentaram melhor desempenho podem ser passadas aos filhos de forma que estes possam vir a ser melhores do que os pais.

A etapa de mutação é um processo de baixa probabilidade que é responsável por aumentar a variabilidade genética da população evitando que mínimos locais venham a ser atingidos, no presente trabalho esta etapa não foi implementada.

É importante destacar que os algoritmos genéticos são métodos de busca aproximada, não podendo garantir com toda certeza que a solução encontrada é a melhor.

Por fim, as condições de validação dos algoritmos genéticos variam muito entre trabalhos e geralmente é baseada em um índice que mede a qualidade das soluções encontradas. No presente trabalho a combinação da herança genética aos algoritmos aleatórios permitiu uma melhora considerável de desempenho, a implementação desses processos ficará mais clara posteriormente quando tratarmos dos algoritmos propostos no Capítulo 3.

## **2.9 Considerações Finais**

O presente capítulo apresentou o estado da arte da identificação de sistemas SISO e MIMO usando representação NARX, apontando que embora tenha havido avanços o problema da seleção da estrutura permanece em aberto, sendo explorado no presente trabalho uma proposta de técnica que vise a melhorias nesta etapa como maior parcimônia na seleção das estruturas, maior taxa de assertividade na seleção da estrutura bem como menor tempo de

convergência. Foi apresentado também os fundamentos básicos para o desenvolvimento da proposta tais como modelos NARX, estimadores ortogonais e taxa de redução do erro, teste-t de Student, algoritmos genéticos e algoritmo da Seleção Aleatória da Estrutura de Modelos. Os dois próximos capítulos irão detalhar os algoritmos que compõem o método proposto para sistemas SISO e MIMO.



# CAPÍTULO III

## IDENTIFICAÇÃO DE SISTEMAS SISO

O presente capítulo se inicia por apresentar a formulação do problema de seleção da estrutura de modelos em termos de distribuição de probabilidade, a qual é mais adequada para a compreensão da Seleção Aleatória da Estrutura de Modelos. Em seguida segue-se a discussão do algoritmo proposto para a identificação de Sistemas SISO.

### 3.1 Formulação do Problema e a Seleção Aleatória da Estrutura de Modelos SISO

#### 3.1.1 Formulação do Problema de Seleção da Estrutura de Modelos SISO

Seja um conjunto de dados incluindo a saída  $y(k)$ , contendo ruído ou não, e a entrada  $u(k)$ , e considerando  $\mathcal{F} = 2^m$  como o conjunto de todas as possibilidades de estruturas de modelos, para um conjunto específico de  $m$  regressores  $\mathcal{R}_m = \{\psi_1, \dots, \psi_m\}$  no qual os regressores são formados pela combinação das entradas passadas de  $y(k)$  e  $u(k)$ , o problema da seleção da estrutura de modelos consiste em encontrar o melhor modelo  $f^*$ , no qual o desempenho do modelo  $\mathcal{J}$  assume o valor máximo, isto é,

$$f^* = \arg \max_{f \in \mathcal{F}} \mathcal{J}(f) \quad (3.1)$$

em que  $\mathcal{J}$  é o desempenho do modelo discutido na seção 2.7.

Neste trabalho supõe-se que existe apenas um único melhor modelo, o qual coincide

com o modelo verdadeiro, o que também implica dizer que o grau de não linearidade do conjunto  $\mathcal{R}_m$  tem no mínimo o mesmo grau de não linearidade do modelo verdadeiro.

Uma solução para o problema na Equação (3.1) por força bruta é intratável, desde que o tamanho do conjunto  $\mathcal{F}$  cresce extremamente rápido com o aumento dos atrasos  $n_u, n_y$  e o grau de não linearidade  $l$ . Assim sendo, uma formulação para o problema da seleção da estrutura pode ser feita em termos de distribuição de probabilidade.

Denote  $\phi$  como uma variável aleatória discreta assumindo valores em  $\mathcal{F}$ , de acordo com a distribuição de probabilidade  $\mathcal{P}_\phi$ . O valor esperado do desempenho de  $\phi$  pode ser dado por uma combinação convexa do índice de desempenho dos modelos ponderado por suas probabilidades como mostra a Equação (3.2).

$$\mathbb{E}[\mathcal{J}(\phi)] = \sum_{f \in \mathcal{F}} \mathcal{J}(f) \mathcal{P}_\phi(f) \quad (3.2)$$

Além disso, o problema de seleção da estrutura do modelo pode ser reformulado em encontrar a distribuição de probabilidade  $\mathcal{P}_\phi^*$  que maximiza a Equação (3.2), como na Equação (3.3).

$$\mathcal{P}_\phi^* = \arg \max_{\mathcal{P}_\phi} \mathbb{E}[\mathcal{J}(\phi)] \quad (3.3)$$

### 3.1.2 Seleção Aleatória da Estrutura de Modelos

Para resolver o problema de maximização da Equação (3.3), Falsone *et al.* (2015) apresentaram um algoritmo para Seleção Aleatória da Estrutura de Modelos (*Randomized Model Structure Selection - RaMSS*). Um procedimento baseado em uma população de modelos, na qual informações são extraídas a partir da presença ou ausência dos regressores, e usadas iterativamente para atualizar uma distribuição de probabilidade  $\mathcal{P}_\phi$  até que  $\mathcal{P}_\phi^* = 1$ , a qual é referenciada neste trabalho como distribuição limite.

Falsone *et al.* (2015) mostraram que uma avaliação mais confiável e robusta da importância dos regressores pode ser obtida considerando um índice derivado de uma coleção de modelos e não de um único modelo, este índice é usado no algoritmo RaMSS.

Considere um conjunto  $\mathcal{R}_m = \{\psi_1, \dots, \psi_m\}$  de  $m$  regressores candidatos. Para cada

regressor  $\psi_j$  pertencente ao conjunto  $\mathcal{R}_m$ , um índice  $\mathcal{I}_j^p$  é calculado sobre um conjunto de modelos candidatos pelas Equações (3.4) a (3.6).

$$\mathcal{I}_j^p = \mathcal{I}_j^{p+} - \mathcal{I}_j^{p-} \quad (3.4)$$

$$\mathcal{I}_j^{p+} = \frac{1}{n_j^+} \sum_{i=1}^{n_j^+} \mathcal{J}_i^+ \quad (3.5)$$

$$\mathcal{I}_j^{p-} = \frac{1}{n_j^-} \sum_{i=1}^{n_j^-} \mathcal{J}_i^- \quad (3.6)$$

em que  $\mathcal{I}_j^{p+}$  é o desempenho médio dos modelos que contêm o j-ésimo regressor,  $n_j^+$  é o número de modelos que contêm o j-ésimo regressor e  $\mathcal{J}_i^+$  é o desempenho dos modelos que contêm o j-ésimo regressor calculado usando MSPE. Na Equação (3.6), analogamente à Equação (3.5),  $\mathcal{I}_j^{p-}$ ,  $n_j^-$  e  $\mathcal{J}_i^-$  se refere ao desempenho média, número de modelos, e desempenho dos modelos que não contêm o j-ésimo regressor, respectivamente.

Como o índice  $\mathcal{I}_j^p$  avalia a importância global do regressor sobre uma população de modelos, a classificação do regressor é mais robusta e confiável em relação àquelas que avaliam apenas um modelo. Como demonstrado em Falsone *et al.* (2015), quando  $\mathcal{P}_\phi^*$  é suficientemente alto, este índice tem valores positivos apenas para os regressores verdadeiros no modelo. Esta condição pode ser atingida por meio de uma sintonia iterativa utilizando o algoritmo RaMSS, mesmo quando  $\mathcal{P}_\phi^* = 0$ , isto é, o método proposto é capaz de extrair informações sobre a estrutura correta processando modelos parciais ou incompletos até que  $\mathcal{P}_\phi^*$  atinja a distribuição limite. Além disso, se o índice  $\mathcal{I}_j^p$  é calculado usando o índice de simulação  $\mathcal{J}_{sim}$ , resultados melhores podem ser obtidos ao custo de um maior esforço computacional.

A distribuição de probabilidade  $\mathcal{P}_\phi$  na Equação (3.2) é parametrizada associando uma variável aleatória de Bernoulli  $\rho_j \sim Be(\mu_j)$ , com  $\mu_j \in ]0, 1[$ , para cada regressor  $\psi_j$ ,  $j = 1, \dots, m$ . A variável  $\mu_j$ , a qual simboliza a probabilidade de inclusão do regressor (*Regressor Inclusion Probability - RIP*), é usada para estabelecer se o regressor será incluído no modelo ou não.

Assumindo que todas as probabilidades  $\rho_j$ ,  $j = 1, \dots, m$  são independentes, a proba-

bilidade de escolher um modelo específico é o produto da RIP de cada regressor que pertence ao modelo, vezes o produto do complemento da RIP de cada regressor que não pertence ao modelo, isto é

$$\mathcal{P}_\phi(f) = \prod_{j:\psi_j \in f} \mu_j \prod_{j:\psi_j \notin f} (1 - \mu_j) \quad (3.7)$$

Baseado na Equação (3.7), uma solução para o problema na Equação (3.3) consiste em atualizar a distribuição de probabilidade  $\mathcal{P}_\phi$ , de forma que esta se concentre no modelo “verdadeiro”.

Assim, o algoritmo RaMSS inicia gerando uma população de modelos de tamanho  $M$ , e a inclusão dos regressores neste modelo é definida pela variável aleatória de Bernoulli  $\rho_j$ , assim, para cada regressor de 1 até  $m$  um número aleatório é gerado, se este número é menor ou igual a  $\mu_j$  do regressor correspondente, este regressor será incluído no modelo. Este processo é repetido até que se tenham os  $M$  modelos que irão compor a população inicial.

Uma vez que a estrutura do modelo foi estabelecida, os parâmetros são calculados utilizando mínimos quadrados, e a significância estatística de cada parâmetro é testada usando um teste-t de Student, aqueles que não são significantes são eliminados e os parâmetros são reestimados.

No próximo passo, o desempenho dos modelos é calculado utilizando os índices preditivo, simulado ou até mesmo os dois (de acordo com o  $\beta$  definido pelo usuário), como na Equação (2.49).

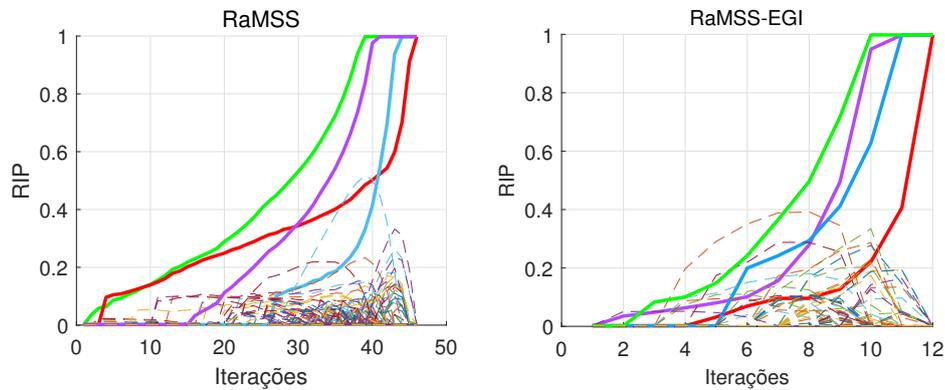
Usando o desempenho dos modelos, o índice que caracteriza a importância dos regressores é calculado utilizando as Equações (3.4) a (3.6), e depois, usados para atualizar as RIPs, Equação (3.8),

$$\mu_j(i+1) = \mu_j(i) + \gamma \mathcal{I}_j \quad (3.8)$$

em que  $\mathcal{I}_j$  é o índice definido pelas Equações (3.4) a (3.6), as quais podem assumir a forma preditiva ou simulada e  $\gamma$  é um parâmetro de sintonia. De acordo com Falsone *et al.* (2015) e Avelina *et al.* (2017), este parâmetro pode ser definido por uma solução adaptativa dada por

$$\gamma = \frac{1}{10 (\mathcal{J}_{max} - \bar{\mathcal{J}}) + 0,1} \quad (3.9)$$

Figura 3.1: Evolução da probabilidade de inclusão dos regressores (RIP), (a) RaMSS e (b) RaMSS-EGI.



em que  $\mathcal{J}_{max}$  é o desempenho do melhor modelo na população atual e  $\bar{\mathcal{J}}$  é o desempenho média dos modelos.

Na próxima iteração, o algoritmo gera as estruturas de modelos baseado na RIP atualizada, e então os mesmos passos são executados até que as RIPs converjam para a distribuição limite. A Figura (3.1) mostra o gráfico da probabilidade de inclusão dos regressores em função da iteração para os métodos RaMSS e para o método proposto RaMSS-EGI. Observa que a RIP de todos os regressores inicia com um valor baixo e, ao longo das iterações, a RIP dos regressores que pertencem ao modelo (linhas contínuas) tendem ao valor limite (próximo de 1), e a RIP dos regressores ruins (linhas pontilhadas) tende para zero.

### 3.2 Métodos propostos para o caso SISO

Duas novas abordagens são apresentadas: uma baseada na combinação da taxa de redução do erro (ERR) e dos algoritmos aleatórios (RaMSS), chamada de RaMSS-ERR, e uma baseada na combinação de herança genética e algoritmos aleatórios, chamada de RaMSS-EGI. Os dois algoritmos são similares ao RaMSS, iniciando por gerar uma população  $M$  usando uma probabilidade de inclusão de regressores (RIP) inicial, e então calculando o desempenho de cada indivíduo, desempenho este que será posteriormente utilizada para atualizar a RIP. Se o desempenho de um modelo contendo o  $j$ -ésimo regressor é, em média, melhor do que o desempenho daqueles modelos que não contêm o  $j$ -ésimo regressor, a RIP do  $j$ -ésimo regressor é aumentada, caso contrário será diminuída. Ambos os algoritmos utilizam a ERR como um

filtro para inclusão dos regressores, excluindo aqueles que possuem uma baixa contribuição para a redução da variância da saída. Cada regressor tem sua ERR calculada e comparada com um limiar  $\delta$ , se a ERR não é maior que este limiar, o regressor é excluído do modelo.

A principal diferença entre os algoritmos RaMSS-ERR e RaMSS-EGI está em como a população é gerada e também no fato de que este último utiliza uma população  $M_{best}$  a qual atua como uma memória social, mantendo um registro dos melhores modelos em cada iteração, esta população  $M_{best}$  é também utilizada para atualizar a RIP da mesma forma que a população  $M$ . Esta dupla atualização reforça a convergência e funciona de forma similar ao elitismo, técnica comumente empregada em algoritmos genéticos de forma a preservar os melhores indivíduos entre gerações.

O principal aspecto que diferencia os dois algoritmos propostos é a forma que cada um gera a população de modelos. Enquanto o RaMSS-ERR gera a população de modelo completamente baseada na distribuição de probabilidade  $\mathcal{P}_\phi$ , o método RaMSS-EGI é adaptado para um framework evolucionário, preservando para as próximas gerações de modelos, parte dos melhores modelos na iteração anterior. A metodologia usada neste trabalho é mostrada no esquema na Figura (3.2).

Na sequência serão apresentados os algoritmos para os métodos propostos no caso SISO e, posteriormente, para o caso MIMO.

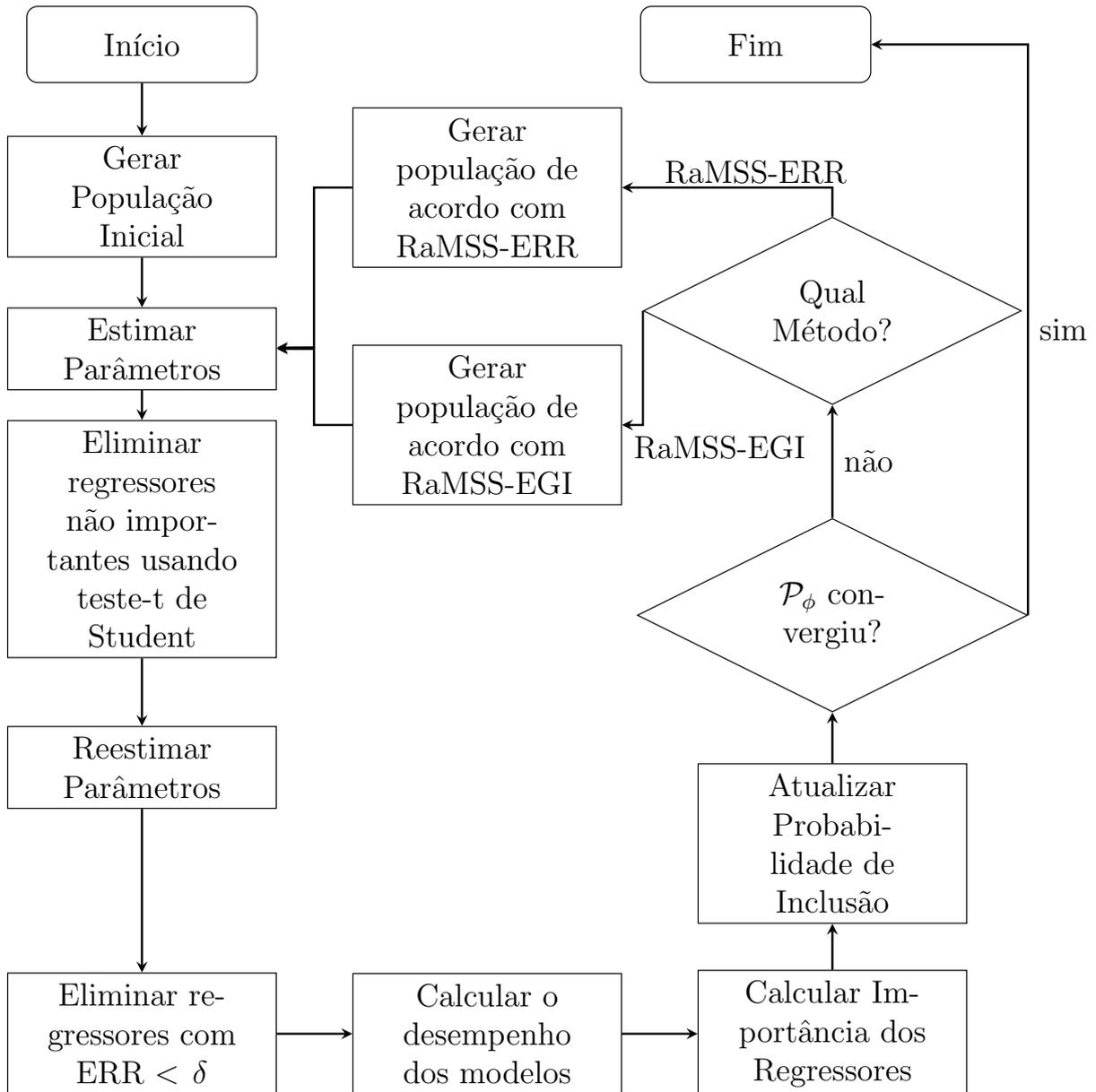
### 3.3 Algoritmo Geral para o caso SISO

Os métodos propostos usam o conceito do RaMSS de atualizar uma distribuição probabilística  $\mathcal{P}_\phi$ , com a diferença que eles são auxiliados pelo critério ERR, o qual estabelece se o regressor irá ser incluído no modelo ou não. Além disso, o método RaMSS-EGI usa duas populações de modelos, uma gerada usando a probabilidade de inclusão do regressor (RIP) combinada com herança genética, população ( $M$ ), e outra extraída dos melhores modelos nas iterações anteriores, chamada aqui de melhor população ( $M_{best}$ ).

A saída do algoritmo proposto é a probabilidade de inclusão dos regressores  $\mu$ , que parametriza a distribuição de probabilidade  $\mathcal{P}_\phi$ .

Para estabelecer o algoritmo, considere a seguinte definição.

Figura 3.2: Algoritmo Geral para RaMSS-ERR e RaMSS-EGI.



**Definição.** Cada modelo na população  $M$  é formado por um conjunto de regressores  $\Psi$  de dimensão  $\tau \times N$ , o número de regressores  $\tau$ , um vetor de parâmetros correspondendo a cada regressor  $\Theta$  de dimensão  $\tau \times 1$ , um vetor contendo a taxa de redução do erro de cada regressor  $ERR$  de dimensão  $\tau \times 1$ , o índice de desempenho  $BIC$  e o desempenho do modelo  $\mathcal{J}$ . Veja Equação (3.10).

$$M_i = [\Psi_i, \tau_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i] \quad (3.10)$$

O Algoritmo 2 sumariza o processo geral de identificação. No Algoritmo 2,  $\mathcal{R}_m$  é o conjunto de regressores candidatos de dimensão  $m \times 1$ ,  $Z_M$  é o número de indivíduos,  $K$  é o parâmetro de sintonia da Equação (2.47) e (2.48),  $\beta$  é um parâmetro de sintonia na Equação (2.49),  $\mu_i$  é a probabilidade de inclusão de regressores inicial, a qual é definida como  $\mu_i = 1/m$ ,  $\mu_{lower}$ ,  $\mu_{upper}$  e  $\delta$  são valores limites pré definidos que serão abordados posteriormente neste trabalho, e  $\kappa$  é um pequeno valor usado para indicar que todos os RIPs convergiram. Além disso, na linha 14,  $\mu$  é atualizado baseado na população  $M_{best}$ , a qual constitui uma população elitista, assim, em cada iteração se os modelos na população  $M$  são melhores que os modelos na população  $M_{best}$ , esses serão copiados para a população  $M_{best}$ . Nas próximas seções cada passo do Algoritmo 2 será detalhado.

É possível observar que o Algoritmo 2 possui como entradas os valores medidos da entrada do sistema  $u(k)$  e da saída do sistema  $y(k)$ , que pode ou não conter ruídos de medição, o conjunto de regressores candidatos  $\mathcal{R}_m$ , o tamanho da população  $Z_m$ , e os parâmetros  $K$ ,  $\beta$ ,  $\mu_i$ ,  $\mu_{lower}$ ,  $\mu_{upper}$ ,  $\delta$  e  $\kappa$ . E como saída a probabilidade de inclusão  $\mu$ .

Para os dois métodos propostos, RaMSS-ERR e RaMSS-EGI, a estimação dos parâmetros, o cálculo dos índices de desempenho  $\mathcal{J}_i$  e  $BIC_i$ , bem como a etapa de atualização de  $\mu$  para cada regressor é comum em ambos (linhas 7 a 11 do Algoritmo 2). Ao se escolher entre os dois métodos pode-se proceder a linha 13 ou a linha 17 do Algoritmo 2, escolhendo o método RaMSS-EGI é criada uma população elitista (linha 13) a qual é utilizada para atualizar a probabilidade de inclusão dos regressores  $\mu$  (linha 14), em seguida a próxima população  $M$  é gerada utilizando o Algoritmo 6. Por outro lado, quando se escolhe o método RaMSS-ERR o algoritmo simplesmente cria a população  $M$  usando o Algoritmo 3.

---

**Algoritmo 2:** Algoritmo Geral RaMSS-ERR e RaMSS-EGI no caso SISO
 

---

```

1 Input:  $\{u(k), y(k), k = 1, \dots, N\}$ ,  $\mathcal{R}_m$ ,  $Z_M$ ,  $K$ ,  $\beta$ ,  $\mu_i$ ,  $\mu_{lower}$ ,  $\mu_{upper}$ ,  $\delta$ ,  $\kappa$ ;
2 Output:  $\mu$ ;
3 Escolha o método RaMSS-ERR ou RaMSS-EGI;
4  $\mu \leftarrow \mu_i$ ;
5 Gerar população inicial  $M$  usando o Algoritmo 3 ;
6 repeat
7   for  $i \leftarrow 1$  to  $Z_M$  do
8     Estimar os parâmetros do modelo usando o Algoritmo 4;
9     Calcular os índices de desempenho  $\mathcal{J}_i$  e  $BIC_i$  usando as Equações (2.47) e
      (2.50) respectivamente;
10  end
11  Atualizar  $\mu$  para cada regressor usando a população  $M$  como entrada no
      Algoritmo 5;
12  if  $Method = RaMSS-EGI$  then
13    Selecione o melhor indivíduo na população  $M$  e o copie para a população
       $M_{best}$ ;
14    Atualizar  $\mu$  para cada regressor usando a população  $M_{best}$  como entrada
      no Algoritmo 5;
15    Gerar a próxima população  $M$  usando  $\mu$  atualizada e o Algoritmo 6;
16  else
17    Gerar a próxima população  $M$  usando  $\mu$  atualizada e o Algoritmo 3;
18 until  $\min_{j=1, \dots, m} |2\mu_j - 1| \geq 1 - \kappa$ ;

```

---

### 3.3.1 Gerar População Inicial

Como no RaMSS, a população inicial é gerada com base em uma variável aleatória de Bernoulli  $\rho_j$ . A população inicial é formada por modelos, com diferentes regressores tendo em vista que todos os regressores iniciam com a mesma probabilidade de serem incluídos no modelo. A inclusão dos regressores nos modelos é definida considerando um conjunto de regressores  $\mathcal{R}_m$  e testando cada um para inclusão. A inclusão é feita gerando um número aleatório para cada regressor candidato e o comparando com a RIP correspondente, o regressor é incluído no modelo somente se o número aleatório pertencer ao intervalo  $[0, \mu_j]$ . O Algoritmo 3 é usado para gerar a população inicial.

O processo de gerar a população no Algoritmo 3 é feito para cada modelo da população  $Z_m$ . Inicialmente todos os modelos nessa população tem os seus conjuntos de regressores esvaziados (linha 5), a seguir para cada um dos regressores candidatos de 1 até  $m$  (linha 8) é feito um teste gerando um número aleatório entre 0 e 1 e comparado com a

---

**Algoritmo 3:** Gerar População Inicial
 

---

```

1 Input:  $Z_M, \mathcal{R}_m, \mu_i$ 
2 Output:  $M$ ;
3  $\mu \leftarrow \mu_i$ ;
4 for  $i \leftarrow 1$  to  $Z_M$  do
5    $\hat{\Psi}_i \leftarrow []$ ;
6   while  $\hat{\Psi}_i = []$  do
7      $\hat{\tau}_i \leftarrow 0$ ;
8     for  $j \leftarrow 1$  to  $m$  do
9       Extrair  $\rho_j$  de  $Be(\mu_j)$ 
10      if  $\rho_j = 1$  then
11         $\hat{\Psi}_i \leftarrow [\hat{\Psi}_i^T \psi_j]^T$ ;
12         $\hat{\tau}_i \leftarrow \hat{\tau}_i + 1$ ;
13      end
14    end
15     $M_i [\Psi_i, \tau_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i] \leftarrow [\hat{\Psi}_i, \hat{\tau}_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i]$ ;
16  end
17 end

```

---

respectiva probabilidade de inclusão do regressor ( $\mu_j$ ) (linha 9), se o número aleatório é menor do que  $\mu_j$ ,  $\rho_j$  é igual a 1, o regressor é incluído no modelo e o tamanho do modelo  $\tau_i$  é aumentado em uma unidade. Por fim os regressores selecionados  $\hat{\Psi}_i$  e o tamanho do modelo  $\hat{\tau}_i$  são copiados para o respectivo modelo  $M_i$ .

### 3.3.2 Estimação dos parâmetros e teste-t de Student

Os parâmetros dos modelos são estimados usando mínimos quadrados e então um teste é feito para determinar se os regressores são significantes. Wei e Billings (2008) mostraram que um teste-t de Student pode ser usado efetivamente para remover regressores não significantes baseados na variância dos parâmetros. Neste trabalho o teste-t de Student é usado para eliminar os regressores não significantes e então os parâmetros são reestimados.

No Algoritmo 4,  $u$  e  $y$  são entradas e saídas, respectivamente, as quais são usadas para montar a matriz de regressores  $\Psi$  (Billings, 2013),  $V$  é uma matriz quadrada  $\tau \times \tau$ , e  $V_{hh}$  são os elementos da diagonal de  $V$ . O percentil da distribuição t de Student com  $N - \tau$  graus de liberdade é dado por  $t_{\alpha, N-\tau}$  (Montgomery, 2013). O nível de significância para o teste estatístico é definido como  $\alpha = 0.01$ .

---

**Algoritmo 4:** Estimaco de parmetros, Eliminao dos regressores e filtro ERR
 

---

```

1 Input:  $M_i, \{u(k), y(k), k = 1, \dots, N\}$ ;
2 Output:  $M_i[\hat{\Psi}_i, \hat{\tau}_i, \hat{\Theta}_i]$ ;
3  $V \leftarrow \left( \sum_{k=1}^N \Psi_i(k) \Psi_i^T(k) \right)^{-1}$ ;
4  $\hat{\Theta}_i \leftarrow V \sum_{k=1}^N \Psi_i(k) y(k)$ ;
5  $\hat{\sigma}_e^2 \leftarrow \frac{1}{N-\tau_i} \sum_{k=1}^N \left( y(k) - \Psi_i^T(k) \hat{\Theta}_i \right)^2$ ;
6  $\hat{\Psi}_i \leftarrow \Psi_i$ ;  $\hat{\tau}_i \leftarrow \tau_i$ ;
7 for  $h \leftarrow 1$  to  $\tau_i$  do
8    $\hat{\sigma}_h^2 \leftarrow \hat{\sigma}_e^2 V_{hh}$ ;
9   if  $|\hat{\Theta}_{i,h}| \leq \hat{\sigma}_h t_{\alpha, N-\tau_i}$  then
10      $\hat{\Psi}_i \leftarrow \hat{\Psi}_i - [\psi_h]$ ; // Remove o regressor  $\psi_h$  de  $\Psi_i$ 
11      $\hat{\tau}_i \leftarrow \hat{\tau}_i - 1$ ;
12   end
13 end
14  $\hat{ERR}_i \leftarrow FROLS(M_i[\Psi_i, u, y])$ ;
15 Elimina regressores com ERR menor que  $\delta$  em  $M_i$ , e ordena os regressores em
   forma descendente de ERR nos indivduos;
16  $\hat{\Theta}_i \leftarrow \left( \sum_{k=1}^N \hat{\Psi}_i(k) \hat{\Psi}_i^T(k) \right)^{-1} \sum_{k=1}^N \hat{\Psi}_i(k) y(k)$ ;
17  $M_i[\Psi_i, \tau_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i] \leftarrow [\hat{\Psi}_i, \hat{\tau}_i, \hat{\Theta}_i, ERR_i, BIC_i, \mathcal{J}_i]$ ;

```

---

O Algoritmo 4   executado para cada um dos  $Z_M$  modelos da populao  $M$ . Inicialmente os parmetros so estimados (linha 4) e a varincia do erro   calculada utilizando a sada medida e a sada estimada do modelo (linha 5). Ento o teste-t   realizado para cada um dos regressores do modelo (linhas 7 a 13). Na linha 14 do Algoritmo 4, a funo ‘‘FROLS()’’ se refere ao algoritmo FROLS clssico, o qual   usado como um pr -filtro para a incluso dos regressores quando o regressor no reduz a varincia da sada em pelo menos um valor especificado, i. e., a taxa ERR do regressor no   maior do que  $\delta$ , este regressor   considerado um candidato ruim e no   includo no modelo. A seleo apropriada de  $\delta$    uma limitao prtica do algoritmo OFR-ERR, desde que esta taxa determina o n mero de termos selecionado pelo algoritmo. Usualmente o valor de  $\delta$    selecionado por tentativa e erro e auxiliado por crit rios de informao, tais como BIC, ou testes estat sticos (Guo *et al.*, 2014). Neste trabalho  $\delta$    definido como 0,1%.

Outra caracter stica do m todo RaMSS-EGI   que este transmite sequ ncias de DNA dos pais (parte dos regressores) para a pr xima gerao, isso requer que os modelos tenham

seus regressores organizados. O algoritmo FROLS é usado para organizar os regressores em ordem decrescente de acordo com a importância do índice ERR de cada regressor.

O algoritmo FROLS é detalhado na seção 2.5 no Algoritmo 1, mais detalhes sobre o algoritmo podem ser encontrado nos trabalhos (Korenberg *et al.*, 1988; Wei *et al.*, 2004; Billings, 2013).

### 3.3.3 Atualização da probabilidade de inclusão de regressores (RIP)

A RIP de cada regressor ( $\mu_j$ ) é atualizada de acordo com a Equação (3.8), a RIP do  $j$ -ésimo regressor será aumentada se o desempenho do modelo contendo o  $j$ -ésimo regressor é em média maior que o desempenho dos modelos que não contêm o  $j$ -ésimo regressor, e diminuída caso contrário, o Algoritmo 5 mostra o processo. É importante notar que o Algoritmo 5 é usado duas vezes no Algoritmo 2 quando o método escolhido é o RaMSS-EGI, na primeira vez ele é utilizado para atualizar a RIP baseado na população  $M$  e na segunda vez é usado para atualizar a RIP baseado na melhor população  $M_{best}$ .

---

#### Algoritmo 5: Algoritmo de atualização da RIP

---

```

1 Input:  $M, \mathcal{R}_m, Z_M, \mu, \mu_{lower}, \mu_{upper}$ ;
2 Output:  $\mu$ ;
3 for  $j \leftarrow 1$  to  $m$  do
4    $\mathcal{J}^+ \leftarrow 0; n^+ \leftarrow 0; \mathcal{J}^- \leftarrow 0; n^- \leftarrow 0;$ 
5   for  $i \leftarrow 1$  to  $Z_M$  do
6     if  $\psi_j \in \Psi_i$  then
7        $\mathcal{J}^+ \leftarrow \mathcal{J}^+ + \mathcal{J}(i); n^+ \leftarrow n^+ + 1;$ 
8     else
9        $\mathcal{J}^- \leftarrow \mathcal{J}^- + \mathcal{J}(i); n^- \leftarrow n^- + 1;$ 
10    end
11  end
12   $\gamma \leftarrow \frac{1}{10(\mathcal{J}_{max} - \bar{\mathcal{J}}) + 0.1};$ 
13   $\mu_j \leftarrow \mu_j + \gamma \left( \frac{\mathcal{J}^+}{\max(n^+, 1)} - \frac{\mathcal{J}^-}{\max(n^-, 1)} \right);$ 
14   $\mu_j \leftarrow \max(\min(\mu_j, \mu_{upper}), \mu_{lower});$ 
15 end

```

---

O Algoritmo 5 inicia avaliando cada um dos  $m$  regressores candidatos (linha 3), considerando a presença do  $j$ -ésimo regressor em cada um dos  $Z_M$  modelos. O desempenho dos modelos que contêm o  $j$ -ésimo regressor são somados (linha 7), bem como o desempenho dos modelos que não contêm o  $j$ -ésimo regressor (linha 9). Estes valores são utilizados para

atualizar a probabilidade de inclusão do regressor, aumentando quando o desempenho médio dos modelos que contém o  $j$ -ésimo regressor é superior ao desempenho médio dos modelos que não contém o  $j$ -ésimo regressor (linha 13).

### 3.3.4 RaMSS-ERR e RaMSS-EGI

Os métodos propostos são similares, sendo o RaMSS-EGI uma modificação do RaMSS-ERR. É possível observar no Algoritmo 2 que os métodos se diferenciam em como a população é gerada a partir da segunda iteração em diante, e pela população elitista  $M_{best}$  usada pelo RaMSS-EGI. Quando o método selecionado é o RaMSS-ERR, o método usado para gerar a próxima população é completamente aleatório e usa o Algoritmo 3.

No caso do RaMSS-EGI, em cada iteração é feito um sorteio com base no  $BIC$  de cada modelo, e então os modelos selecionados transmitem parte de seus regressores para o próximo modelo. Estes regressores selecionados para ser transmitidos ao próximo modelo serão parte do modelo e então os outros regressores em  $\mathcal{R}$  são testados se serão incluídos no modelo ou não, utilizando a variável aleatória de Bernoulli  $\rho_j$  e a probabilidade de inclusão de regressores  $\mu$ . A Figura (GenPopRaMSS-EGI) mostra o processo de criação de um indivíduo no método RaMSS-EGI.

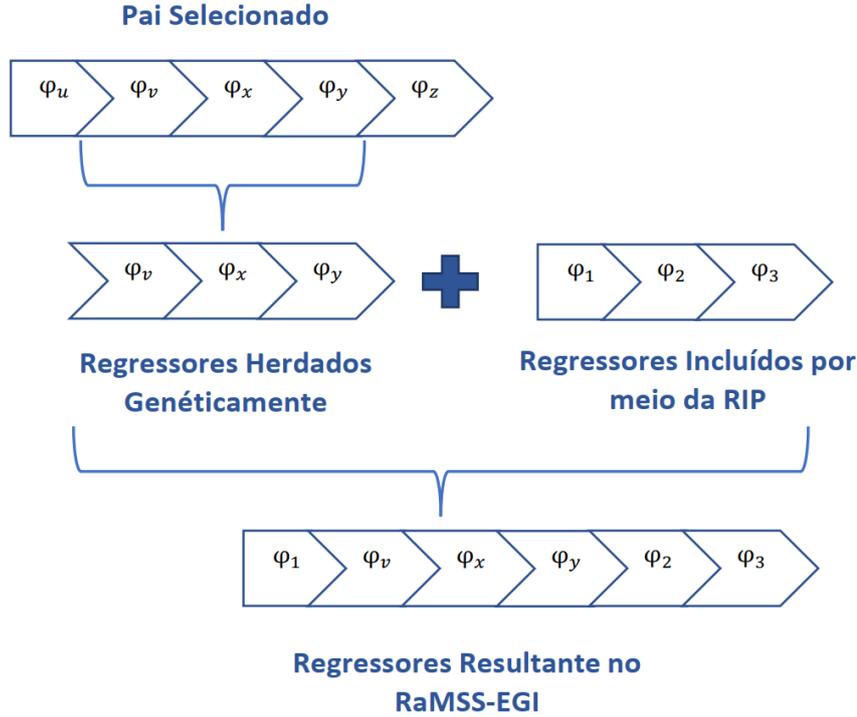
Para explicar a ideia principal no método RaMSS-EGI, considere um modelo com cinco regressores e um conjunto de regressores candidatos de tamanho  $m = 165$ . Então a Equação (3.7) pode ser escrita como,

$$\mathcal{P}_{\phi, RaMSS}(f) = \prod_{j:\psi_j \in f} \mu_j \prod_{j:\psi_j \notin f} (1 - \mu_j) \quad (3.11)$$

É esperado que após um dado número de iterações, os melhores regressores tenham uma probabilidade maior de serem selecionados. Com este ponto em mente, é esperado que o método RaMSS-EGI transmita, pelo menos em parte, os regressores “verdadeiros” do modelo, processo aqui referido como herança genética. Assim, após algumas iterações, e considerando que o processo genético tenha transmitido três dos regressores “verdadeiros”, denotados aqui como  $\psi_{true}$ , a Equação (3.7) pode ser escrita como,

$$\mathcal{P}_{\phi, EGI}(f) = \prod_{j:\psi_j \in (f - \psi_{true})} \mu_j \prod_{j:\psi_j \notin f} (1 - \mu_j) \quad (3.12)$$

Figura 3.3: Processo para gerar indivíduo no algoritmo RaMSS-EGI



Como três dos regressores verdadeiros já foram selecionados pelo algoritmo de herança genética, o problema agora consiste em selecionar os dois regressores restantes não do conjunto  $f$ , mas do conjunto  $(f - \psi_{true})$ , e como  $\mu_j \leq 1$ , tem-se que,

$$\mathcal{P}_{\phi, RaMSS}(f) = \prod_{j:\psi_j \in f} \mu_j \prod_{j:\psi_j \notin f} (1 - \mu_j) \leq \mathcal{P}_{\phi, EGI}(f) = \prod_{j:\psi_j \in (f - \psi_{true})} \mu_j \prod_{j:\psi_j \notin f} (1 - \mu_j) \quad (3.13)$$

Embora  $\mathcal{P}_{\phi, EGI}$  não seja substancialmente maior que  $\mathcal{P}_{\phi, RaMSS}$ , em alguns casos, que serão discutidos no próximo capítulo, nota-se que o uso do RaMSS-EGI ao invés do RaMSS-ERR pode ser muito benéfico. O Algoritmo 6 mostra o processo para gerar a população no método RaMSS-EGI.

No Algoritmo 6, para cada um dos  $Z_M$  modelos, é selecionado um modelo pai  $M_p$  a partir da população  $M$  (linha 4) com base no  $BIC$  de cada modelo, este modelo pai irá transmitir uma sequência aleatória dos regressores para um modelo filho (linhas 5 a 8) e em seguida os  $m$  regressores candidatos serão testados para serem incluídos ou não no modelo (linhas 9 a 15). Os modelos que já foram incluídos no processo anterior de herança genética não são testados para inclusão. Por fim os regressores selecionados  $\hat{\Psi}_i$  e o número de regressores  $\hat{\tau}_i$  são armazenados no modelo  $M_i$  da população.

---

**Algoritmo 6:** Gerar população no método RaMSS-EGI
 

---

```

1 Input:  $M, Z_M, \mathcal{R}_m, \mu$ 
2 Output:  $M$ ;
3 for  $i \leftarrow 1$  to  $Z_M$  do
4   Seleccione o modelo pai  $M_p$  a partir da população  $M$ , usando sorteio;
5    $i_p \leftarrow \text{rand}([1, \tau_i/2])$ ;
6    $f_p \leftarrow \text{rand}([\tau_i/2, \tau_i])$ ;
7    $\hat{\Psi}_i \leftarrow M_p([\psi_{i_p}, \dots, \psi_{f_p}])$ ;
8    $\hat{\tau}_i \leftarrow$  número de regressores em  $\hat{\Psi}_i$ ;
9   for  $j \leftarrow 1$  to  $m$  do
10    Extraia  $\rho_j$  de  $Be(\mu_j)$ 
11    if  $\rho_j = 1$  E  $\psi_j \notin \hat{\Psi}_i$  then
12       $\hat{\Psi}_i \leftarrow [\hat{\Psi}_i^T \psi_j]^T$ ;
13       $\hat{\tau}_i \leftarrow \hat{\tau}_i + 1$ ;
14    end
15  end
16   $M_i[\Psi_i, \tau_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i] \leftarrow [\hat{\Psi}_i, \hat{\tau}_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i]$ ;
17 end

```

---

### 3.4 Considerações Finais

Neste capítulo foram apresentados os algoritmos para implementação do método proposto para sistemas SISO em um cenário probabilístico. O procedimento proposto pode ser sintetizado nos Algoritmos 2 a 6. No próximo capítulo são apresentados os algoritmos para implementação do método proposto para sistemas MIMO.



# CAPÍTULO IV

## IDENTIFICAÇÃO DE SISTEMAS MISO E MIMO

O presente capítulo apresenta a formulação do problema de seleção da estrutura de modelos em termos de distribuição de probabilidade para o caso de sistemas com múltiplas entradas e múltiplas saídas, sendo os sistemas com  $q$  entradas e  $s$  saídas ( $q \times s$ ) os sistemas MIMO, e os sistemas com  $q$  entradas e saída única ( $q \times 1$ ) um caso particular denominado MISO (*Multiple Input and Single Output*). A formulação do problema da seleção da estrutura em termos de distribuição de probabilidades é mais adequada para a compreensão da Seleção Aleatória da Estrutura de Modelos e por isso é detalhada a seguir. Em um segundo momento, segue-se a discussão do algoritmo proposto para a identificação de Sistemas MISO/MIMO.

### 4.1 Formulação do Problema e a Seleção Aleatória da Estrutura de Modelos MIMO

#### 4.1.1 Formulação do Problema de Seleção da Estrutura de Modelos MIMO

Seja um conjunto de medidas experimentais de todas as  $s$  saídas do sistema em questão  $y(k) = [y_1(k), y_2(k), \dots, y_s(k)]$ , contendo ruído ou não, e todas as entradas  $u(k) = [u_1(k), u_2(k), \dots, u_q(k)]$ , de um sistema MIMO, define-se o conjunto  $\mathcal{F}_t = 2^{m_t}$  como o conjunto de todas as possibilidades de estruturas de modelos associado à saída  $t$ , para um conjunto específico de  $m_t$  regressores  $\mathcal{R}_{m_t} = \{\psi_{t,1}, \dots, \psi_{t,m_t}\}$  no qual os regressores são formados pela combinação das entradas passadas de  $y(k)$  e  $u(k)$ , o problema da seleção da estrutura de

modelos para o caso MIMO consiste em encontrar para cada saída  $s$  o melhor modelo  $f_t^*$ , no qual o desempenho do modelo  $\mathcal{J}_t$  assume o valor máximo, isto é,

$$\begin{bmatrix} f_1^* \\ \vdots \\ f_t^* \end{bmatrix} = \begin{bmatrix} \arg \max_{f_1 \in \mathcal{F}_1} \mathcal{J}_1(f_1) \\ \vdots \\ \arg \max_{f_t \in \mathcal{F}_t} \mathcal{J}_t(f_t) \end{bmatrix}, \quad (4.1)$$

em que  $t \in [1, 2, \dots, s]$  e  $\mathcal{J}_t$  é o desempenho do  $t$ -ésimo modelo associado com a saída  $t$ , discutido na sessão 2.7.

Assim como no caso SISO, para o caso MIMO também é considerado que existe apenas um único melhor modelo, o qual coincide com o modelo verdadeiro, o que também implica dizer que o grau de não linearidade do conjunto  $\mathcal{R}_{m_t}$  tem no mínimo o mesmo grau de não linearidade do modelo verdadeiro associado à saída  $t$ .

Definindo  $\phi_t$  como uma variável aleatória discreta assumindo valores em  $\mathcal{F}_t$ , de acordo com a distribuição de probabilidade  $\mathcal{P}_{\phi_t}$ . O valor esperado do desempenho de  $\phi_t$  pode ser dado por uma combinação convexa do índice de desempenho dos modelos ponderado por suas probabilidades como mostra a Equação (4.2).

$$\begin{bmatrix} \mathbb{E}[\mathcal{J}_1(\phi_1)] \\ \vdots \\ \mathbb{E}[\mathcal{J}_t(\phi_t)] \end{bmatrix} = \begin{bmatrix} \sum_{f_1 \in \mathcal{F}_1} \mathcal{J}_1(f_1) \mathcal{P}_{\phi_1}(f_1) \\ \vdots \\ \sum_{f_t \in \mathcal{F}_t} \mathcal{J}_t(f_t) \mathcal{P}_{\phi_t}(f) \end{bmatrix}, \quad (4.2)$$

com  $t \in [1, 2, \dots, s]$ .

Desta forma, observa-se que o problema de seleção da estrutura dos modelos MIMO com  $s$  saídas pode ser dividido em  $s$  problemas independentes, considerando em cada modelo todos os conjuntos de entradas  $q$  e saídas  $s$ . Além disso, o problema também pode ser reformulado em encontrar a distribuição de probabilidade  $\mathcal{P}_{\phi_t}^*$  que maximiza a Equação (4.2)

para cada saída  $s$ , como na Equação (4.3).

$$\begin{bmatrix} \mathcal{P}_{\phi_1}^* \\ \vdots \\ \mathcal{P}_{\phi_t}^* \end{bmatrix} = \begin{bmatrix} \arg \max_{\mathcal{P}_{\phi_1}} \mathbb{E} [\mathcal{J}_1(\phi_1)] \\ \vdots \\ \arg \max_{\mathcal{P}_{\phi_t}} \mathbb{E} [\mathcal{J}_t(\phi_t)] \end{bmatrix}, \quad (4.3)$$

#### 4.1.2 Seleção Aleatória da Estrutura de Modelos MIMO

Assim como no caso SISO, o problema de seleção da estrutura de modelos para o caso MIMO pode ser resolvido utilizando os algoritmos para a Seleção Aleatória da Estrutura de Modelos (RaMSS).

Considere um conjunto  $\mathcal{R}_{m_t} = \{\psi_{t,1}, \dots, \psi_{t,m}\}$  de  $m_t$  regressores candidatos associados com o  $t$ -ésimo modelo. Para cada regressor  $\psi_{t,j}$  pertencente ao conjunto  $\mathcal{R}_{t,m}$ , um índice  $\mathcal{I}_{t,j}^p$  é calculado sobre um conjunto de modelos candidatos pelas Equações (4.4) a (4.6).

$$\mathcal{I}_{t,j}^p = \mathcal{I}_{t,j}^{p+} - \mathcal{I}_{t,j}^{p-} \quad (4.4)$$

$$\mathcal{I}_{t,j}^{p+} = \frac{1}{n_{t,j}^+} \sum_{i=1}^{n_{t,j}^+} \mathcal{J}_{t,i}^+ \quad (4.5)$$

$$\mathcal{I}_{t,j}^{p-} = \frac{1}{n_{t,j}^-} \sum_{i=1}^{n_{t,j}^-} \mathcal{J}_{t,i}^- \quad (4.6)$$

em que  $\mathcal{I}_{t,j}^{p+}$  é o desempenho médio dos modelos que contém o  $j$ -ésimo regressor associado com a saída  $t$ ,  $n_{t,j}^+$  é o número de modelos que contém o  $j$ -ésimo regressor e  $\mathcal{J}_{t,j}^+$  é o desempenho dos modelos que contém o  $j$ -ésimo regressor calculado usando  $MSP E_t$ . Na Equação (4.6), analogamente à Equação (4.5),  $\mathcal{I}_{t,j}^{p-}$ ,  $n_{t,j}^-$  e  $\mathcal{J}_{t,j}^-$  se referem ao desempenho médio, número de modelos, e desempenho dos modelos que não contém o  $j$ -ésimo regressor, respectivamente.

A distribuição de probabilidade  $\mathcal{P}_{\phi_t}$  na Equação (4.2) é parametrizada associando uma variável aleatória de Bernoulli  $\rho_{t,j} \sim Be(\mu_{t,j})$ , com  $\mu_{t,j} \in ]0, 1[$ , para cada regressor  $\psi_{t,j}$ ,  $t = 1, \dots, s$  e  $j = 1, \dots, m$ . A variável  $\mu_{t,j}$  estabelece se o regressor será incluído no modelo ou não.

Assim, supondo todas as probabilidades  $\rho_{t,j}$ ,  $t = 1, \dots, s$  e  $j = 1, \dots, m$  independentes, a probabilidade de escolher um modelo específico associado à saída  $t$  é o produto da RIP de cada regressor que pertence ao modelo  $t$ , vezes o produto do complemento da RIP de cada regressor que não pertence ao modelo  $t$ , isto é

$$\mathcal{P}_{\phi_t}(f) = \prod_{j:\psi_{t,j} \in f_t} \mu_{t,j} \prod_{j:\psi_{t,j} \notin f_t} (1 - \mu_{t,j}) \quad (4.7)$$

Posto o problema na Equação (4.7) observa-se que este é muito semelhante ao problema na Equação (3.7), de forma que pode ser solucionado atualizando a distribuição de probabilidade  $\mathcal{P}_{\phi_t}$ , de forma que esta se concentre no modelo “verdadeiro”, ou seja, utilizando o algoritmo RaMSS.

Para o caso MIMO, o algoritmo RaMSS inicia gerando uma população de modelos de tamanho  $M$  para cada saída  $t$  e considerando para cada um dos modelos os regressores formados por todas as  $q$  entradas e  $s$  saídas. A inclusão dos regressores neste modelo é definida pela variável aleatória de Bernoulli  $\rho_{t,j}$ , assim, para cada modelo  $t$  e regressores de 1 até  $m_t$ , um número aleatório é gerado, se este número é menor ou igual a  $\mu_{t,j}$  do regressor correspondente, este regressor será incluído no modelo  $t$ . Este processo é repetido até que se tenha os  $M$  modelos que irão compor a população inicial com relação à saída  $t$ .

Definido a estrutura do modelo, os parâmetros são calculados utilizando mínimos quadrados, e a significância estatística de cada parâmetro é testada usando um teste-t de Student, aqueles que não são significantes são eliminados e os parâmetros são reestimados.

O desempenho dos modelos são calculados utilizando os índices preditivo, simulado ou até mesmo os dois (de acordo com o  $\beta$  definido pelo usuário), como na Equação (2.49).

A RIPs dos regressores é então atualizada usando o desempenho dos modelos como na Equação (4.8),

$$\mu_{t,j}(i+1) = \mu_{t,j}(i) + \gamma \mathcal{I}_{t,j} \quad (4.8)$$

em que  $\mathcal{I}_{t,j}$  é o índice definido pelas Equações (4.4) a (4.6) e  $\gamma$  é dado por

$$\gamma_t = \frac{1}{10 (\mathcal{J}_{tmax} - \bar{\mathcal{J}}_{\square}) + 0.1} \quad (4.9)$$

em que  $\mathcal{J}_{tmax}$  é o desempenho do melhor modelo para a saída  $t$  na população atual e  $\bar{\mathcal{J}}_t$  é o desempenho médio dos modelos para a saída  $t$ .

Na próxima iteração, o algoritmo gera as estruturas de modelos baseado na RIP atualizada, e então os mesmos passos são executados até que as RIPs converjam para a distribuição limite.

## 4.2 Métodos propostos para o caso MIMO

A seguir as duas novas abordagens propostas são estendidas para o caso MIMO. Os algoritmos propostos para o caso MIMO seguem as mesmas ideias centrais dos já apresentados anteriormente para o caso SISO. O problema MIMO é dividido em  $s$  problemas de múltiplas entradas e saída única ( $s$  problemas MISO), e o índice  $t$  representa o  $t$ -ésimo modelo associado com a  $t$ -ésima saída.

## 4.3 Algoritmo Geral para o caso MIMO

Para tratar o problema de uma forma multidimensional, a estrutura de dados proposta para o caso MIMO consiste em uma população  $M_t$ , Equação (4.10), formada por um conjunto de regressores  $\Psi_t$  de dimensão  $\tau_t \times N$ , o número de regressores  $\tau_t$ , um vetor de parâmetros correspondendo a cada regressor  $\Theta_t$  de dimensão  $\tau_t \times 1$ , um vetor contendo o ERR associado a cada regressor no modelo  $ERR_t$  de dimensão  $\tau_t \times 1$ , o valor do BIC associado ao modelo  $BIC_t$  e por fim o desempenho do modelo dado por  $\mathcal{J}_t$ . O Algoritmo 7 sumariza o processo geral de identificação para o caso MIMO.

$$M_{t,i} = [\Psi_{t,i}; \tau_{t,i}; \Theta_{t,i}; ERR_{t,i}; BIC_{t,i}; \mathcal{J}_{t,i}] \quad (4.10)$$

No Algoritmo 7,  $[\mathbf{u}(k), \mathbf{y}(k)]$  são definidos como os conjuntos de todas as entradas e saídas, isto é,  $\{[u_1(k), \dots, u_q(k)], [y_1(k), \dots, y_s(k)], k = 1, \dots, N\}$ , e são todas incluídas no conjunto de regressores candidatos  $\mathcal{R}_{m_t}$  de dimensão  $m_t \times 1$ ,  $Z_M$  é o número de indivíduos,  $K$  é o parâmetro de sintonia nas Equações (2.47) e (2.48),  $\beta$  é um parâmetro de sintonia na Equação (2.49),  $\mu_i$  é a RIP inicial,  $\mu_{lower}$ ,  $\mu_{upper}$  são os limites inferiores e superiores para a RIP e  $\delta$  é o limiar definido para o ERR. O critério de parada é definido pelo valor de  $\kappa$ , o

qual define as faixas inferiores e superiores para considerar que a RIP de todos os regressores convergiu. A saída do Algoritmo 7 é a probabilidade de inclusão dos regressores  $\mu_t$  a qual parametriza a distribuição de probabilidade  $\mathcal{P}_{\phi,t}$  relacionada com o modelo  $t$ .

---

**Algoritmo 7:** Algoritmo Geral RaMSS-ERR e RaMSS-EGI no caso MIMO

---

```

1 Input:  $\{\mathbf{u}(k), \mathbf{y}(k)\}, k = 1, \dots, N\}$ ,  $\mathcal{R}_m$ ,  $Z_M$ ,  $K$ ,  $\beta$ ,  $\mu_i$ ,  $\mu_{lower}$ ,  $\mu_{upper}$ ,  $\delta$ ,  $\kappa$ ;
2 Output:  $\mu_1, \dots, \mu_s$ ;
3 Escolha o método RaMSS-ERR ou RaMSS-EGI;
4  $\mu \leftarrow \mu_i$ ;
5 for  $t \leftarrow 1$  to  $s$  do
6   Gerar população Inicial  $M_t$  usando o Algoritmo 8 ;
7   repeat
8     for  $i \leftarrow 1$  To  $Z_M$  do
9       Estimar os parâmetros do modelo usando o Algoritmo 9;
10      Calcular o  $BIC_{t,i}$  e o desempenho do modelo  $\mathcal{J}_{t,i}$  usando as Equações
11      (2.51) e (2.49);
12    end
13    Atualizar  $\mu_t$  para cada regressor usando a população  $M_t$  como entrada no
14    Algoritmo 10;
15    if  $Method = RaMSS-EGI$  then
16      Selecionar o melhor indivíduo da população  $M_t$  e copiá-lo para a
17      população  $M_{t,best}$ ;
18      Atualizar  $\mu_t$  para cada regressor usando a população  $M_{t,best}$  como
19      entrada no Algoritmo 10;
20      Gerar a próxima população  $M_t$  usando  $\mu_t$  atualizada, e o Algoritmo 11;
21    else
22      Gerar a próxima população  $M_t$  usando  $\mu$  atualizada e o Algoritmo 8;
23    until  $\min_{j=1, \dots, m} |2\mu_{t,j} - 1| \geq 1 - \kappa$ ;
24  end

```

---

O Algoritmo 7 tem como entrada os valores medidos das entradas  $\mathbf{u}$  e saídas  $\mathbf{y}$  do sistema, os conjuntos de regressores  $\mathcal{R}_{m_t}$ , o tamanho da população  $Z_M$ , e os parâmetros  $K$ ,  $\beta$ ,  $\mu_{t,i}$ ,  $\mu_{lower}$ ,  $\mu_{upper}$ ,  $\delta$  e  $\kappa$ . Como saída é gerado um vetor de probabilidades de inclusão de regressores  $[\mu_1, \dots, \mu_s]$  para cada saída  $s$ . O algoritmo é executado  $s$  vezes gerando uma população  $M_t$  para cada uma das saídas do sistema (linha 5 e 6). Em seguida os parâmetros dos modelos são calculados para cada um dos modelos presentes na população  $Z_M$  (linha 9), em seguida o  $BIC_{t,i}$  e o desempenho  $\mathcal{J}_{t,i}$  são calculados (linha 10). Calculado os índices de desempenho, a probabilidade de inclusão dos regressores  $\mu_t$  é atualizada com base na população  $M_t$ .

De acordo com o método selecionado RaMSS-EGI ou RaMSS-ERR o algoritmo

realiza etapas distintas, no primeiro caso, é gerado uma população  $M_{t,best}$  que armazena os melhores modelos da população  $M_t$ , os quais são utilizados para atualizar a probabilidade de inclusão dos regressores  $\mu_t$  e a população é gerada com base no Algoritmo 11 (linhas 14 a 16). No caso do algoritmo RaMSS-ERR a população é gerada de forma aleatória usando o Algoritmo 8 (linha 19).

#### 4.3.1 Gerar população inicial no caso MIMO

O Algoritmo 8 é usado para gerar a população inicial no caso MIMO. As entradas do algoritmo são o tamanho da população  $Z_M$ , o conjunto de regressores candidatos  $\mathcal{R}_{m_t}$  e a probabilidade de inclusão dos regressores  $\mu_t$ , como saída o algoritmo gera uma população  $M_{t,i}$ . De forma semelhante ao Algoritmo 3, os regressores candidatos são esvaziados e o tamanho do modelo é zerado (linhas 4 e 5), a seguir cada um dos  $m_t$  regressores candidatos são testados para a inclusão, isto é feito gerando um número aleatório entre 0 e 1, se o número é menor do que  $\mu_{t,j}$ ,  $\rho_j$  é igual a 1 e o  $j$ -ésimo regressor é incluído no modelo, bem como o tamanho do modelo é incrementado em uma unidade (linhas 7 a 13). A seguir os valores dos regressores selecionados  $\hat{\Psi}_i$  e tamanho do modelo  $\hat{\tau}_i$  são copiados para o modelo  $M_{t,i}$  (linha 14).

---

#### Algoritmo 8: Gerar população inicial no caso MIMO

---

```

1 Input:  $Z_M, \mathcal{R}_{m_t}, \mu_t$ 
2 Output:  $M_{t,i}$ ;
3 for  $i \leftarrow 1$  To  $Z_M$  do
4    $\hat{\Psi}_i \leftarrow [ ]$ ;
5    $\hat{\tau}_i \leftarrow 0$ ;
6   while  $\hat{\Psi}_i = [ ]$  do
7     for  $j \leftarrow 1$  To  $m_t$  do
8       Extraia  $\rho_j$  a partir de  $Be(\mu_{t,j})$ 
9       if  $\rho_j = 1$  then
10          $\hat{\Psi}_i \leftarrow [ \hat{\Psi}_i^T \ \psi_j ]^T$ ;
11          $\hat{\tau}_i \leftarrow \hat{\tau}_i + 1$ ;
12       end
13     end
14      $M_{t,i} [\Psi_{t,i}, \tau_{t,i}, \Theta_{t,i}, ERR_{t,i}, BIC_{t,i}, \mathcal{J}_{t,i}] \leftarrow [ \hat{\Psi}_i, \hat{\tau}_i, \Theta_i, ERR_i, BIC_i, \mathcal{J}_i ]$ ;
15   end
16 end

```

---

### 4.3.2 Estimação dos parâmetros e teste-t de Student no caso MIMO

Os parâmetros dos modelos são estimados usando o método dos mínimos quadrados e posteriormente submetidos a um teste-t de Student para averiguar a significância dos parâmetros. O processo de estimação e eliminação dos parâmetros é mostrado no Algoritmo 9.

---

**Algoritmo 9:** Estimação dos parâmetros e Eliminação dos Regressores para o caso MIMO

---

```

1 Input:  $M_{t,i}$ ,  $\{[\mathbf{u}(k), \mathbf{y}(k)], k = 1, \dots, N\}$ ,  $Z_M$ ;
2 Output:  $M_{t,i}[\hat{\Psi}_i, \hat{\tau}_i, \hat{\Theta}_i]$ ;
3  $V \leftarrow \left( \sum_{k=1}^N \Psi_{t,i}(k) \Psi_{t,i}^T(k) \right)^{-1}$ ;
4  $\hat{\Theta}_i \leftarrow V \sum_{k=1}^N \Psi_{t,i}(k) y_t(k)$ ;
5  $\hat{\sigma}_e^2 \leftarrow \frac{1}{N - \tau_{t,i}} \sum_{k=1}^N \left( y_t(k) - \Psi_{t,i}^T(k) \hat{\Theta}_i \right)^2$ ;
6  $\hat{\Psi}_i \leftarrow \Psi_{t,i}$ ;  $\hat{\tau}_i \leftarrow \tau_{t,i}$ ;
7 for  $h \leftarrow 1$  to  $\tau_{t,i}$  do
8    $\hat{\sigma}_h^2 \leftarrow \hat{\sigma}_e^2 V_{hh}$ ;
9   if  $|\hat{\Theta}_{i,h}| \leq \hat{\sigma}_h T_{\alpha, N - \tau_{t,i}}$  then
10      $\hat{\Psi}_i \leftarrow \hat{\Psi}_i - [\psi_h]$ ; // Remove regressor  $\psi_h$  de  $\hat{\Psi}_i$ 
11      $\hat{\tau}_i \leftarrow \hat{\tau}_i - 1$ ;
12   end
13 end
14  $\hat{ERR}_{t,i} \leftarrow \text{FROLS}(\hat{\Psi}_i, [u, y])$ ;
15 Elimine os regressores em  $\hat{\Psi}_i$  com um ERR menor que  $\delta$ , e ordene os indivíduos em ordem decrescente de ERR;
16  $\hat{\Theta}_i \leftarrow \left( \sum_{k=1}^N \hat{\Psi}_i(k) \hat{\Psi}_i^T(k) \right)^{-1} \sum_{k=1}^N \hat{\Psi}_i(k) y_t(k)$ ;
17  $M_{t,i}[\Psi_{t,i}, \tau_{t,i}, \Theta_{t,i}, ERR_{t,i}, BIC_{t,i}, \mathcal{J}_{t,i}] \leftarrow [\hat{\Psi}_i, \hat{\tau}_i, \hat{\Theta}_i, ERR_{t,i}, BIC_{t,i}, \mathcal{J}_{t,i}]$ ;

```

---

No Algoritmo 9,  $V$  é uma matriz quadrada de dimensão  $\tau_t \times \tau_t$ , e  $V_{hh}$  é o  $h$ -ésimo elemento da diagonal de  $V$ . O percentil da distribuição t de Student com  $N - \tau_t$  graus de liberdade é dado por  $T_{\alpha, N - \tau_t}$  (Montgomery, 2013). O nível de significância para o teste estatístico é definido como  $\alpha = 0,01$ .

O algoritmo é executado para cada um dos modelos na população  $M_t$ , os parâmetros são estimados e a variância do erro é calculada com base na saída  $y_t$  e na saída estimada  $\hat{y}_t$  (linhas 4 e 5). Em seguida o nível de significância dos parâmetros são testados usando o teste-t (linhas 7 a 13). Na linha 14 a taxa de redução do erro é calculada para cada um dos

regressores candidatos e então os regressores com ERR menor do que  $\delta$  são eliminados do modelo (linha 14 e 15). Por fim os parâmetros são novamente estimados e os valores de  $\hat{\Psi}_i$ ,  $\hat{\tau}_i$ ,  $\hat{\Theta}_i$ ,  $ERR_{t,i}$ ,  $BIC_{t,i}$ ,  $\mathcal{J}_{t,i}$  são atualizados no modelo  $M_{t,i}$  (linhas 16 e 17).

#### 4.3.3 Atualização da probabilidade de inclusão do regressor (RIP) no caso MIMO

A probabilidade de inclusão do regressor relacionado com a  $t$ -ésima saída, e o  $j$ -ésimo regressor em  $\mathcal{R}_{m_t}$  é representada por  $\mu_{t,j}$ . O índice RIP é atualizado de acordo com a Equação (4.8), e é aumentado se, em média, o desempenho dos modelos que contém o  $j$ -ésimo regressor é maior do que o desempenho daqueles modelos que não contém o  $j$ -ésimo regressor, caso contrário a RIP é diminuída. O Algoritmo 10 mostra o procedimento.

O Algoritmo 4.8 pode ter como base a população  $M_t$  ou a população  $M_{t,best}$ , em ambos os casos o algoritmo procede analisando cada regressor pertencente ao conjunto  $\mathcal{R}_{m_t}$ . Cada um dos  $Z_M$  modelos são avaliados se possuem ou não o  $j$ -ésimo regressor, o desempenho dos modelos que possuem o  $j$ -ésimo regressor são somados e os que não possuem o  $j$ -ésimo regressor são somados em separado (linhas 5 a 11). Posteriormente tais valores são utilizados para atualizar a probabilidade de inclusão do regressor  $\mu_{t,j}$  aumentando se, em média, o desempenho dos modelos que contém o  $j$ -ésimo regressor é maior do que o desempenho dos modelos que não contém o  $j$ -ésimo regressor e diminuindo em caso contrário (linha 13).

---

#### Algoritmo 10: Algoritmo para atualização da RIP para o caso MIMO

---

```

1 Input:  $M_t, \mathcal{R}_{m_t}, Z_M, \mu_t, \mu_{lower}, \mu_{upper}$ ;
2 Output:  $\mu_t$ ;
3 for  $j \leftarrow 1$  to  $m_t$  do
4    $\mathcal{J}^+ \leftarrow 0; n^+ \leftarrow 0; \mathcal{J}^- \leftarrow 0; n^- \leftarrow 0$ ;
5   for  $i \leftarrow 1$  to  $Z_M$  do
6     if  $\psi_j \in \Psi_{t,i}$  then
7        $\mathcal{J}^+ \leftarrow \mathcal{J}^+ + \mathcal{J}_{t,i}; n^+ \leftarrow n^+ + 1$ ;
8     else
9        $\mathcal{J}^- \leftarrow \mathcal{J}^- + \mathcal{J}_{t,i}; n^- \leftarrow n^- + 1$ ;
10    end
11  end
12   $\gamma \leftarrow \frac{1}{10(\mathcal{J}_{max} - \bar{\mathcal{J}}) + 0.1}$ ;
13   $\mu_{t,j} \leftarrow \mu_{t,j} + \gamma \left( \frac{\mathcal{J}^+}{\max(n^+, 1)} - \frac{\mathcal{J}^-}{\max(n^-, 1)} \right)$ ;
14   $\mu_{t,j} \leftarrow \max(\min(\mu_{t,j}, \mu_{upper}), \mu_{lower})$ ;
15 end

```

---

#### 4.3.4 Gerar população para o método RaMSS-EGI no caso MIMO

Da mesma forma que no caso SISO, a população no método RaMSS-EGI no caso MIMO é gerada em um framework evolutivo. Em cada iteração é feito um sorteio utilizando o método roleta e com base no  $BIC$  de cada modelo, é selecionado um modelo pai  $M_p$  a partir da população  $M_{best}$ , e então parte desse modelo é selecionada para ser transmitida para um modelo da próxima geração. Estes regressores farão parte do modelo e então os outros regressores em  $\mathcal{R}_m$  serão testados para ser incluídos no modelo com base na variável aleatória de Bernoulli  $\rho_j$  e a probabilidade de inclusão do regressor  $\mu_t$ .

No Algoritmo 11 para cada um dos  $Z_M$  modelos é selecionado um modelo pai  $M_p$  a partir da população  $M_{t,best}$  que irá transferir uma sequência dos regressores para um modelo filho (linhas 4 a 8), em seguida cada um dos  $m_t$  regressores são testados para inclusão no modelo (linhas 10 a 16). Por fim o modelo  $M_{t,i}$  é atualizado com o novo conjunto de regressores  $\hat{\Psi}_{t,i}$  e com o tamanho do novo modelo  $\hat{\tau}_{t,i}$  (linha 17).

---

#### Algoritmo 11: Gerar população RaMSS-EGI para o caso MIMO

---

```

1 Input:  $M_t, M_{t,best}, Z_M, \mathcal{R}_m, \mu_t$ 
2 Output:  $M_t$ ;
3 for  $i \leftarrow 1$  To  $Z_M$  do
4   Seleccione o modelo pai  $M_p$  a partir da população  $M_{t,best}$ , usando o método
   Roleta;
5    $\tau \leftarrow M_p[\tau]$ ;
6    $i_p \leftarrow rand([1, \tau/2])$ ;
7    $f_p \leftarrow rand([\tau/2, \tau])$ ;
8    $\hat{\Psi}_{t,i} \leftarrow M_p([\psi_{ip}, \dots, \psi_{fp}])$ ;
9    $\hat{\tau}_{t,i} \leftarrow$  número de regressores em  $\hat{\Psi}_{t,i}$ ;
10  for  $j \leftarrow 1$  To  $m_t$  do
11    Extraia  $\rho_j$  a partir de  $Be(\mu_{t,j})$ 
12    if  $\rho_j = 1$  AND  $\psi_j \notin \hat{\Psi}_{t,i}$  then
13       $\hat{\Psi}_{t,i} \leftarrow [\hat{\Psi}_{t,i}^T \ \psi_j]^T$ ;
14       $\hat{\tau}_{t,i} \leftarrow \hat{\tau}_{t,i} + 1$ ;
15    end
16  end
17   $M_{t,i} [\Psi_{t,i}, \tau_{t,i}, \Theta_{t,i}, ERR_{t,i}, BIC_{t,i}, \mathcal{J}_{t,i}] \leftarrow [\hat{\Psi}_{t,i}, \hat{\tau}_{t,i}, \Theta_{t,i}, ERR_{t,i}, BIC_{t,i}, \mathcal{J}_{t,i}]$ ;
18 end

```

---

#### **4.4 Considerações Finais**

O presente Capítulo tratou dos algoritmos para implementação dos dois métodos para seleção da estrutura de modelos para sistemas com múltiplas entradas e múltiplas saídas em um cenário probabilístico, com o auxílio da ERR e de herança genética. O procedimento proposto pode ser sintetizado nos Algoritmos 7 a 11.

No próximo capítulo os algoritmos propostos são avaliados para diversos sistemas de referência, tanto SISO quanto MIMO. São também abordadas questões relativas aos parâmetros essenciais para funcionamento dos métodos propostos.



# CAPÍTULO V

## AVALIAÇÃO DOS ALGORITMOS DE IDENTIFICAÇÃO PROPOSTOS

Este capítulo tem por objetivo apresentar os resultados obtidos com o uso do RaMSS-ERR e RaMSS-EGI no problema da seleção da estrutura de diferentes sistemas, tanto para o caso de uma entrada e uma saída (SISO) quanto para o caso de múltiplas entradas e uma saída (MISO) e também para o caso de múltiplas entradas e múltiplas saídas (MIMO). O caso SISO será tratado na seção 5.1 e os casos MISO e MIMO logo em seguida na seção 5.2.

### 5.1 Sistemas SISO

Na seção 5.1.1 são apresentados alguns modelos que são utilizados como referência para se medir o desempenho dos métodos propostos, tais modelos são retirados da literatura e apresentam características distintas como não linearidades na entrada ou na saída, ou ainda não linearidade entre entrada e saída. Na seção 5.1.2 alguns parâmetros para ambos os métodos são mostrados, estes parâmetros serão usados nos testes nas próximas seções.

O primeiro teste na seção 5.1.3 investiga a influência do ERR nos algoritmos aleatórios. Na próxima seção 5.1.4, sendo o tamanho da população um parâmetro de grande importância na velocidade computacional, é feito um teste para determinar o tamanho mais adequado da população. Na seção 5.1.5, os métodos propostos são aplicados nos cinco modelos da Tabela (5.1) com o objetivo de comparar o desempenho dos métodos propostos com

o RaMSS clássico. Seguindo para a seção 5.1.6, alguns modelos de referência retirados da literatura são utilizados para verificar o desempenho dos métodos propostos em identificar modelos sujeitos a entradas pouco excitativas. Na seção 5.1.7, os algoritmos RaMSS-ERR e RaMSS-EGI são aplicados na identificação de um sistema com grandes atrasos e alto grau de não linearidade, usando um grande conjunto de regressores candidatos. Finalmente, é feito a identificação de um Reator Tanque Agitado Contínuo (*Continuous Stirred-tank Reactor - CSTR*) usando os métodos propostos na seção 5.1.8.

### 5.1.1 Modelos SISO utilizados como referência para testes

Alguns sistemas de referência são utilizados nos testes que se seguem. Tais sistemas contam com diferentes tipos de estruturas e não linearidades como: não linearidades na entrada, sistema  $\mathcal{S}_1$  (Bonin *et al.*, 2010); um sistema linear, sistema  $\mathcal{S}_2$  (Wei e Billings, 2008); não linearidades na entrada e na saída, sistema  $\mathcal{S}_3$  (Mao e Billings, 1997); não linearidades entre entrada e saída, sistema  $\mathcal{S}_4$  (Baldacchino *et al.*, 2013); e até mesmo um sistema pertencente à classe NARMAX que apresenta ruído de média móvel, sistema  $\mathcal{S}_5$  (Baldacchino *et al.*, 2013). Esses modelos são mostrados na Tabela (5.1) em que *WGN* e *WUN* se referem respectivamente ao ruído branco gaussiano (*White Gaussian Noise*) e ruído branco uniforme (*White Uniform Noise*).

Tabela 5.1: Modelos de Benchmark usados para testar o desempenho dos métodos RaMSS-ERR e RaMSS-EGI.

Sistema	Estrutura Conhecida	Entrada( $u$ )	Ruído( $e$ )
$\mathcal{S}_1$	$y(k) = 0, 8y(k-1) + 0, 4u(k-1) + 0, 4u^2(k-1) + 0, 4u^3(k-1) + e(k)$	WGN(0;1)	WGN(0; 0, 33 <sup>2</sup> )
$\mathcal{S}_2$	$y(k) = -1, 7y(k-1) - 0, 8y(k-2) + u(k-1) + 0, 81u(k-2) + e(k)$	WUN(-2;2)	WGN(0; 0, 01)
$\mathcal{S}_3$	$y(k) = 0, 2y^3(k-1) + 0, 7y(k-1)u(k-1) + 0, 6u^2(k-2) - 0, 7y(k-2)u^2(k-2) - 0, 5y(k-2) + e(k)$	WUN(-1;1)	WGN(0;0,01)
$\mathcal{S}_4$	$y(k) = 0, 7y(k-1)u(k-1) - 0, 5y(k-2) + 0, 6u^2(k-2) - 0, 7y(k-2)u^2(k-2) + e(k)$	WUN(-1;1)	WGN(0;0,004)
$\mathcal{S}_5$	$y(k) = 0, 7y(k-1)u(k-1) - 0, 5y(k-2) + 0, 6u^2(k-2) - 0, 7y(k-2)u^2(k-2) + 0, 2e(k-1) - 0, 3u(k-1)e(k-2) + e(k)$	WUN(-1;1)	WGN(0;0,02)

Estes modelos enfrentam algumas dificuldades para serem identificados utilizando o algoritmo FROLS e apenas alguns regressores corretos são selecionados por este algoritmo, resultando em problemas na seleção da estrutura (Falsone *et al.*, 2015). No que se segue, os

algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI são aplicados nestes sistemas com o objetivo de testar sua capacidade de identificar a estrutura correta dos modelos.

### 5.1.2 Parâmetros para o caso SISO

Há muitos parâmetros que precisam ser definidos nos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI. Alguns parâmetros são fixos tais como  $K$  no índice de desempenho  $\mathcal{J}_p$  o qual é definido como 1, sendo o índice puramente preditivo ( $\beta = 0$ ). O limiar  $\delta$  no algoritmo ERR foi definido como 0,1% para a inclusão dos regressores, isto é, o regressor é incluído no modelo se o mesmo melhorar o MSPE em pelo menos 0,1%. Para os valores iniciais de RIP, duas estratégias foram adotadas, na maioria dos testes o valor de  $\mu_j = 1/m$ , em que  $m$  é o tamanho do conjunto de regressores candidatos, foi usado. Para sistemas com conjunto de candidatos regressores muito grande ( $n_u = n_y > 10$ ), ver seção 5.1.7, os valores iniciais da RIP foram definidos como  $\mu_j = 1/Z_M$ , em que  $Z_M$  é o tamanho da população, com o objetivo de garantir uma exploração do espaço de busca mais efetiva (Avellina *et al.*, 2017). Finalmente, os limites inferior ( $\mu_{lower}$ ) e superior ( $\mu_{upper}$ ) foram definidos respectivamente como  $1/m$  e  $1 - 1/m$ , a fim de evitar uma exclusão total dos regressores. Os parâmetros restantes são especificados nos respectivos testes.

### 5.1.3 Sensibilidade dos algoritmos aleatórios ao uso da taxa de redução do erro (ERR)

O uso da taxa de redução do erro (ERR) como um filtro para a seleção dos regressores nos algoritmos aleatórios é uma das principais características deste trabalho.

Para testar os efeitos do ERR nos algoritmos aleatórios, um teste usando RaMSS, RaMSS-ERR e RaMSS-EGI foi feito usando o modelo  $\mathcal{S}_1$  retirado de Bonin *et al.* (Bonin *et al.*, 2010), o qual é dado na Tabela (5.1). Os parâmetros usados foram o tamanho da população  $Z_M = 100$ , atrasos das entradas e saídas  $n_y = n_u = 8$ , grau de não linearidade  $l = 3$ , resultando em 969 regressores candidatos, isto é  $m = 969$ . Por fim foi utilizado uma RIP inicial de  $\mu_j = 1/m$ .

O teste consistiu em 50 simulações de Monte Carlo com um número fixo de 100 iterações. A Tabela (5.2) mostra a taxa de convergência  $CR$ , calculada como a fração do nú-

Tabela 5.2: Influência do ERR nos algoritmos aleatórios com diferentes valores iniciais de RIP

			ERR	CR	A	IT	ET	EM
RaMSS	$1/R_m$	$\mathcal{S}_1$	0	1,00	1,00	39,15	25,43	3240
			1	1,00	1,00	37,21	18,52	2095
RaMSS-EGI	$1/R_m$	$\mathcal{S}_1$	0	1,00	1,00	30,73	29,51	789
			1	1,00	1,00	19,47	19,51	674

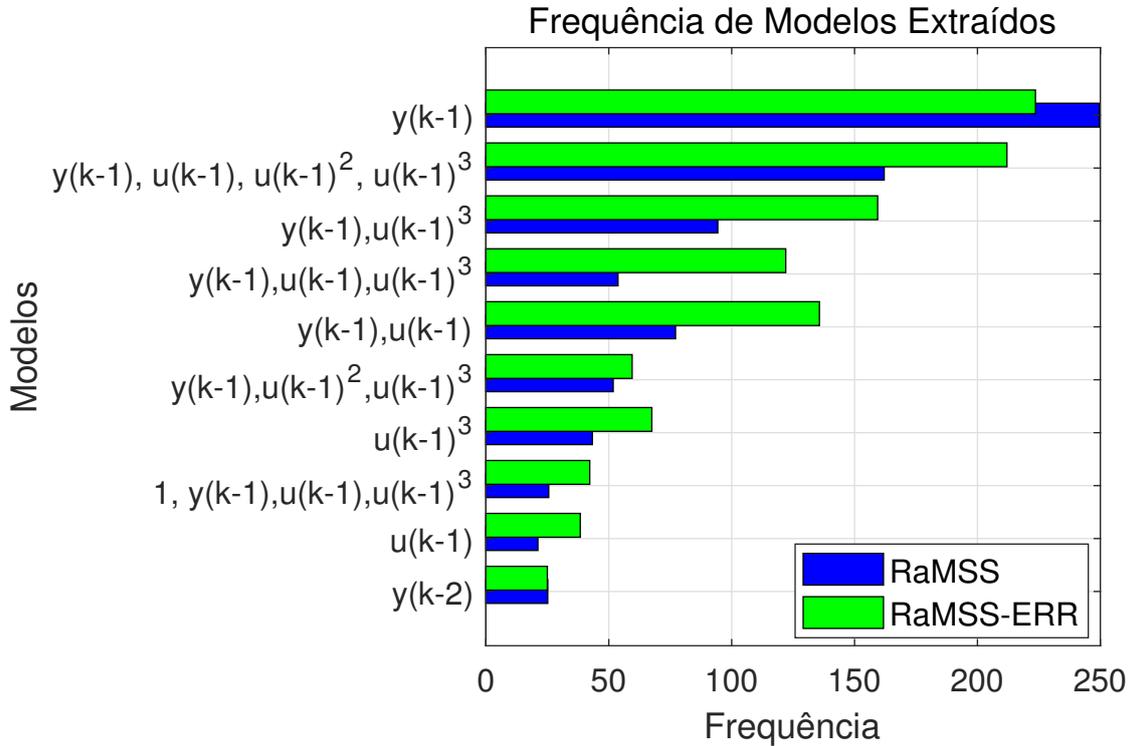
mero de simulações que convergiram antes do algoritmo atingir o número máximo de iterações, a precisão  $A$ , calculada aqui como a fração de simulações que resultaram na seleção correta da estrutura do modelo,  $IT$  é o número de iterações até o algoritmo atingir a convergência, o tempo decorrido  $ET$  que é o tempo necessário para que a distribuição de probabilidade esteja concentrada nos regressores verdadeiros, e  $EM$ , que é o número de diferentes modelos explorados. A convergência da distribuição de probabilidade  $\mathcal{P}_\phi$  foi considerada quando o algoritmo atingiu o critério de parada. Na coluna do  $ERR$  0 indica o algoritmo rodando sem ERR e 1 algoritmo rodando com o ERR. Repare que o algoritmo RaMSS usando o ERR é o algoritmo proposto RaMSS-ERR.

Os resultados apresentados na Tabela (5.2) mostram que o uso do ERR nos algoritmos aleatórios foi capaz de diminuir o número de iterações e o número total de modelos explorados. Para o algoritmo RaMSS, o número de iterações é levemente diminuído, enquanto o tempo decorrido  $ET$  é reduzido em 27% e o número total de modelos explorados é reduzido em 35%. Já no algoritmo RaMSS-EGI, todos os índices diminuíram com o uso do ERR, o número de iterações foi reduzido por 36%, o tempo decorrido para a convergência em 33% e o número total de modelos explorados diminuiu aproximadamente em 15%.

Essa melhora resultante do uso do ERR pode ser explicada pela frequência em que os modelos que contribuem para o aumento da RIP são extraídos. A Figura (5.1) mostra o número médio dos dez modelos mais extraídos durante 50 simulações de Monte Carlo.

Como pode ser visto, exceto pelo modelo formado apenas por  $y(k-1)$ , o qual é frequentemente extraído, o algoritmo RaMSS-ERR extraiu modelos mais completos em uma frequência maior do que o RaMSS, o que acelera o processo de convergência já que quanto mais completo é o modelo, maior é o seu impacto na convergência da RIP.

Figura 5.1: Frequência de modelos extraídos pelos algoritmos RaMSS e RaMSS-ERR



#### 5.1.4 Tamanho da População

O tamanho da população tem uma alta influência no tempo de processamento. Uma população grande usualmente aumenta o custo computacional e consequentemente o tempo de convergência (Abbas e Bayoumi, 2004), por outro lado, uma população muito pequena pode tornar o processo de convergência muito lento ou mesmo impedir que esta aconteça. Desta forma, foi necessário estabelecer um valor padrão para o número de indivíduos nos algoritmos propostos.

Neste estudo, para analisar os efeitos do tamanho da população nos métodos propostos, 50 simulações de Monte Carlo foram feitas utilizando o sistema  $\mathcal{S}_1$ , variando o tamanho da população  $Z_M$  de 5 até 100, para cinco diferentes tamanhos de conjuntos de regressores candidatos  $m = [165, 455, 969, 1771, 2925]$  e usando dois valores iniciais de RIP,  $\mu_j = 1/m$  e  $\mu_j = 10/m$ . O número máximo de iterações foi definido como 200, nas simulações em que o algoritmo atingiu esse número foi considerado que o algoritmo não convergiu.

A Figura (5.2) mostra o tempo decorrido para convergência e a precisão, a qual é a porcentagem de simulações em que o algoritmo selecionou a estrutura correta do modelo.

Analisando as Figuras (5.2a), (5.2b), (5.2c) e (5.2d), é possível notar que para ambos os métodos, o tempo decorrido para convergir é mais influenciado pelo tamanho do conjunto de regressores candidatos  $m$  do que pelo tamanho da população  $Z_M$ , até mesmo para valores altos de RIP inicial. Entretanto, observando as Figuras (5.2e), (5.2f), (5.2g) e (5.2h), é possível notar que o tamanho da população tem grande influência na precisão do algoritmo. Para ambos os métodos um aumento no tamanho da população produziu um aumento na precisão. Usando o RaMSS-EGI e  $m = 2925$ , a precisão com uma população de 100 é 98%, enquanto com uma população de 5 a precisão é de 18%.

Outro ponto interessante é que aumentando os valores da RIP inicial, também aumenta a precisão do algoritmo. No pior cenário de  $m = 2925$  usando o algoritmo RaMSS-EGI e  $\mu_j = 10/m$ , uma precisão de 100% é atingida utilizando  $Z_M = 60$  enquanto para o RaMSS-ERR a precisão usando  $Z_M = 100$  é aumentada de 80% para 98%. Essa melhora pode ser explicada pelo fato de que uma RIP inicial maior permite uma exploração do espaço de busca mais efetiva do que com valores pequenos.

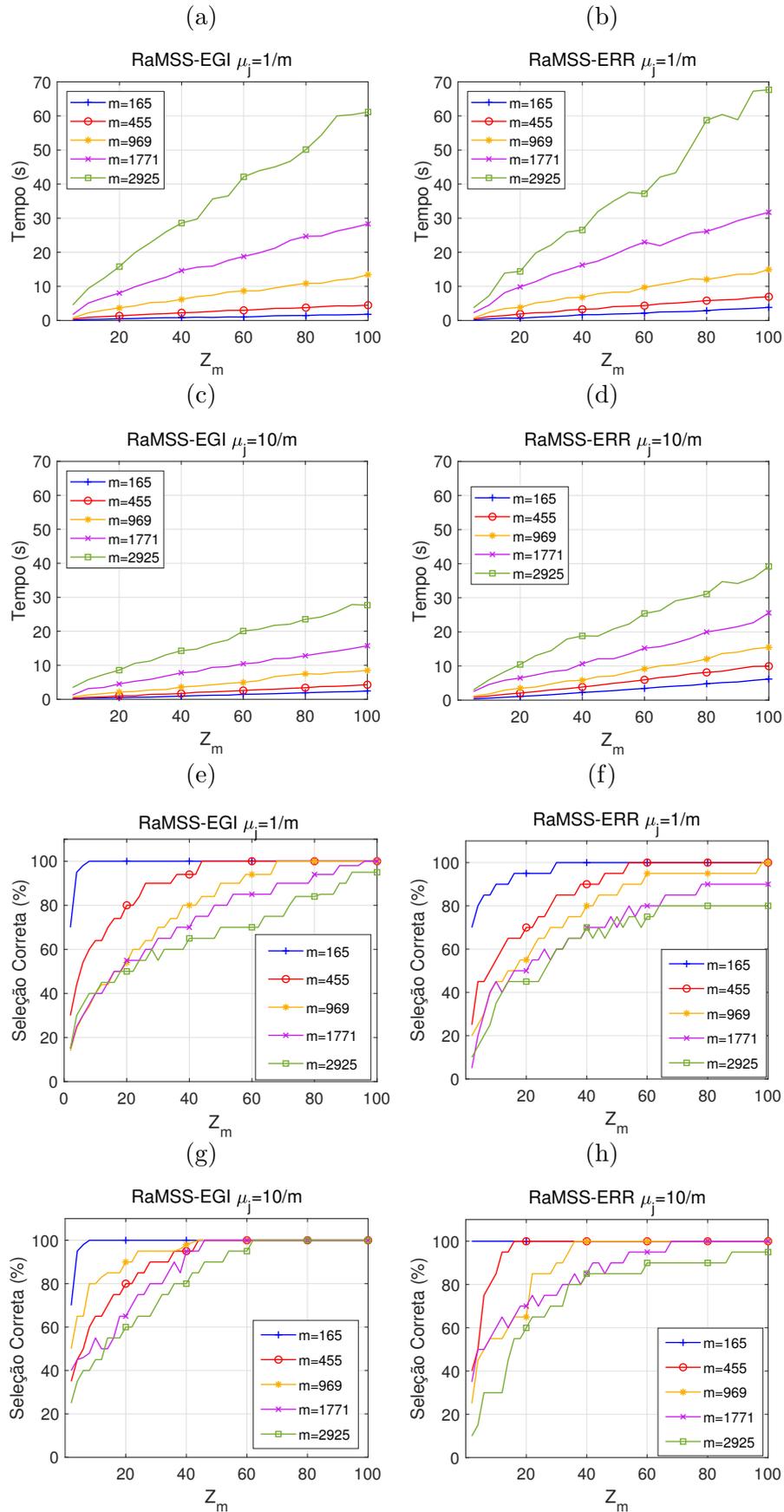
A análise do tamanho da população mostra que há um balanceamento entre a precisão e o tempo de convergência, isto é, para aumentar a precisão é necessário aumentar o tamanho da população, o que acaba por aumentar também o tempo necessário para se atingir a convergência. Para a maioria dos cenários uma população de 100 indivíduos apresentou uma boa precisão, 80% para o pior caso, e um tempo de convergência razoável, 70 segundos para o pior caso, assim a menos que seja especificado outro valor, o tamanho de 100 indivíduos será utilizado como o tamanho da população padrão para ambos os métodos, em todos os casos de estudos neste trabalho.

#### 5.1.5 Análise e testes comparativos para o caso SISO

Com o objetivo de medir o desempenho dos métodos propostos RaMSS-ERR e RaMSS-EGI, cinco sistemas da literatura são identificados e comparados com o RaMSS. Os testes nesta seção foram similares aos apresentados por Falsone *et al.* (2015).

Um total de 100 simulações de Monte Carlo foram realizadas para cada sistema na Tabela (5.1), usando os mesmos 500 dados de entrada e saída de cada sistema. Os

Figura 5.2: Tempo decorrido para convergir e seleção correta em função do tamanho da população, (a), (c), (e) (g) para RaMSS-EGI, (b), (d), (f) (h) para RaMSS-ERR.



parâmetros usados nos testes foram  $Z_M = 100$ ,  $n_u = n_y = 8$  e grau de não linearidade  $l = 3$ , resultando em  $m = 969$ . Além disso, os testes foram feitos utilizando dois valores iniciais de RIP,  $\mu_j = 1/m$  e  $\mu_j = 10/m$ , sendo o primeiro valor o mesmo usado em Falsone *et al.* (2015), e o último escolhido considerando que um valor inicial de RIP maior aumenta a capacidade exploratória do algoritmo melhorando a precisão (Avellina *et al.*, 2017). Os outros parâmetros são os mesmos especificados na seção 5.2.2. Os testes foram feitos usando os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI, e são sumarizados na Tabela (5.3).

Além dos índices introduzidos na seção 5.1.3, dois índices relacionados ao tamanho dos modelos são incluídos, o *M.AMS* que é o tamanho médio máximo dos modelos e indica o tamanho médio máximo dos modelos explorados pelos algoritmos, e o *F.AMS* que representa o tamanho médio final dos modelos, indicando o tamanho médio que os modelos extraídos pelo algoritmo convergiram.

Analisando a Tabela (5.3) é possível notar que os algoritmos RaMSS-EGI e RaMSS-ERR superaram o RaMSS clássico em termos de iterações, tempo decorrido até a convergência e número de modelos explorados nos testes usando  $\mu_j = 1/m$ . Para  $\mu_j = 10/m$ , o RaMSS e RaMSS-EGI tiveram uma melhora em todos os índices, entretanto, o mesmo não é verdade para o RaMSS-ERR. Exceto pelo sistema  $\mathcal{S}_4$ , quando  $\mu_j$  é definido como  $10/m$ , o tempo decorrido para convergir para o RaMSS-ERR é aumentado. Este aumento no tempo decorrido até a convergência é devido ao processamento extra que é requerido pelo algoritmo ERR. Com base nestes testes, pode ser observado que para um pequeno grupo de regressores candidatos, em termos de número de iteração e tempo decorrido até a convergência, o RaMSS-ERR tem um desempenho similar ao RaMSS. No entanto, para um conjunto de regressores candidatos com muitos termos, como será mostrado na seção 5.1.7.2, o uso do RaMSS-ERR mostra uma melhora considerável quando comparado com o RaMSS.

Além disso, uma desvantagem do método proposto RaMSS-EGI quando comparado ao RaMSS clássico é a sua precisão. Para o sistema  $\mathcal{S}_3$ , é possível observar que a precisão é de 95% usando o RaMSS-EGI com  $\mu_j = 10/m$ , enquanto que para o RaMSS clássico a precisão é de 100% para o mesmo valor de  $\mu_j$ . Esta diminuição na precisão do algoritmo é devido ao fato de que o algoritmo mantém na memória os melhores modelos, os quais por vezes podem ser

Tabela 5.3: Análise comparativa dos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI.

Método	$\mu_j$	Sistema	CR	A	IT	ET	M. AMS	F. AMS	EM
RaMSS	1/m	$\mathcal{S}_1$	1,00	1,00	39,15	19,34	4,81	4,04	1994
		$\mathcal{S}_2$	1,00	1,00	44,00	25,96	5,12	4,05	2932
		$\mathcal{S}_3$	1,00	0,88	54,18	32,26	4,27	5,00	4179
		$\mathcal{S}_4$	1,00	1,00	68,05	36,52	5,10	4,00	2533
		$\mathcal{S}_5$	1,00	1,00	40,89	21,96	4,31	4,00	1500
	10/m	$\mathcal{S}_1$	1,00	1,00	27,73	18,97	5,15	4,10	1235
		$\mathcal{S}_2$	1,00	1,00	26,47	21,46	10,21	4,20	2185
		$\mathcal{S}_3$	1,00	1,00	48,00	28,22	8,00	5,18	3670
		$\mathcal{S}_4$	1,00	1,00	40,73	31,42	5,73	4,30	1498
		$\mathcal{S}_5$	1,00	1,00	25,15	18,49	5,21	4,05	774
RaMSS-ERR	1/m	$\mathcal{S}_1$	1,00	1,00	37,21	18,52	4,00	4,00	1479
		$\mathcal{S}_2$	1,00	1,00	39,26	22,57	4,36	4,01	1642
		$\mathcal{S}_3$	1,00	1,00	43,73	31,15	5,26	4,99	2733
		$\mathcal{S}_4$	1,00	1,00	64,15	35,62	4,31	4,00	2198
		$\mathcal{S}_5$	1,00	1,00	38,52	18,22	4,42	4,00	1463
	10/m	$\mathcal{S}_1$	1,00	1,00	29,68	20,47	5,60	4,10	981
		$\mathcal{S}_2$	1,00	1,00	35,15	27,02	6,73	3,98	1423
		$\mathcal{S}_3$	1,00	1,00	41,68	39,27	5,84	4,97	1711
		$\mathcal{S}_4$	1,00	1,00	32,73	26,89	5,00	4,30	1278
		$\mathcal{S}_5$	1,00	1,00	25,05	25,71	4,55	4,20	841
RaMSS-EGI	1/m	$\mathcal{S}_1$	1,00	1,00	19,47	18,51	5,05	4,00	674
		$\mathcal{S}_2$	1,00	1,00	20,31	22,75	5,78	4,02	1031
		$\mathcal{S}_3$	1,00	0,90	41,84	30,81	6,26	5,02	921
		$\mathcal{S}_4$	1,00	1,00	22,73	28,33	5,57	4,10	638
		$\mathcal{S}_5$	1,00	1,00	18,52	20,88	4,57	4,05	665
	10/m	$\mathcal{S}_1$	1,00	1,00	10,47	12,53	6,20	4,16	560
		$\mathcal{S}_2$	1,00	1,00	12,78	16,17	7,21	4,25	642
		$\mathcal{S}_3$	1,00	0,95	14,14	19,89	6,15	5,12	838
		$\mathcal{S}_4$	1,00	1,00	11,15	20,47	5,78	4,00	813
		$\mathcal{S}_5$	1,00	1,00	9,47	12,03	6,10	4,00	515

modelos sobreparametrizados. Nestes casos o algoritmo tenta modelar o ruído, o que acaba por levar o algoritmo a convergir para modelos incorretos.

Embora o RaMSS-EGI seja mais rápido, convergindo em um número reduzido de iterações e explorando uma quantidade menor de modelos, o algoritmo RaMSS-ERR tem uma melhor precisão que o RaMSS clássico e sua implementação é mais simples quando comparado com o RaMSS-EGI.

### 5.1.6 Sistemas com entrada pouco excitativa

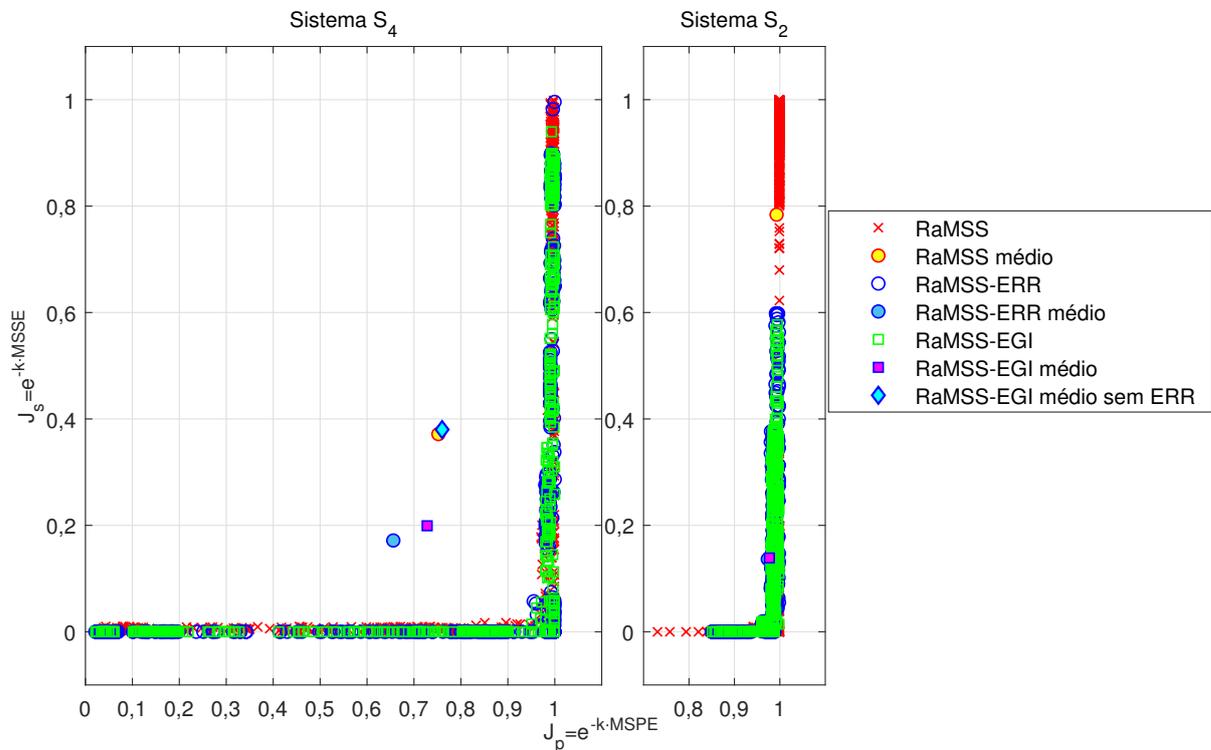
Na presente discussão os sistemas  $\mathcal{S}_2$  e  $\mathcal{S}_4$  são usados para verificar o desempenho dos métodos propostos quando sujeitos a sinais pouco excitativos. O teste consiste em 500 simulações de Monte Carlo limitadas em 100 iterações, aplicados a cada sistema, com os mesmos parâmetros usados na seção 5.1.5. Um conjunto de 500 pares de entradas e saídas foram geradas a partir de cada sistema para cada simulação de Monte Carlo. Os sistemas são excitados por um ruído branco filtrado por um filtro passa baixa dado pela Equação (5.1) (Falsone *et al.*, 2015).

$$u(k) = 1,85u(k-1) - 0,855u(k-2) + \xi(k) \quad (5.1)$$

com  $\xi(k) \sim WGN(0; 2,28 \cdot 10^{-2})$ .

A Figura (5.3) mostra o valor do desempenho de cada modelo identificado pelos três métodos. Para ambos os modelos, o algoritmo RaMSS clássico teve um desempenho melhor do que os métodos propostos. Isto pode ser explicado pelo fato de que o sinal pouco excitativo produz também uma saída pouco excitada, resultando em uma diferença muito pequena entre sinais adjacentes na saída, o que faz com que vários regressores pareçam similares. Sob estas condições, o ERR dá uma maior importância aos regressores denominados “preditores triviais” como  $y(k-1)$  e  $y(k-2)$ , tendo esses altos valores de ERR, enquanto outros regressores recebem baixos valores de ERR. Como os métodos propostos utilizam o ERR como um filtro para a inclusão dos regressores, aqueles com baixo ERR e que muitas vezes são importantes para descrever a dinâmica do sistema em questão, são excluídos. No presente teste, na maioria das vezes, os modelos selecionados eram formados usualmente de  $y(k-1)$  para o sistema  $\mathcal{S}_2$

Figura 5.3: Comparação dos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI nos sistemas  $\mathcal{S}_4$  e  $\mathcal{S}_2$  com sinal de entrada pouco excitativo.



e  $y(k-2)$  para o sistema  $\mathcal{S}_4$ .

Um teste usando o algoritmo RaMSS-EGI sem o filtro ERR foi feito para o sistema  $\mathcal{S}_4$  nas mesmas condições, o valor médio para o MSPE e MSSE é mostrado como um diamante azul na Figura (5.3). Como pode ser observado, o RaMSS-EGI sem o filtro ERR teve um desempenho similar ao RaMSS clássico, mostrando que os métodos propostos são afetados pela deficiência do ERR em lidar com sinais pouco excitativos.

### 5.1.7 Identificação de sistemas com alto grau de não-linearidade

Nesta seção dois tipos de sistemas são considerados: o primeiro tipo consiste em dois sistemas com resposta inversa (um sistema linear e um não linear), e o segundo tipo consiste em um sistema artificial com alto grau de não linearidade. Estes sistemas requerem um grande conjunto de regressores candidatos, já que apresentam longos atrasos e alto grau de não linearidade, respectivamente.

Os sistemas são identificados usando os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI. Para comparar os métodos, 50 simulações de Monte Carlo foram feitas com o número

máximo de iterações fixados em 500. Alguns parâmetros são descritos nas respectivas seções e os outros são definidos de acordo com a seção 5.2.2.

#### 5.1.7.1 Sistemas com Resposta Inversa

Na indústria, sistemas com resposta inversa são muito comuns estando presente em processos químicos como boilers e reatores químicos (Marlin, 2000).

De acordo com Balaguer *et al.* (2011), sistemas com resposta inversa impõem sérios desafios para o projeto de controladores por ter algumas limitações nos níveis de desempenho e robustez. Portanto, este tipo de sistema usualmente requer melhores técnicas de controle, e consequentemente é necessário conhecer o modelo do sistema.

No que se segue, os parâmetros foram definidos como  $n_y = n_u = 12$  e o grau de não linearidade como  $l = 3$ , resultando em  $m = 2925$ ,  $\mu_j = 1/m$ .

Os seguintes sistemas artificiais foram considerados

$$\begin{aligned} \mathcal{S}_6 : y(k) = & 1,646y(k-1) - 0,670y(k-2) - 0,293u(k-11) + \\ & + 0,359u(k-12) + e(k) \end{aligned} \quad (5.2)$$

com  $e(k) \sim WGN(0; 0,004)$  e  $u(k) \sim WGN(0; 1)$ .

$$\begin{aligned} \mathcal{S}_7 : y(k) = & 1,641y(k-1) - 0,665y(k-2) + 0,051u(k-10) \\ & - 0,182u(k-12)y(k-3) - 0,291u(k-11) \\ & + 0,360u(k-12) + e(k) \end{aligned} \quad (5.3)$$

com  $e(k) \sim WGN(0; 0,004)$  e  $u(k) \sim WGN(0; 1)$ .

A Tabela (5.4) mostra os resultados médios dos testes realizados com os sistemas  $\mathcal{S}_6$  e  $\mathcal{S}_7$  com a população de 100 e de 500 para os critérios de taxa de convergência, precisão na seleção do modelo, número de iterações, tempo até a convergência e número de modelos explorados.

Como pode ser visto na Tabela (5.4), para o sistema com atrasos longos, os algoritmos propostos RaMSS-ERR e RaMSS-EGI tiveram uma melhor precisão do que o RaMSS clássico. Para uma população  $Z_M = 100$ , a precisão para o RaMSS foi de 87% e 67% para

Tabela 5.4: Resultados médios para os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI aplicados aos sistemas  $\mathcal{S}_6$  e  $\mathcal{S}_7$ .

Sistema	Método	$Z_M$	CR	A	IT	ET	EM
$\mathcal{S}_6$	RaMSS	100	1,00	0,87	28,03	51,57	1913
		500	1,00	1,00	20,04	81,27	9086
	RaMSS-ERR	100	1,00	1,00	20,43	43,06	1413
		500	1,00	1,00	17,87	87,18	7997
	RaMSS-EGI	100	1,00	1,00	9,33	20,02	771
		500	1,00	1,00	8,14	29,33	3238
$\mathcal{S}_7$	RaMSS	100	1,00	0,67	85,23	102,11	4554
		500	1,00	1,00	75,12	167,14	24412
	RaMSS-ERR	100	1,00	1,00	75,24	97,99	1527
		500	1,00	1,00	65,45	135,15	7439
	RaMSS-EGI	100	1,00	1,00	14,15	33,65	1018
		500	1,00	1,00	12,10	55,51	3507

os sistemas  $\mathcal{S}_6$  e  $\mathcal{S}_7$  respectivamente, enquanto para os métodos propostos a precisão foi de 100%. Aumentando o tamanho da população para  $Z_M = 500$ , todos os métodos atingiram uma precisão de 100% ao custo de um maior número de iterações, tempo decorrido e modelos explorados. Apesar da redução no número de iterações para os métodos propostos, aumentar o tamanho da população não parece ter nenhum benefício, já que o tempo decorrido até a convergência e o número de modelos explorados também aumentaram.

Em termos de iteração e modelos explorados, os métodos propostos mostraram uma melhora quando comparados com o RaMSS clássico. No melhor caso, o número de iterações para o RaMSS-EGI foi menor que 20% do total de iterações para o RaMSS (sistema  $\mathcal{S}_7$  com  $Z_M = 500$ ). Considerando o número de modelos explorados, o RaMSS-EGI explorou cerca de 15% do número total de modelos explorados pelo RaMSS, e metade do RaMSS-ERR (sistema  $\mathcal{S}_7$  com  $Z_M = 500$ ).

De uma forma geral, os métodos propostos também tiveram uma melhora no tempo decorrido quando comparados com o RaMSS clássico, para o melhor caso (sistema  $\mathcal{S}_7$  com  $Z_M = 500$ ), o tempo decorrido até a convergência para o RaMSS-EGI foi em torno de 30% do tempo do RaMSS e 41% do tempo do RaMSS-ERR. Todos estes benefícios mostram claramente a vantagem dos métodos propostos quando comparados com o RaMSS clássico.

### 5.1.7.2 Modelo Artificial com alto grau de não linearidade

Nesta seção, os três métodos são aplicados em um problema de seleção da estrutura com um grande conjunto de regressores candidatos. O modelo considerado  $\mathcal{S}_8$  foi retirado de Avellina *et al.* (2017) e tem um grau de não linearidade  $l = 6$ . O modelo é dado pela Equação (5.4).

$$\mathcal{S}_8 : y(k) = 0,5y(k-1) + 0,8u(k-2)^3 + u(k-2)^6 - 0,05y(k-2)^2 + 0,5 + e(k) \quad (5.4)$$

com  $u(k) = \text{sign}(w(k)) \cdot |w(k)|^{1/3}$ ,  $w(k) \sim WGN(0; 0, 3)$ ,  $e(k) \sim WGN(0; 0, 01)$ .

O teste para o sistema  $\mathcal{S}_8$  consistiu em 50 simulações de Monte Carlo para cada um dos três métodos e para dois níveis diferentes de tamanho da população. Um total de 1000 pares de entrada e saída foram gerados. Os parâmetros usados pelo sistema  $\mathcal{S}_8$  foram  $n_u = n_y = 6$  e grau de não linearidade  $l = 6$ , resultando em  $m = 18564$ . Como um valor de RIP inicial de  $1/m$  se mostrou muito insuficiente para explorar um espaço de busca tão grande,  $\mu_j$  foi definido como  $1/Z_M$ , permitindo assim uma exploração do espaço de busca mais eficiente.

Para este caso, o algoritmo RaMSS teve um baixo desempenho, com uma precisão de 40% e 52% para  $Z_M = 100$  e  $Z_M = 500$  respectivamente. Além disso, o número de iterações e o tempo decorrido foram bem maiores, para  $Z_M = 500$  o algoritmo levou em média 16 minutos para convergir. O algoritmo RaMSS-ERR usando uma população de  $Z_M = 100$ , teve um desempenho similar ao RaMSS clássico, embora o número de modelos explorados tenha sido menor. Quando o tamanho da população foi aumentado para  $Z_M = 500$ , o algoritmo RaMSS-ERR aumentou sua precisão para 85%, houve também uma melhora no número de iterações e tempo decorrido até a convergência.

Neste problema com um grande conjunto de regressores candidatos o algoritmo RaMSS-EGI se mostrou muito eficiente. A precisão para ambas as populações foi de 100%, o número de iterações e tempo decorrido até a convergência foi bem menor que os competidores. O tempo médio de convergência foi em média 2 minutos, e o número de modelos explorados foi menor que 5% do número de modelos explorados pelo algoritmo RaMSS clássico.

Tabela 5.5: Resultados médios para os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI aplicados ao sistema  $\mathcal{S}_8$ .

System	Method	$Z_M$	CR	A	IT	ET	EM
$\mathcal{S}_8$	RaMSS	100	1,00	0,40	211,44	212,25	20130
		500	1,00	0,52	215,83	997,15	102500
	RaMSS-ERR	100	1,00	0,41	190,11	385,12	4190
		500	1,00	0,85	68,13	92,23	8345
	RaMSS-EGI	100	1,00	1,00	59,21	94,28	1570
		500	1,00	1,00	65,34	105,38	4573

### 5.1.8 Identificação do Reator Tanque Agitado

O Reator Tanque Agitado (*Continuous Stirred Tank Reactor - CSTR*) é um equipamento industrial muito comum usado primariamente nas reações em fase líquida (Fogler, 2016). Neste exemplo é considerado um CSTR similar ao apresentado por Balaguer *et al.* (2011), o reator é isotérmico, com as seguintes reações de van der Vusse.



O reator pode ser visto na Figura (5.4).

Supondo uma mistura perfeita, i.e., não existe dependência entre a temperatura, concentração e taxa de reação com a posição dentro do CSTR, e aplicando um balanço de massa, o sistema pode ser descrito pelas Equações (5.6a) e (5.6b),

$$\frac{dC_A(t)}{dt} = \frac{F_r(t)}{V} [C_{Ai} - C_A(t)] - k_1 C_A(t) - k_3 C_A^2(t) \quad (5.6a)$$

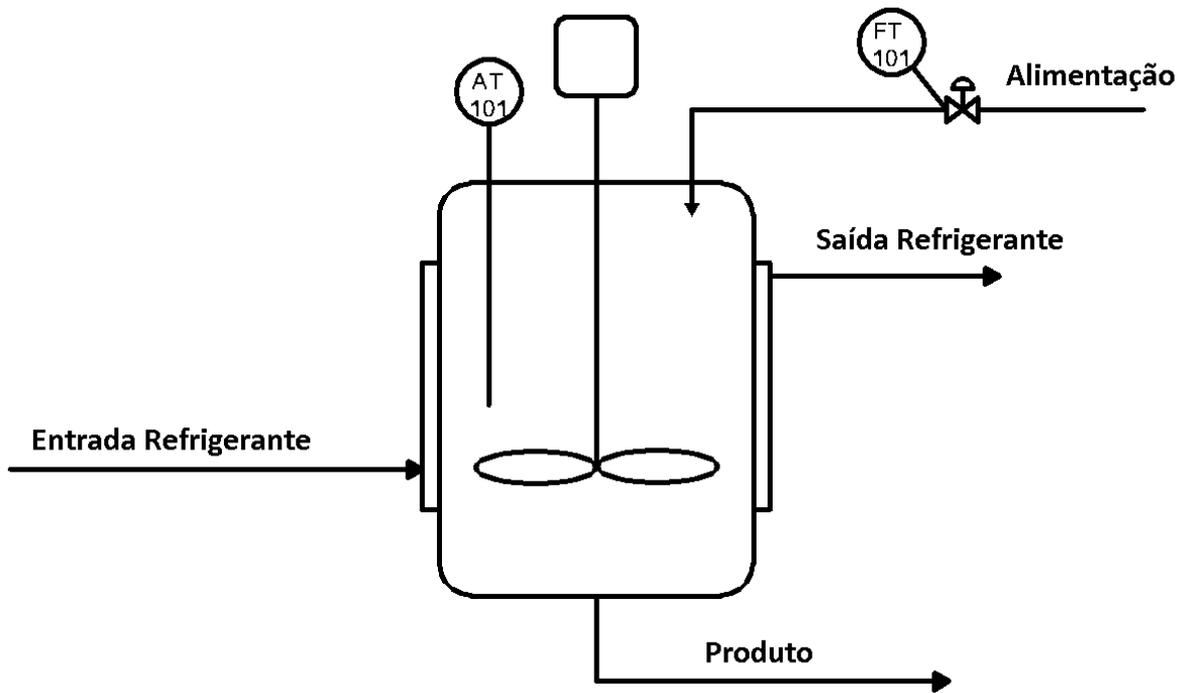
$$\frac{dC_B(t)}{dt} = -\frac{F_r(t)}{V} C_B(t) + k_1 C_A(t) - k_2 C_B(t) \quad (5.6b)$$

em que  $F_r$  é a taxa de alimentação do produto,  $V$  é o volume do reator,  $C_{Ai}$  é a concentração nominal de  $A$  na alimentação,  $C_A$  e  $C_B$  são as concentrações no reator, e  $k_1$ ,  $k_2$  e  $k_3$  são as constantes de velocidade de reação do processo no CSTR e são mostradas na Tabela (5.6).

O sistema está inicialmente em estado estacionário com  $C_{Ao} = 2,92 \text{ mol l}^{-1}$  e  $C_{Bo} = 1,10 \text{ mol l}^{-1}$ .

Para identificar o modelo, o sistema foi excitado usando um sinal aleatório  $u(k) \sim$

Figura 5.4: Sistema CSTR.



Fonte: Adaptado de Balaguer *et al.*, 2011.

Tabela 5.6: Valores dos parâmetros para o processo no CSTR.

$k_1$	$k_2$	$k_3$	$C_{Ai}$	$V$
$5/6 \text{ min}^{-1}$	$5/3 \text{ min}^{-1}$	$1/6 \text{ l mol}^{-1} \text{ min}^{-1}$	$10 \text{ mol l}^{-1}$	$700 \text{ l}$

$WUN(0;1)$ , o sistema foi amostrado em  $0,166\text{ Hz}$  e 600 amostras foram salvas. A Figura (5.5) mostra a entrada e as duas saídas  $C_A$  e  $C_B$  usadas na identificação.

Ambos os algoritmos RaMSS-ERR e RaMSS-EGI foram aplicados na identificação do CSTR usando os parâmetros da seção 5.2.2, com  $n_u = n_y = 12$  e grau de não linearidade  $l = 3$ . Todos os resultados apresentados foram feitos em simulação livre e são mostrados nas seções 5.1.8.1 e 5.1.8.2.

### 5.1.8.1 Identificação usando RaMSS-ERR

Equações (5.7) e (5.8) mostra os modelos identificados usando o algoritmo RaMSS-ERR, um modelo com 5 regressores foi identificado para a concentração  $C_A$  e um modelo com 8 regressores para a concentração  $C_B$ .

$$C_A(k) = 0,613C_A(k-2) + 0,569C_A(k-3) + 0,011C_A(k-2)C_A(k-1) + \\ - 0,109C_A(k-3)C_A(k-2) + 1,259u(k-1) \quad (5.7)$$

$$C_B(k) = 0,858C_B(k-4) - 0,325C_B(k-4)C_B(k-4) + 0,475u(k-4) + \\ + 0,109u(k-10)C_B(k-8) - 0,175u(k-4)u(k-1) + 0,105C_B(k-5) + \\ - 0,166u(k-4)C_B(k-3)C_B(k-1) + 0,319C_B(k-1) \quad (5.8)$$

Três características dos modelos identificados para o CSTR são mostrados nas Figuras (5.6) e (5.7), o comportamento dinâmico que mostra a capacidade do modelo em representar as dinâmicas do sistema, a resposta ao degrau, a qual mede a resposta em estado estacionário e a polarização.

Na Figura (5.7) pode-se notar que o comportamento dinâmico do sistema é bem representado pelo modelo identificado para a resposta  $C_A$ , o mesmo é verdade para a resposta em degrau. As Figuras (5.6c) e (5.6d) mostram que não há polarização nos modelos.

Para a resposta  $C_B$ , a Figura (5.7) mostra que o modelo identificado é capaz de representar as respostas dinâmicas e em estado estacionário. É possível observar também nas Figuras (5.7a) e (5.7b) que o modelo identificado apresenta uma resposta inversa. Os resíduos na Figura (5.7c) seguem uma distribuição normal, e a primeira bissetriz na Figura (5.7) está

Figura 5.5: Sinais usados na identificação do sistema CSTR. Entrada  $u$ , saídas  $C_A$  e  $C_B$ , respectivamente em a, b e c.

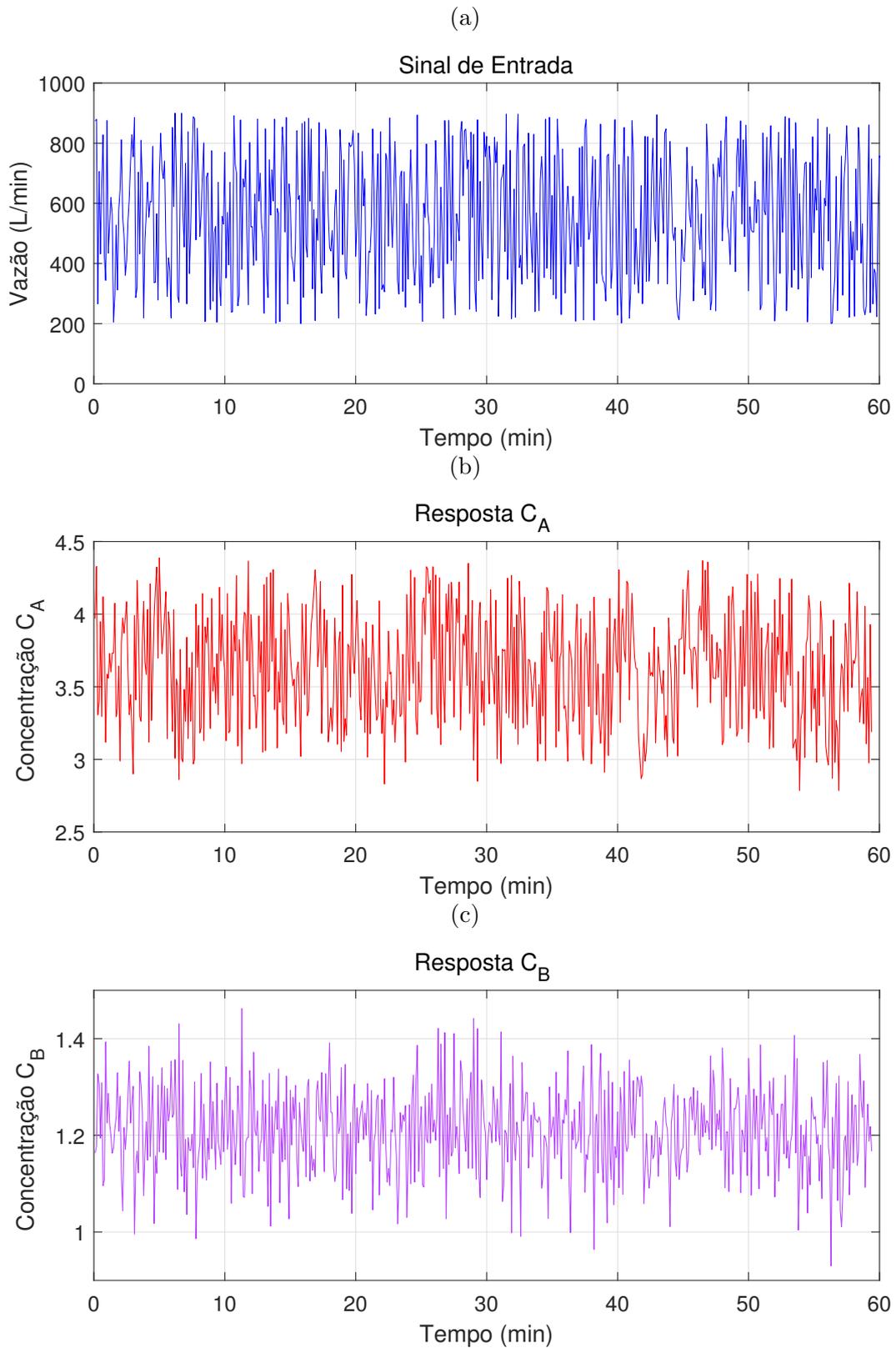


Tabela 5.7: Métricas obtidas na identificação do CSTR usando ambos os métodos RaMSS-ERR and RaMSS-EGI.

Método		MSSE	Iteração	Tempo (s)	Modelos Explorados
RaMSS-ERR	$C_A$	0,9816	32	25	1213
	$C_B$	0,9965	82	48	1934
RaMSS-EGI	$C_A$	0,9811	18	21	872
	$C_B$	0,9960	34	45	953

bem espalhada devido a presença de ruído.

### 5.1.8.2 Identificação usando RaMSS-EGI

O mesmo sistema CSTR foi identificado usando o algoritmo RaMSS-EGI. As Equações (5.9) e (5.10) mostram o modelo identificado.

$$C_A(k) = 0,700 + 0,504C_A(k-2) + 0,112C_A(k-3)C_A(k-1) - 0,020C_A(k-3)C_A(k-2)C_A(k-1) + 1,158u(k-1) \quad (5.9)$$

$$C_B(k) = 1,255C_B(k-4) - 0,502C_B(k-4)C_B(k-4) + 0,641u(k-4) + 0,156u(k-5) + 0,194C_B(k-5)C_B(k-2) - 0,412u(k-4)C_B(k-2) + 0,165u(k-4)u(k-1)C_B(k-5) + 0,065u(k-10)C_B(k-3)C_B(k-3) \quad (5.10)$$

Os mesmos testes feitos para o algoritmo RaMSS-ERR foram também feitos para o RaMSS-EGI, Figuras (5.8) e (5.9) mostram os resultados.

Como pode ser observado, ambos os algoritmos RaMSS-ERR e RaMSS-EGI puderam identificar um modelo apropriado para o CSTR. A Tabela (5.7) mostra algumas métricas obtidas na identificação do CSTR usando os algoritmos propostos. Para este caso, embora o RaMSS-EGI seja mais rápido e tenha explorado um número menor de modelos que o RaMSS-ERR, o MSSE para ambas as saídas é melhor utilizando este último.

### 5.1.9 Discussão sistemas SISO

A partir dos resultados apresentados na seção 5.1 pode-se notar que os algoritmos RaMSS-ERR e RaMSS-EGI tiveram um melhor desempenho que o RaMSS. Esta melhora pode ser explicada considerando dois aspectos distintos, a inclusão da taxa de redução do erro (ERR)

Figura 5.6: Resposta dos modelos identificados usando RaMSS-ERR para a concentração  $C_A$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz  $C_{A_{Real}} \times C_{A_{Estimado}}$  (d).

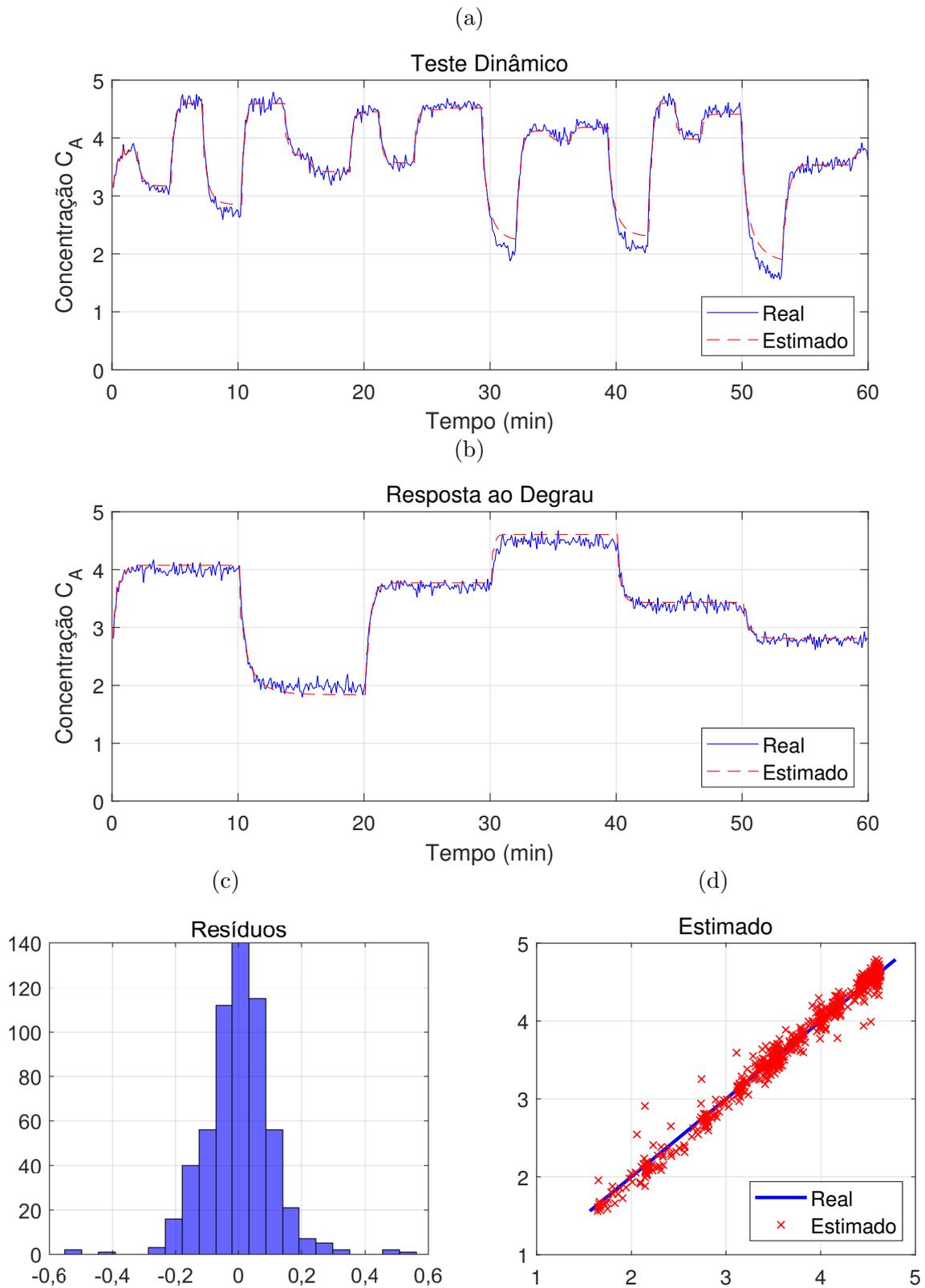


Figura 5.7: Resposta dos modelos identificados usando RaMSS-ERR para a concentração  $C_B$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz  $C_{B_{Real}} \times C_{B_{Estimado}}$  (d).

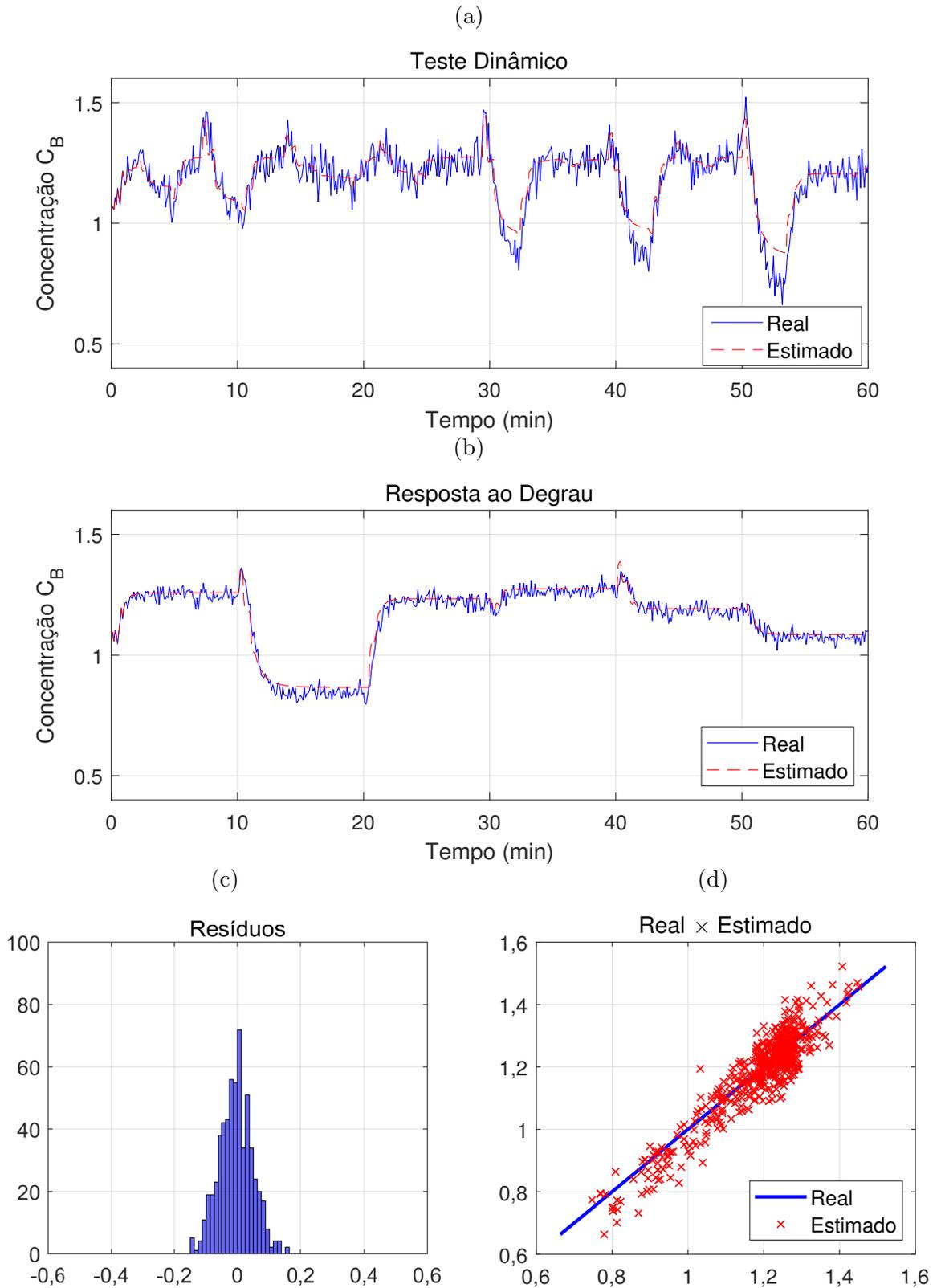


Figura 5.8: Resposta dos modelos identificados usando RaMSS-EGI para a concentração  $C_A$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bissetriz  $C_{A_{Real}} \times C_{A_{Estimated}}$  (d).

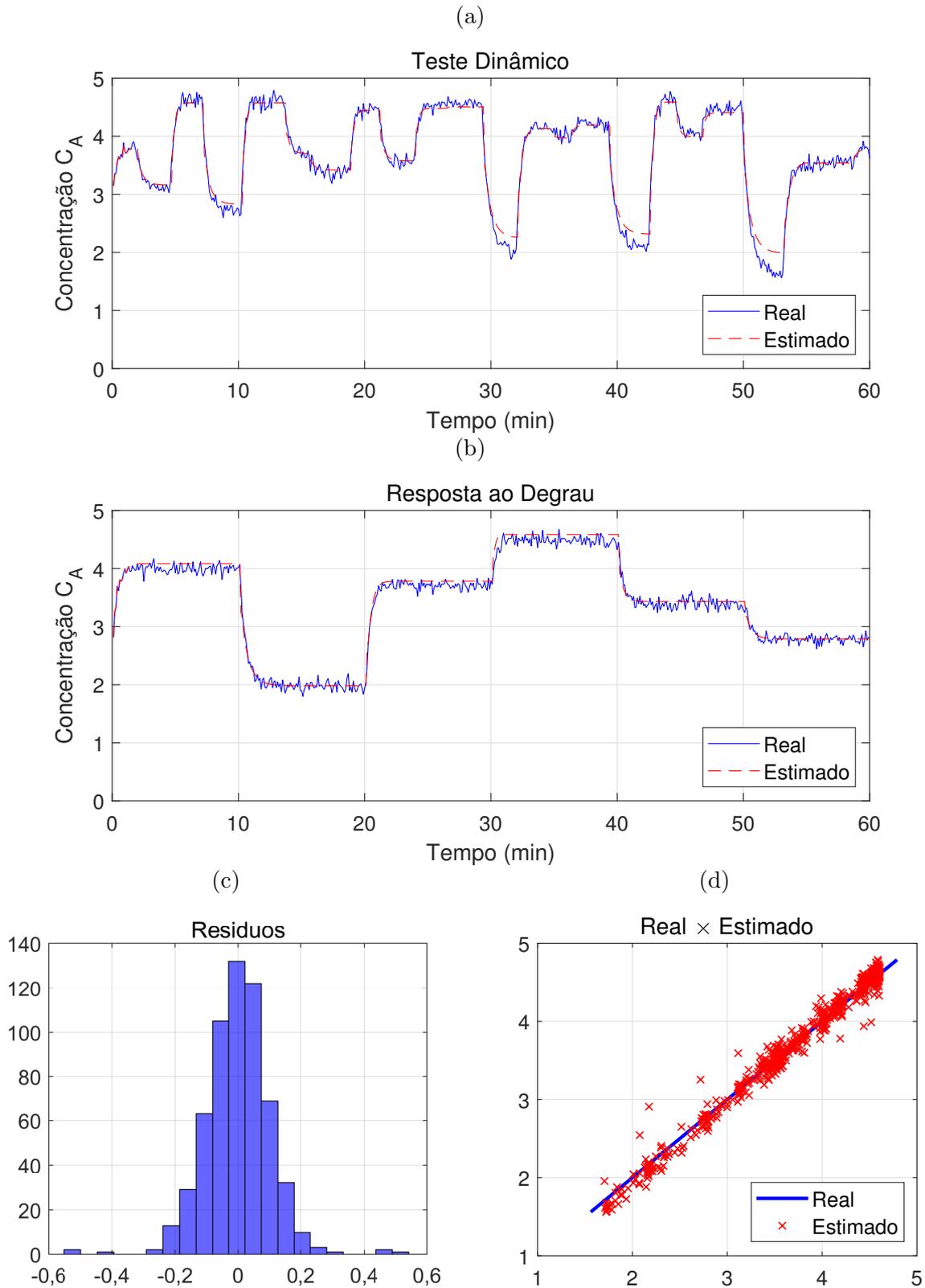
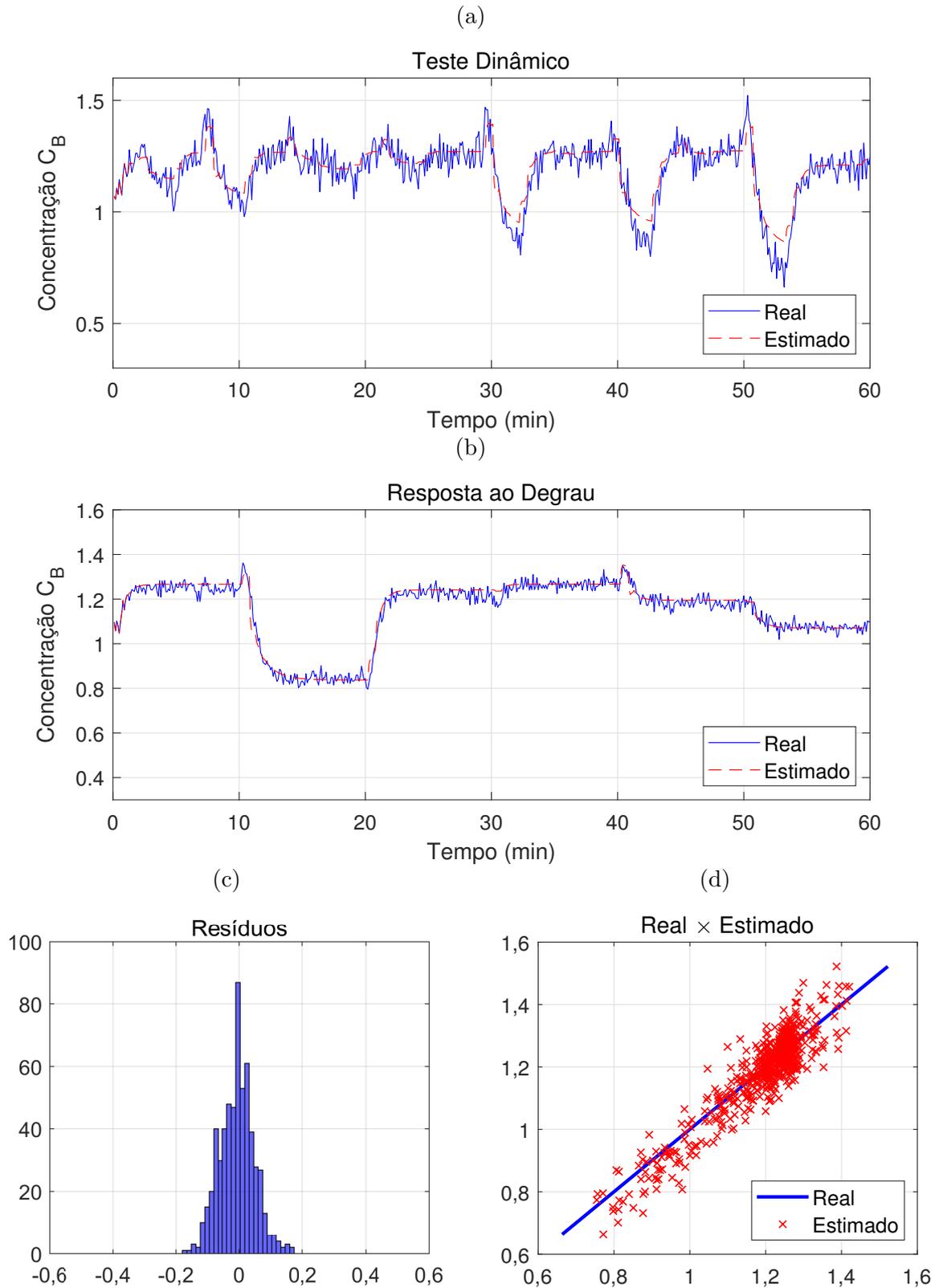


Figura 5.9: Resposta dos modelos identificados usando RaMSS-EGI para a concentração  $C_B$ . Teste dinâmico (a), resposta ao degrau (b), resíduos (c), primeira bisettriz  $C_{B_{Real}} \times C_{B_{Estimated}}$  (d).



no processo e a contribuição de uma segunda população  $M_{best}$  para o processo de atualização da RIP no método RaMSS-EGI.

O ERR age como um pré-filtro no processo de inclusão dos regressores, eliminando aqueles que apresentam uma baixa contribuição para a diminuição da variância na saída.

Como estes regressores são eliminados, é possível notar pela Equação (3.6) que o índice  $\mathcal{I}_j^p$  para estes regressores será maior, assim reduzindo a importância dos regressores e consequentemente irá também diminuir a probabilidade de inclusão do regressor (RIP).

No que concerne à validação da população  $M_{best}$ , como esta população mantém um registro dos melhores modelos que surgem ao longo das iterações, os regressores destes modelos têm suas RIPs aumentadas rapidamente, e os regressores que não pertencem a estes melhores modelos tem suas RIPs diminuídas rapidamente. Essa herança melhora o processo de busca e reduz drasticamente o número de modelos explorados, por outro lado, ela pode também gerar convergências prematuras para modelos sobreparametrizados, o que faz com que o RaMSS-EGI tenha uma maior variância no número de iterações.

Para mitigar esta desvantagem, uma possibilidade é reduzir a contribuição da população  $M_{best}$  na atualização da RIP usando um fator de redução  $\epsilon \in [0, 1]$  como na Equação (5.11).

$$\mu_j(i+1) = \mu_j(i) + \epsilon\gamma\mathcal{I}_j \quad (5.11)$$

A inclusão desse fator diminui a convergência prematura, mas também aumenta o número de modelos explorados, e consequentemente o tempo decorrido até a convergência.

Como pode ser visto na seção 5.1.6, os métodos propostos herdam a deficiência do ERR em lidar com sinais de entrada pouco excitativos. Os métodos propostos apresentam dificuldades em lidar com matrizes mal condicionadas e, como consequência, os preditores triviais como  $y(k-1)$  e  $y(k-2)$  são supervalorizados em termos de ERR, e outros regressores importantes para descrever a dinâmica do processo recebem baixos valores de ERR e são removidos do modelo.

Considerando a aplicação dos métodos propostos no sistema CSTR, pode-se notar que ambos os métodos foram capazes de identificar modelos com uma boa capacidade de

simulação. Embora o modelo identificado pelo RaMSS-ERR tenha um melhor desempenho que o modelo identificado pelo RaMSS-EGI, este último apresenta uma convergência mais rápida e explora um número menor de modelos, como pode ser visto na Tabela (5.7).

## 5.2 Sistemas MISO e MIMO

O objetivo desta seção é validar o desempenho dos algoritmos RaMSS-ERR e RaMSS-EGI no problema da seleção da estrutura de diferentes sistemas MISO e MIMO. Inicialmente são apresentados na seção 5.2.1 alguns modelos de referência retirados da literatura e adaptados para os casos MISO e MIMO, tais sistemas são utilizados para avaliar o desempenho dos algoritmos propostos na identificação de sistemas MISO e MIMO. A seção 5.2.2 mostra os parâmetros usados para o RaMSS, RaMSS-ERR e RaMSS-EGI, estes parâmetros serão usados nos testes das seções que se seguem. O tamanho da população e a probabilidade de inclusão dos regressores são parâmetros que possuem alta influência na velocidade computacional, desta forma, para entender a influência destes parâmetros nos algoritmos propostos, foram feitos alguns testes variando o tamanho da população, com diferentes conjuntos de regressores candidatos, usando dois valores de RIP na seção 5.2.3. Na seção 5.2.4 os métodos propostos são aplicados na identificação de cinco sistemas MIMO, adaptados da literatura, para comparar com o desempenho do RaMSS clássico. Finalmente, a seção 5.2.5 mostra o processo de identificação de uma coluna C3/C4 usando os métodos propostos.

### 5.2.1 Modelos MISO e MIMO utilizados como referência para testes

Alguns sistemas de referência são utilizados nos testes que se seguem. Tais sistemas foram adaptados da literatura para contemplar múltiplas entradas e múltiplas saídas. Estes sistemas contam ainda com diferentes tipos de estruturas e não linearidades, tais como um sistema linear, um sistema com não linearidades na entrada, um com não linearidades na entrada e na saída, um com não linearidades entre entrada e saída e um sistema pertencente à classe NARMAX que apresenta ruído de média móvel.

Os sistemas  $\mathcal{S}_9$ ,  $\mathcal{S}_{10}$ ,  $\mathcal{S}_{11}$ ,  $\mathcal{S}_{14}$  foram adaptados de Wei e Billings (2008), Bonin *et al.* (2010), Mao e Billings (1997), Billings *et al.* (1989), respectivamente, o sistema  $\mathcal{S}_{12}$ , e  $\mathcal{S}_{13}$

Tabela 5.8: Modelos de Benchmark MIMO usados para testar o desempenho dos métodos propostos.

Sistema	Estrutura Conhecida	Entradas( $u_1, u_2$ )	Ruído( $e$ )
$\mathcal{S}_9$	$y(k) = 0, 8y(k-1) + 0, 4u_1(k-1) + 0, 4u_1^2(k-1) + 0, 4u_2(k-1) + 0, 4u_2^2(k-1) + e(k)$	WGN(0;1)	WGN(0; 0, 33 <sup>2</sup> )
$\mathcal{S}_{10}$	$y(k) = -1, 7y(k-1) - 0, 8y(k-2) + u_1(k-1) + 0, 81u_1(k-2) + 0, 3u_2(k-1) + u_2(k-2) + e(k)$	WUN(-2;2)	WGN(0;0, 1 <sup>2</sup> )
$\mathcal{S}_{11}$	$y(k) = 0, 2y^3(k-1) + 0, 6u_1^2(k-2) - 0, 5y(k-2) + -0, 7y(k-2)u_1^2(k-2) + 0, 7y(k-1)u_2(k-1) + e(k)$	WUN(-1;1)	WGN(0;0, 1 <sup>2</sup> )
$\mathcal{S}_{12}$	$y(k) = 0, 7y(k-1)u_1(k-1) - 0, 5y(k-2) + 0, 6u_2^2(k-2) - 0, 7y(k-2)u_2^2(k-2) + e(k)$	WUN(-1;1)	WGN(0;0,004)
$\mathcal{S}_{13}$	$y(k) = 0, 7y(k-1)u_1(k-1) + 0, 5y(k-2) + 0, 6u_2^2(k-2) - 0, 7y(k-2)u_2^2(k-2) + 0, 2e(k-1) - 0, 3u_1(k-1)e(k-2) + e(k)$	WUN(-1;1)	WGN(0;0,02)
$\mathcal{S}_{14a}$	$y_1(k) = 0, 5y_1(k-1) + 0, 1y_2(k-1)u_1(k-1) + u_1(k-2) + e(k)$	WGN(-1;1)	WGN(0;0,04)
$\mathcal{S}_{14b}$	$y_2(k) = 0, 9y_2(k-2) + 0, 2y_2(k-1)u_2(k-2) + u_2(k-1) + e(k)$	WGN(-1;1)	WGN(0;0,04)

de Baldacchino *et al.* (2013). O sistema  $\mathcal{S}_{14}$  é um modelo NARMAX e é usado para testar a capacidade dos algoritmos propostos em identificar a parte NARX do sistema. A Tabela (5.8) mostra os modelos modificados considerados.

### 5.2.2 Parâmetros para o caso MIMO

Os parâmetros utilizados nos algoritmos propostos são similares aos parâmetros utilizados no caso SISO. O parâmetro  $K$  nas Equações (2.44) e (2.47) foram definidos como 1, e o  $\beta$  definido como 0, i.e., o índice de desempenho é puramente preditivo. O tamanho da população e o RIP inicial é discutido na Subseção 5.2.3. Os limites inferiores e superiores para a RIP foram definidos como  $\mu_{t,j}$  e  $1 - \mu_{t,j}$ , respectivamente, a fim de evitar uma exclusão total dos regressores. O limiar do ERR ( $\delta$ ) foi definido como 0,1%.

No que se segue, a menos que seja especificado, estes parâmetros serão utilizados.

### 5.2.3 Tamanho da População e Probabilidade de Inclusão dos Regressores

A combinação do tamanho da população e da probabilidade de inclusão de regressores (RIP) são os principais parâmetros responsáveis por explorar o espaço de busca eficientemente. Assim, os testes nesta subseção mostram a influência destes dois parâmetros no tempo de convergência e na seleção correta dos modelos.

Para entender a influência do tamanho da população e da RIP nos métodos propostos, 50 simulações de Monte Carlo foram feitas utilizando o sistema  $\mathcal{S}_9$  (Tabela (5.8)), variando o tamanho da população de 5 a 150. Quatro conjuntos de regressores candidatos foram usados  $m = [455, 1330, 2925, 5456]$ . O número máximo de iterações foi definido como 200, quando o algoritmo extrapola estes valores, foi considerado que o algoritmo não convergiu.

Analisando as Figuras (5.10a), (5.10b), (5.10c), (5.10d) pode-se inferir que uma população menor é mais benéfica do que uma maior, devido ao fato do tempo decorrido até a convergência aumentar com o aumento da população. Contudo, observando as Figuras (5.10e), (5.10f), (5.10g) e (5.10h), pode-se notar que a seleção correta do modelo é extremamente afetada pelo tamanho da população, o que é devido ao fato de uma população pequena não permitir uma exploração eficiente do espaço de busca. Nas Figuras (5.10e), (5.10f), (5.10g) e (5.10h), é possível também observar que quando se aumenta o conjunto de regressores candidatos a seleção correta do modelo diminui, desta forma é necessário que haja uma combinação adequada do tamanho da população e da RIP inicial para garantir uma eficiente exploração do espaço de busca.

Aumentando o RIP inicial  $\mu_j$  como nas Figuras (5.10b) e (5.10d), é possível observar que o tempo decorrido até a convergência diminui quando comparados às Figuras (5.10a) e (5.10c). Esta melhora pode ser explicada pelo fato de que quando se aumenta a RIP inicial, o algoritmo explora modelos com um maior número de regressores, o que também permite uma melhor exploração do espaço de busca, aumentando também a taxa de seleção correta dos modelos, o que pode ser visto nas Figuras (5.10f) e (5.10h).

Para o caso de  $m = 5456$  o tamanho da população de  $Z_M = 150$  e RIP inicial de  $\mu_j = 10/m$  foi insuficiente para atingir 100% de seleção correta do modelo no caso RaMSS-EGI, Figura (5.10h). Para este caso, foi feito um teste adicional com  $Z_M = 100$  e RIP inicial de  $\mu_j = 1/Z_M$ . O algoritmo RaMSS-EGI convergiu em 2,23 segundos, e atingiu 100% de seleção correta. Assim a menos que seja especificado um valor diferente, o tamanho da população será de 100 e o RIP inicial de  $1/Z_M$  para os testes realizados nas próximas subseções.

Figura 5.10: Tempo decorrido até a convergência (a) e (b), e Seleção Correta (c) e (d), todos em função do tamanho da população para o método RaMSS-EGI.

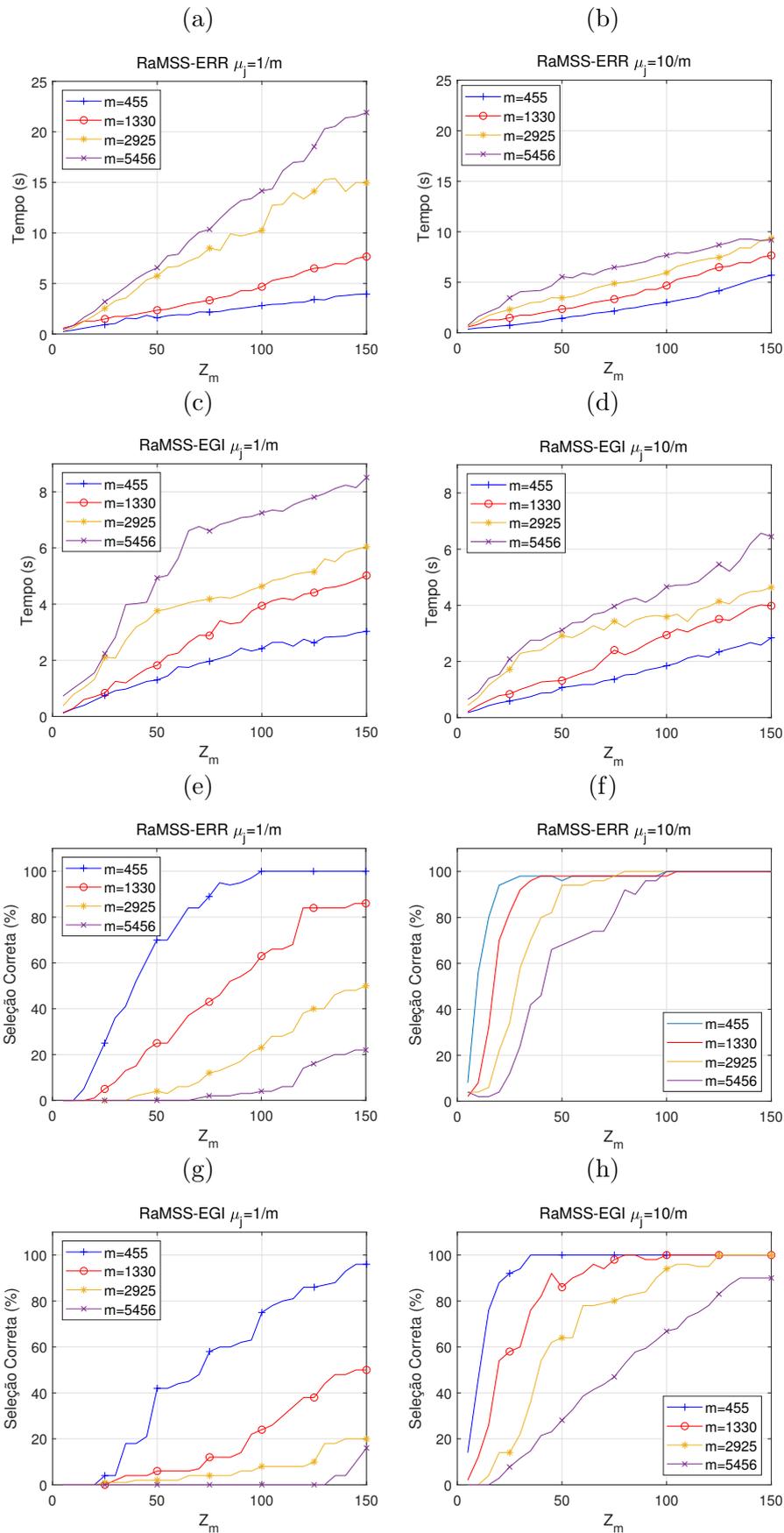


Tabela 5.9: Análise comparativa dos algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI para os sistemas MIMO.

Método	Sistema	Z <sub>m</sub>	CR	A	IT	ET	M.AMS	F.AMS	EM
RaMSS	$\mathcal{S}_9$	100	1,00	0,35	65,75	8,35	8,80	6,40	2929
	$\mathcal{S}_{10}$	100	1,00	0,40	71,45	8,04	13,15	6,97	4401
	$\mathcal{S}_{11}$	100	1,00	0,25	134,70	32,54	7,25	7,14	2418
	$\mathcal{S}_{12}$	100	1,00	0,30	87,65	36,54	6,25	6,12	2017
	$\mathcal{S}_{13}$	100	1,00	0,45	51,80	16,90	4,90	4,15	1591
	$\mathcal{S}_{14a}$	100	0,25	0,00	444,25	142,19	45,50	44,74	38689
	$\mathcal{S}_{14b}$	100	1,00	0,00	479,12	460,37	205,75	199,47	45093
	$\mathcal{S}_{14a}$	500	1,00	0,43	218,95	214,17	17,85	7,65	14368
	$\mathcal{S}_{14b}$	500	1,00	0,51	142,15	107,58	12,15	6,12	11887
	RaMSS-ERR	$\mathcal{S}_9$	100	1,00	0,80	33,55	13,11	5,12	4,91
$\mathcal{S}_{10}$		100	1,00	1,00	49,80	18,19	8,00	6,00	4296
$\mathcal{S}_{11}$		100	1,00	0,35	114,50	25,62	7,30	6,50	2384
$\mathcal{S}_{12}$		100	1,00	1,00	26,02	20,97	6,20	4,01	1802
$\mathcal{S}_{13}$		100	1,00	0,60	53,20	24,70	4,95	4,45	1490
$\mathcal{S}_{14a}$		100	1,00	1,00	21,25	6,65	6,45	3,00	1457
$\mathcal{S}_{14b}$		100	1,00	1,00	22,95	10,54	5,80	3,03	1682
$\mathcal{S}_9$		1,00	1,00	1,00	14,20	5,38	5,25	5,03	866
RaMSS-EGI	$\mathcal{S}_{10}$	100	1,00	1,00	18,60	8,08	7,15	6,01	1609
	$\mathcal{S}_{11}$	100	1,00	1,00	23,30	5,36	5,30	5,00	889
	$\mathcal{S}_{12}$	100	1,00	1,00	9,50	8,07	5,50	4,01	756
	$\mathcal{S}_{13}$	100	1,00	1,00	12,85	8,41	4,30	4,00	668
	$\mathcal{S}_{14a}$	100	1,00	1,00	8,25	3,74	6,05	3,07	639
	$\mathcal{S}_{14b}$	100	1,00	1,00	8,85	5,51	5,30	3,39	704

#### 5.2.4 Análise e testes comparativos para o caso MIMO

Similarmente aos testes apresentados na seção 5.1.5, nesta seção os algoritmos RaMSS, RaMSS-ERR e RaMSS-EGI são aplicados aos sistemas da Tabela (5.8) com o intuito de testar suas capacidades de identificar corretamente a estrutura dos modelos. O teste consistiu de 100 simulações de Monte Carlo para cada sistema na Tabela (5.8), com o limite máximo de 500 iterações. Na Tabela (5.9) os índices  $CR$ ,  $A$ ,  $IT$ ,  $ET$ ,  $M.AMS$ ,  $F.AMS$  e  $EM$  correspondem a taxa de convergência que indica a fração de vezes que o algoritmo convergiu, a precisão que indica a porcentagem de vezes que o algoritmo convergiu para o modelo correto, número de iterações, tempo até a convergência, tamanho médio máximo dos modelos, tamanho médio final dos modelos e número de modelos explorados, respectivamente.

Na Tabela (5.9) a diferença entre os métodos é notória. Em relação ao número de iterações, o algoritmo RaMSS-ERR convergiu em média 15% a menos do que o número

de iterações necessários para o algoritmo RaMSS convergir no pior caso (sistema  $\mathcal{S}_{11}$ ). Já o algoritmo RaMSS-EGI convergiu em um número de iterações quatro vezes menor do que o RaMSS, também para o pior caso. Em relação ao tempo de convergência, o método RaMSS-ERR superou o RaMSS clássico apenas nos sistemas  $\mathcal{S}_{12}$  e  $\mathcal{S}_{14}$ , mostrando que os cálculos exigidos pelo ERR, apesar de melhorarem a convergência e explorarem um número menor de modelos, podem aumentar o tempo demandado para a convergência. Já os tempos de convergência para o método RaMSS-EGI foram em todos os casos menores que os apresentados pelos outros métodos. Outro destaque que se observa na Tabela (5.9) é com relação ao número de modelos explorados, que para os métodos propostos são bem menores do que do algoritmo clássico.

Por fim, observa-se que para os sistemas  $\mathcal{S}_{14a}$  e  $\mathcal{S}_{14b}$  a precisão dos métodos propostos supera a do método clássico, enquanto o RaMSS utilizando  $Z_M = 100$  não convergiu para o modelo correto nenhuma vez, aumentando-se  $Z_M$  para 500 a precisão do método aumentou para 0.43 e 0.51 respectivamente para os sistemas  $\mathcal{S}_{14a}$  e  $\mathcal{S}_{14b}$ . No caso dos métodos propostos, ambos apresentaram uma precisão de 100% para ambos os modelos usando apenas uma população com  $Z_M = 100$ , mostrando assim serem mais precisos que o RaMSS clássico.

### 5.2.5 Identificação da Coluna de Separação C3/C4

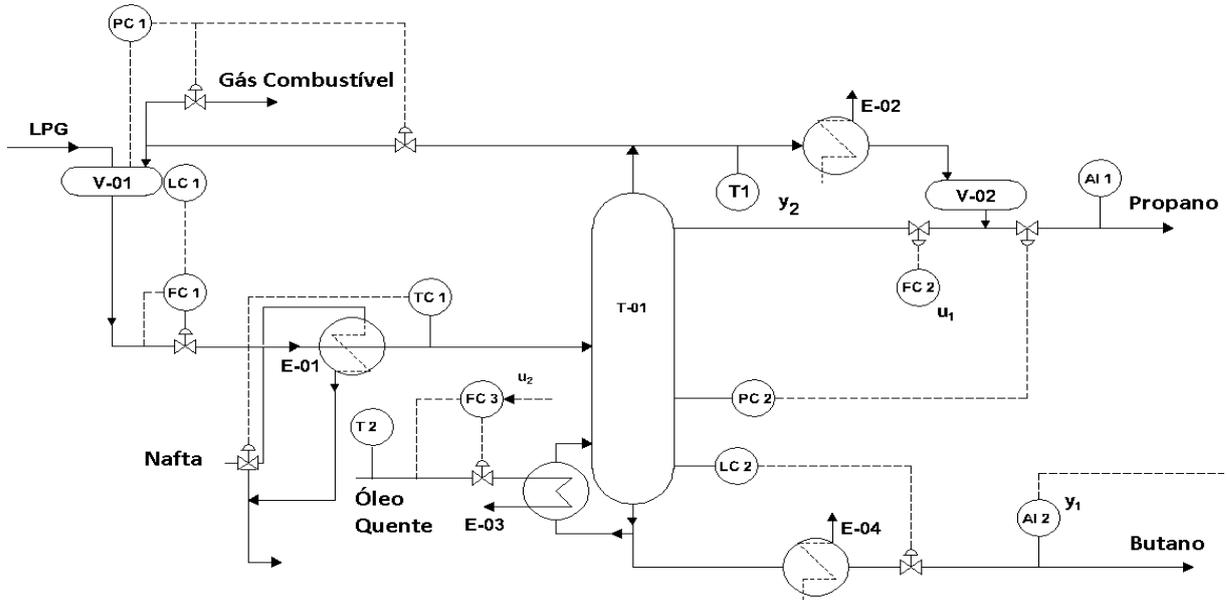
Para validar os métodos, estes foram aplicados na identificação de uma coluna de separação C3/C4, a qual consiste em um subsistema de uma unidade de Craqueamento Catalítico. Esta coluna separa a corrente C3, a qual contém em sua composição hidrocarbonetos com três carbonos, como propano e propeno, da corrente C4, a qual contém em sua composição hidrocarbonetos com quatro carbonos, como butano e buteno.

A corrente de alimentação da coluna é o gás liquefeito de petróleo (*Liquefied Petroleum Gas - LPG*) proveniente de uma coluna debutanizadora. O principal objetivo do processo é a maximização da recuperação de C4, a qual é usada para se obter gasolina com alta octanagem.

O sistema pode ser visto na Figura (5.11).

As variáveis controladas são a composição molar de C3 na corrente de fundo ( $y_1$ )

Figura 5.11: Sistema de separação C3/C4.



Fonte: Adaptado de Porfirio *et al.*, 2003.

medida pelo analisador AI2 em % e a temperatura de topo no primeiro estágio da seção da coluna T-01 ( $y_2$ ) medida por T1 em  $^{\circ}\text{C}$ . As entradas manipuladas são a vazão de refluxo na coluna de topo T-01 ( $u_1$ ) em  $\text{m}^3/\text{d}$  e a vazão de óleo quente no reboiler ( $u_2$ ) também em  $\text{m}^3/\text{d}$ .

A Equação (5.12) (Porfirio *et al.*, 2003) representa o modelo contínuo da coluna C3/C4 e foi usado para gerar os dados de entrada e saída usados no processo de identificação.

$$G(s) = \begin{bmatrix} G_{1,1}(s) & G_{1,2}(s) \\ G_{2,1}(s) & G_{2,2}(s) \end{bmatrix} e^{-1s} \quad (5.12)$$

e as funções de transferência  $G_{z,w}(s)$  são definidas como,

$$G_{1,1}(s) = \frac{-2,218 \times 10^{-4}s + 5,656 \times 10^{-4}}{s^2 + 3,4948s + 0,5902}$$

$$G_{1,2}(s) = \frac{0,7413 \times 10^{-4}s - 1,452 \times 10^{-3}}{s^2 + 2,6987s + 0,4023}$$

$$G_{2,1}(s) = \frac{-1,135 \times 10^{-3}s - 1,235 \times 10^{-3}}{s^2 + 1,6380s + 0,09852}$$

$$G_{2,2}(s) = \frac{-0,3 \times 10^{-3}s + 0,2 \times 10^{-2}}{s^2 + 2,4298s + 0,0651}$$

Tabela 5.10: Métricas obtidas na identificação do sistema de separação C3/C4 usando RaMSS-ERR e RaMSS-EGI.

Saída	$\delta$ (%)	Método	MSSE	IT	ET	EM
$y_1$	0,1	RaMSS-ERR	0,9978	377	1395,07	2287
		RaMSS-EGI	0,9978	366	1621,95	1360
$y_2$	0,1	RaMSS-ERR	0,0686	8	35,38	208
		RaMSS-EGI	0,8332	12	42,29	187
$y_1$	0,01	RaMSS-ERR	0,9980	923	2931,16	18251
		RaMSS-EGI	0,9978	327	1178,24	7783
$y_2$	0,01	RaMSS-ERR	0,9268	170	653,31	916
		RaMSS-EGI	0,9273	45	195,08	399

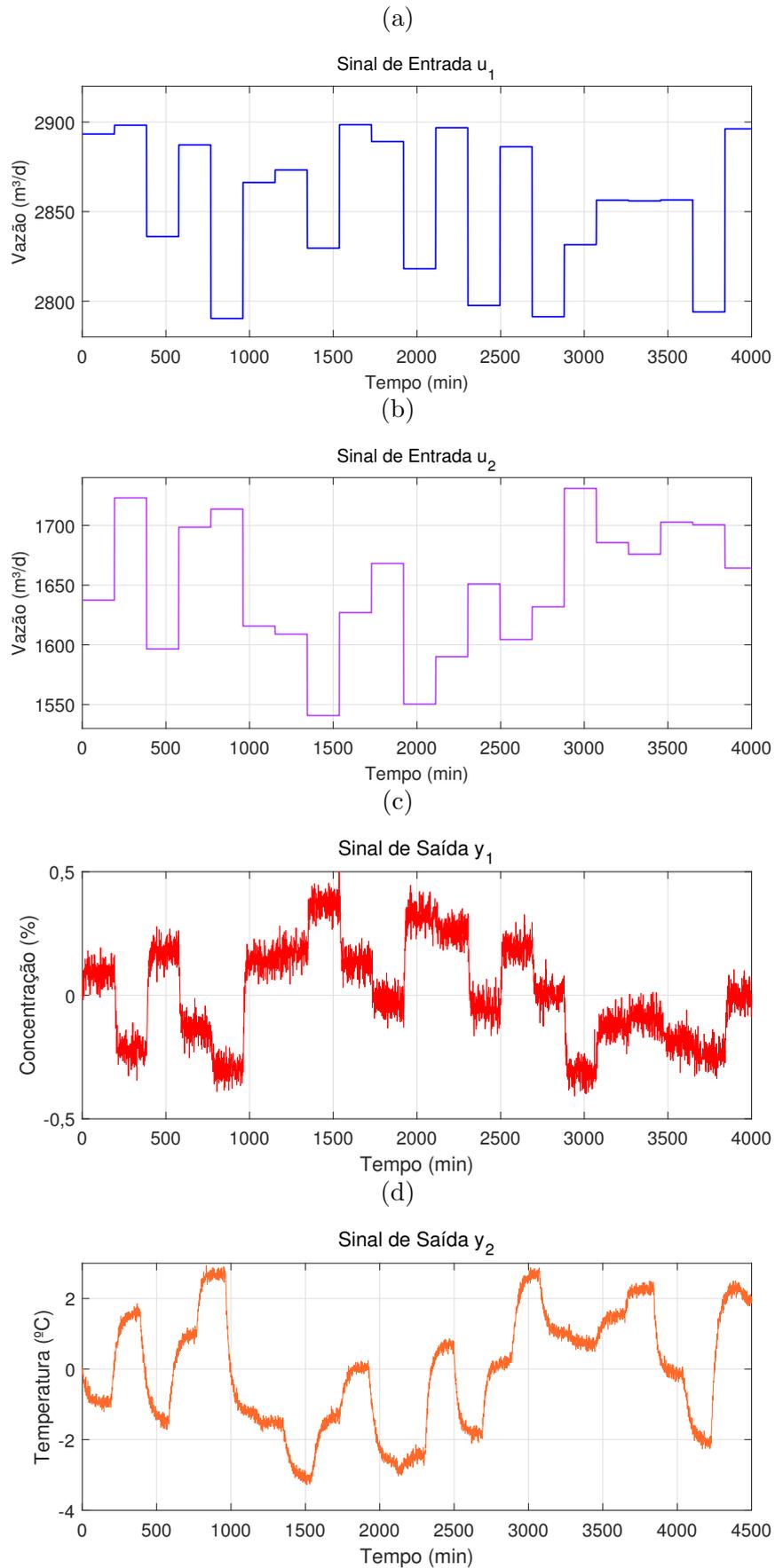
O sistema foi excitado próximo de 75% para a composição molar da corrente de C3 e 49 °C para a temperatura no primeiro estágio da seção da coluna de topo. O sinal de entrada foi gerado usando um sinal pseudoaleatório de multi-nível (Hrenya, 1999), mantendo a entrada  $u_1$  no intervalo 2700 a 3000  $m^3/d$ , e a entrada  $u_2$  no intervalo 1500 a 1800  $m^3/d$ . Os testes foram feitos em uma janela de tempo de 4000 minutos, sendo amostrados em  $5,55 \times 10^{-3} Hz$ , resultando em um total de 1333 amostras para dois conjuntos de dados, um utilizado na identificação e outro para validação. Os sinais de entradas e saídas usados no processo de identificação são mostrados na Figura (5.12).

Os algoritmos RaMSS-ERR e RaMSS-EGI foram aplicados na identificação do sistema de separação C3/C4 usando os parâmetros da seção 5.2.2, com  $n_u = n_y = 10$  e grau de não linearidade  $l = 3$ . Também foram feitos testes usando dois valores para o limiar  $\delta$  do ERR. Os modelos identificados usando o primeiro valor de 0,1%, que também foi utilizado em todos os testes predecessores, apresentou um comportamento dinâmico insatisfatório. Desta forma, foram feitos ajustes ao valor de  $\delta$  até o valor de 0,01%, em que um modelo mais adequado foi identificado.

A Tabela (5.10) mostra os resultados para os modelos identificados usando os métodos propostos em dois níveis de  $\delta$ , 0,1% e 0,01%.

Na Tabela (5.10) observa-se que para os valores de  $\delta = 0,1\%$  ambos os métodos identificaram modelos com MSSE acima de 0,99 para a saída  $y_1$ , o mesmo não acontecendo para a saída  $y_2$  em que o método RaMSS-ERR identificou um modelo com MSSE muito baixo (0,0686). Para a saída  $y_2$  o valor de  $\delta = 0,1\%$  se mostrou um valor muito alto, tendo

Figura 5.12: Sinais usados na identificação do sistema C3/C4. Entradas  $u_1$  e  $u_2$ , saídas  $y_1$  e  $y_2$ , respectivamente em (a), (b), (c) e (d).



em vista que para este caso o modelo identificado pelo método se resumia ao preditor trivial  $y(k-1)$ . Assim, procedeu-se diminuindo o valor de  $\delta$  por um fator de 10. Com esse valor de  $\delta$  o algoritmo RaMSS-ERR identificou um modelo composto por mais termos além do preditor trivial  $y(k-1)$  e que apresentou um melhor desempenho em simulação, com um MSSE de 0,9268.

Ainda de acordo com os testes na Tabela (5.10) nota-se que o número de iterações, tempo até a convergência e número de modelos explorados para  $\delta = 0,1\%$  se destoam dos valores encontrados nos demais testes. Para a saída  $y_2$  a convergência se deu de forma muito rápida em relação a saída  $y_1$ , menos de 1 minuto, o que pode ser indício de uma convergência prematura.

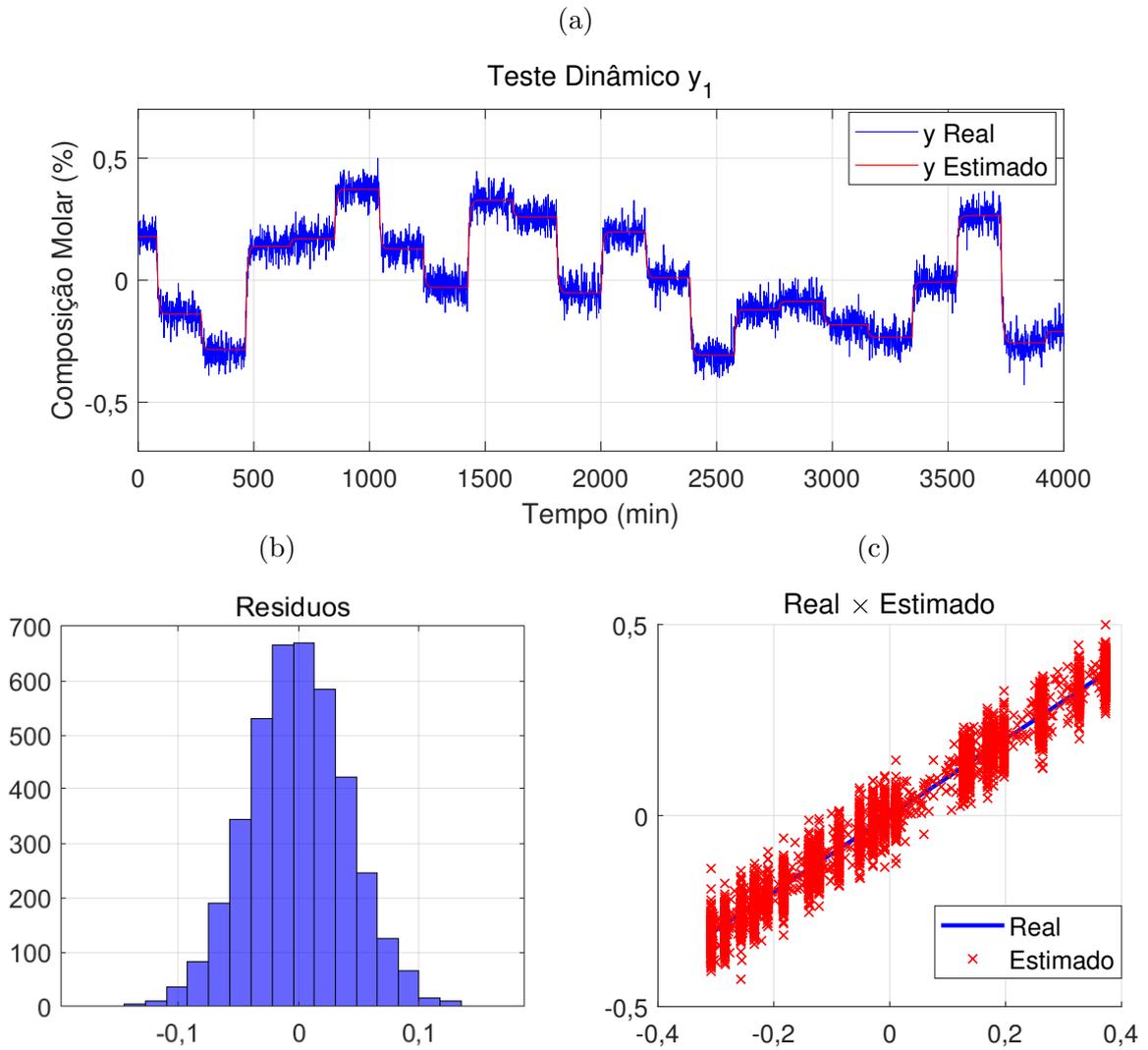
Os modelos identificados usando RaMSS-ERR são mostrados nas Equações (5.13) e (5.14) e os modelos identificados usando RaMSS-EGI são mostrados nas Equações (5.15) e (5.16).

$$\begin{aligned}
 y_1(k) = & 0.116y_1(k-8) + 0.104y_1(k-11) + 3,99 \times 10^{-5}u_1(k-4) \\
 & + 3,41 \times 10^{-5}u(k-7) - 3,33 \times 10^{-5}u_2(k-2) \\
 & - 4,90 \times 10^{-5}u_2(k-3) - 4,00 \times 10^{-4}u_2(k-4) \\
 & - 4,72 \times 10^{-4}u_2(k-5) - 3,98 \times 10^{-4}u_2(k-7) \\
 & - 3,26 \times 10^{-4}u_2(k-8) - 3,85 \times 10^{-5}u_2(k-10)
 \end{aligned} \tag{5.13}$$

$$\begin{aligned}
 y_2(k) = & 0.359y_2(k-1) + 0.300y_2(k-4) + 0.257y_2(k-5) \\
 & - 1,248 \times 10^{-3}u_1(k-1) + 9,73 \times 10^{-4}u_2(k-1) \\
 & + 1,549 \times 10^{-3}u_2(k-6)
 \end{aligned} \tag{5.14}$$

$$\begin{aligned}
 y_1(k) = & 0.163y_1(k-8) + 0.102y_1(k-10) \\
 & + 6.960 \times 10^{-4}u_1(k-5) - 9.940 \times 10^{-4}u_2(k-3) \\
 & - 8.540 \times 10^{-4}u_2(k-5) - 7.890 \times 10^{-4}u_2(k-8)
 \end{aligned} \tag{5.15}$$

Figura 5.13: Resposta dos modelos identificados usando RaMSS-EGI para a concentração  $y_1$ . Teste dinâmico (a), resíduos (b), primeira bisettriz  $y_{1_{Real}} \times y_{1_{Estimado}}$  (c).

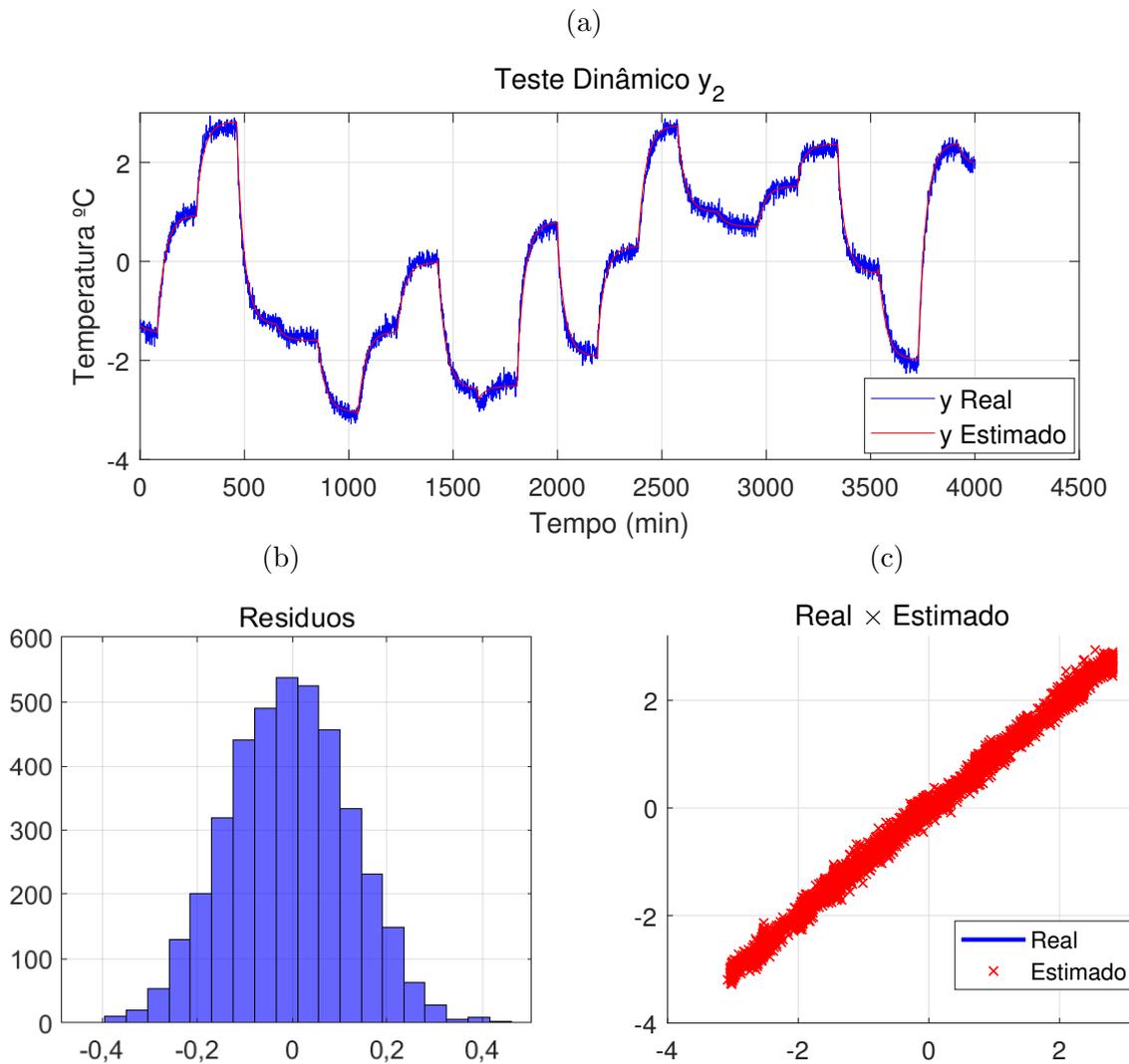


$$\begin{aligned}
 y_2(k) = & 0.269y_2(k-1) + 0.225y_2(k-2) + 0.217y_2(k-3) \\
 & + 0.224y_2(k-4) - 9.680 \times 10^{-4}u_1(k-1) \\
 & + 1.930 \times 10^{-3}u_2(k-4)
 \end{aligned} \tag{5.16}$$

Embora os modelos apresentem desempenhos semelhantes (Tabela (5.10)), observa-se que os modelos obtidos usando o RaMSS-EGI são mais compactos, sendo preferíveis aos identificados pelo método RaMSS-ERR. Os modelos identificados usando o RaMSS-EGI foram validados usando um conjunto de dados de validação e a resposta dinâmica dos modelos obtidos é mostrada na Figura (5.13) e na Figura (5.14). Todos os testes foram feitos em simulação.

O teste dinâmico nas Figuras (5.13a) e (5.14a) mostram que o modelo identificado

Figura 5.14: Resposta dos modelos identificados usando RaMSS-EGI para a concentração  $y_2$ . Teste dinâmico (a), resíduos (b), primeira bisettriz  $y_{2_{Real}} \times y_{2_{Estimado}}$  (c).



conseguiu representar adequadamente a dinâmica do modelo, tanto em relação à saída  $y_1$  quanto à saída  $y_2$ . É possível ainda notar que nas Figuras (5.13b) e (5.14b) os resíduos apresentam uma distribuição normal, indicando uma adequada identificação do modelo. Já na Figura (5.13c) nota-se no gráfico da primeira bisettriz uma quantidade de ruído em alguns patamares da resposta, o que justifica os pontos verticalmente dispostos para um dado nível de composição molar, comportamento este que não é tão claro para a Figura (5.14c) por apresentar uma resposta sem patamares claramente definidos.

### 5.2.6 Discussão sistemas MISO e MIMO

Os resultados apresentados na seção 5.2 mostram que assim como no caso SISO, os algoritmos RaMSS-ERR e RaMSS-EGI tiveram um melhor desempenho que o RaMSS no caso MIMO. Sendo esta melhora devido aos fatores já previamente mencionados, sendo eles a inclusão da taxa de redução do erro (ERR) no processo e a contribuição de uma segunda população  $M_{best}$  para o processo de atualização da RIP no método RaMSS-EGI.

Outro ponto de convergência para o caso MIMO e SISO diz respeito ao tamanho da população, para ambos foi notado que uma população  $Z_m = 100$  apresentou um desempenho satisfatório. Já em relação ao valor inicial do RIP, pelo fato dos sistemas multivariáveis apresentarem várias entradas e possivelmente várias saídas, o tamanho dos conjuntos de regressores candidatos é maior que para o caso SISO, o que faz com que o espaço de busca possa se tornar por vezes muito grande. Assim, para o caso MIMO o uso de uma RIP inicial de  $\mu_{t,j} = 1/Z_M$  se mostrou mais efetiva, permitindo uma melhor exploração do espaço de busca e consequentemente uma melhora na convergência e precisão dos métodos propostos.

Na seção 5.2.4 foi possível notar que, embora os algoritmos propostos superem o algoritmo RaMSS clássico em termos de número de iterações, o tempo de convergência do método RaMSS-ERR pode em algumas vezes ultrapassar o do RaMSS clássico, o que mostra que os cálculos adicionais para se aplicar o filtro pelo ERR, embora possam acelerar a convergência, podem também tornar o tempo até a convergência maior. Já para o RaMSS-EGI tanto o número de iterações quanto o tempo até a convergência foram ambos menores do que o algoritmo RaMSS clássico.

Na identificação da coluna de separação C3/C4, embora a identificação do modelo tenha sido satisfatória, foi possível observar uma desvantagem na metodologia proposta, que diz respeito à necessidade de se sintonizar o limiar  $\delta$  do filtro ERR. Inicialmente para o valor de  $\delta = 0,1\%$  foi notado que os modelos identificados para a saída  $y_2$  apresentavam apenas o preditor trivial  $y(k-1)$ , que apresentava um valor de ERR muito elevado, assim todos os outros regressores eram eliminados do modelo. Ao aumentar o valor de  $\delta$  para  $0,01\%$  foi possível obter modelos mais completos que apresentavam melhor desempenho em simulação.

Embora exista essa necessidade de sintonizar o valor de  $\delta$  em alguns casos, os

métodos propostos se mostraram eficientes na identificação de diversos modelos propostos em um número de iterações e tempo de convergência menores, e ainda explorando um menor número de modelos que o algoritmo RaMSS clássico.

# CAPÍTULO VI

## CONCLUSÕES E SUGESTÕES

O presente trabalho propôs dois algoritmos para a seleção da estrutura de modelos que trabalham em um cenário probabilístico com o auxílio da taxa de redução do erro e de herança genética.

A taxa de redução do erro é utilizada como um pré-filtro para se remover dos modelos aqueles regressores que não contribuem significativamente para a diminuição da variância, fazendo com que os regressores ruins tenham a sua probabilidade de inclusão reduzida, diminuindo assim a probabilidade de serem selecionados na iteração seguinte.

Já com relação à Herança Genética, esta trabalha de forma a transmitir para as próximas gerações os melhores regressores, de forma que a probabilidade de inclusão do regressor seja aumentada e conseqüentemente se tenha uma maior probabilidade de seleção destes regressores nas próximas iterações.

Foram apresentados alguns resultados utilizando os métodos na identificação de modelos SISO e MIMO. Diversos sistemas de benchmark da literatura que usualmente apresentam dificuldade em serem identificados utilizando o algoritmo FROLS foram trabalhados, demonstrando uma melhora dos métodos propostos em relação ao algoritmo RaMSS clássico. Os sistemas foram identificados com um menor número de iterações, menor tempo e explorando uma quantidade menor de modelos.

Foi observado também que os algoritmos propostos apresentam um melhor desem-

penho quando o processo de identificação requer um grande conjunto de regressores candidatos (acima de 10.000 regressores candidatos), como no caso de sistemas com resposta inversa, sujeito a longos atrasos, ou com alto grau de não linearidades.

Como exemplo de aplicações, foi identificado no caso SISO um Reator Tanque Agitado (CSTR) e no caso MIMO uma coluna C3/C4, em ambos os casos os métodos propostos foram capazes de identificar modelos satisfatórios. Para o CSTR o modelo identificado pelo algoritmo RaMSS-ERR apresentou erro quadrático médio de simulação de 0,9816 e 0,9965 para as saídas  $C_A$  e  $C_B$ , com tempos de convergência de 25 e 48 segundos respectivamente. Já para a coluna C3/C4, o erro quadrático de simulação foi de 0,9980 e 0,9268 para a concentração  $y_1$  e temperatura  $y_2$  respectivamente. O maior tempo de convergência foi de 653 segundos.

Em contrapartida, uma desvantagem observada nos métodos propostos foi uma piora na robustez em relação ao RaMSS clássico na identificação de sistemas sujeitos a entradas pouco excitativas, isto é, o RaMSS clássico por não utilizar a taxa de redução do erro consegue obter modelos melhores quando sujeito a entradas pouco excitativas. A inclusão da taxa de redução do erro como pré-filtro se mostrou deficiente, uma vez que torna os métodos suscetíveis a problemas de condicionamento na matriz de estimação dos parâmetros.

Alguns pontos nos métodos propostos podem ainda ser explorados em trabalhos futuros, tais como:

- Utilização de outra métrica como pré-filtro no lugar da taxa de redução de erro, tal como a distância de correlação, ou a taxa de redução do erro de simulação;
- O uso de técnicas de identificação caixa cinza, que permitem que informações da natureza do processo sejam incorporadas no processo de identificação, diminuindo o espaço de busca e conseqüentemente aumentando a probabilidade de seleção correta da estrutura;
- É possível ainda explorar o uso de outros critérios de seleção do modelo, tendo em vista que este é um ponto crucial para determinar o quão satisfatório o modelo encontrado é.

Conclui-se o presente trabalho observando que os algoritmos propostos RaMSS-

ERR e RaMSS-EGI cumpriram com o objetivo principal que era de acelerar o processo de convergência e identificação do modelo, considerando para isso o número de iterações, o tempo até a convergência e o número de modelos explorados. Embora os métodos propostos tenham obtido um bom desempenho na maioria dos sistemas testados, observou-se uma piora na identificação de sistemas sujeitos a entradas pouco excitativas, de forma que ainda é possível que o tema seja explorado visando melhorias pontuais.



## Referências Bibliográficas

- ABBAS, H. M.; BAYOUMI, M. M. Volterra system identification using adaptive genetic algorithms. *Applied Soft Computing*, Elsevier BV, v. 5, n. 1, p. 75–86, dec 2004. <http://dx.doi.org/10.1016/j.asoc.2004.05.001>.
- AGUIRRE; BILLINGS, S. Dynamical effects of overparametrization in nonlinear models. *Physica D: Nonlinear Phenomena*, Elsevier BV, v. 80, n. 1-2, p. 26–40, jan 1995. [http://dx.doi.org/10.1016/0167-2789\(95\)90053-5](http://dx.doi.org/10.1016/0167-2789(95)90053-5).
- AGUIRRE, L. *Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. [S.l.]: Editora UFMG, 2015. ISBN 9788570415844.
- AGUIRRE, L. A.; BARBOSA, B. H.; BRAGA, A. P. Prediction and simulation errors in parameter estimation for nonlinear systems. *Mechanical Systems and Signal Processing*, Elsevier BV, v. 24, n. 8, p. 2855–2867, nov 2010. <http://dx.doi.org/10.1016/j.ymsp.2010.05.003>.
- AGUIRRE, L. A.; BILLINGS, S. A. Improved structure selection for nonlinear models based on term clustering. *International Journal of Control*, Informa UK Limited, v. 62, n. 3, p. 569–587, sep 1995. <http://dx.doi.org/10.1080/00207179508921557>.
- AKINOLA, T. E. et al. Non-linear system identification of solvent-based post-combustion CO<sub>2</sub> capture process. *Fuel*, Elsevier BV, v. 239, p. 1213–1223, mar 2019. <http://dx.doi.org/10.1016/j.fuel.2018.11.097>.
- AL-NAUMANI, Y.; ROSSITER, J. Distributed MPC for upstream oil & gas fields - a practical view. *IFAC-PapersOnLine*, Elsevier BV, v. 48, n. 8, p. 325–330, 2015. <http://dx.doi.org/10.1016/j.ifacol.2015.08.202>.
- AVELLINA, M.; BRANKOVIC, A.; PIRODDI, L. Distributed randomized model structure selection for NARX models. *International Journal of Adaptive Control and Signal Processing*, Wiley, v. 31, n. 12, p. 1853–1870, aug 2017. <http://dx.doi.org/10.1002/acs.2803>.
- BALAGUER, P.; ALFARO, V.; ARRIETA, O. Second order inverse response process identification from transient step response. *ISA Transactions*, Elsevier BV, v. 50, n. 2, p. 231–238, apr 2011. <http://dx.doi.org/10.1016/j.isatra.2010.11.005>.
- BALDACCHINO, T.; ANDERSON, S. R.; KADIRKAMANATHAN, V. Structure detection and parameter estimation for NARX models in a unified EM framework. *Automatica*, Elsevier BV, v. 48, n. 5, p. 857–865, may 2012. <http://dx.doi.org/10.1016/j.automatica.2012.02.021>.

BALDACCHINO, T.; ANDERSON, S. R.; KADIRKAMANATHAN, V. Computational system identification for bayesian NARMAX modelling. *Automatica*, Elsevier BV, v. 49, n. 9, p. 2641–2651, sep 2013. <http://dx.doi.org/10.1016/j.automatica.2013.05.023>.

BILLINGS, S. A. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. [S.l.]: PAPERBACKSHOP UK IMPORT, 2013. ISBN 1119943590.

BILLINGS, S. A.; CHEN, S.; KORENBERG, M. J. Identification of MIMO non-linear systems using a forward-regression orthogonal estimator. *International Journal of Control*, Informa UK Limited, v. 49, n. 6, p. 2157–2189, jun 1989. <http://dx.doi.org/10.1080/00207178908559767>.

BILLINGS, S. A. et al. Identification of nonlinear systems with non-persistent excitation using an iterative forward orthogonal least squares regression algorithm. *International Journal of Modelling, Identification and Control*, Inderscience Publishers, v. 23, n. 1, p. 1, 2015. <http://dx.doi.org/10.1504/ijmic.2015.067496>.

BILLINGS, S. A.; KORENBERG, M. J.; CHEN, S. Identification of non-linear output-affine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, Informa UK Limited, v. 19, n. 8, p. 1559–1568, jan 1988. <http://dx.doi.org/10.1080/00207728808964057>.

BOMBERGER, J. D.; SEBORG, D. E. Determination of model order for NARX models directly from input-output data. *Journal of Process Control*, Elsevier BV, v. 8, n. 5-6, p. 459–468, oct 1998. [http://dx.doi.org/10.1016/s0959-1524\(98\)00013-4](http://dx.doi.org/10.1016/s0959-1524(98)00013-4).

BONIN, M.; PIRODDI, L.; SEGHEZZA, V. NARX model selection based on simulation error minimisation and LASSO. *IET Control Theory & Applications*, Institution of Engineering and Technology (IET), v. 4, n. 7, p. 1157–1168, jul 2010. <http://dx.doi.org/10.1049/iet-cta.2009.0217>.

BOYNTON, R. et al. Applications of NARMAX in space weather. In: *Machine Learning Techniques for Space Weather*. [S.l.]: Elsevier, 2018. p. 203–236. <http://dx.doi.org/10.1016/b978-0-12-811788-0.00008-1>.

CHEN, S.; BILLINGS, S. A. Representations of non-linear systems: the NARMAX model. *International Journal of Control*, Informa UK Limited, v. 49, n. 3, p. 1013–1032, mar 1989. <http://dx.doi.org/10.1080/00207178908559683>.

FALSONE, A.; PIRODDI, L.; PRANDINI, M. A randomized algorithm for nonlinear model structure selection. *Automatica*, Elsevier BV, v. 60, p. 227–238, oct 2015. <http://dx.doi.org/10.1016/j.automatica.2015.07.023>.

FEIL, B.; ABONYI, J.; SZEIFERT, F. Model order selection of nonlinear input–output models—a clustering based approach. *Journal of Process Control*, Elsevier BV, v. 14, n. 6, p. 593–602, sep 2004. <http://dx.doi.org/10.1016/j.jprocont.2004.01.005>.

FOGLER, H. S. *Elements of Chemical Reaction Engineering*. [S.l.]: Pearson Education (US), 2016. ISBN 0133887510.

GHABOUSSI, J.; WU, X. *Numerical Methods in Computational Mechanics*. CRC Press, 2016. ISBN 9781315351643. Disponível em: [⟨https://books.google.com.br/books?id=WuCVdQAAQBAJ⟩](https://books.google.com.br/books?id=WuCVdQAAQBAJ).

GOLDBERG, D. E.; HOLLAND, J. H. Genetic algorithms and machine learning. *Machine Learning*, Springer Nature, v. 3, n. 2/3, p. 95–99, 1988. <http://dx.doi.org/10.1023/a:1022602019183>.

GU, Y.; WEI, H.-L. A robust model structure selection method for small sample size and multiple datasets problems. *Information Sciences*, Elsevier BV, v. 451-452, p. 195–209, jul 2018. <http://dx.doi.org/10.1016/j.ins.2018.04.007>.

GUO, Y. et al. An iterative orthogonal forward regression algorithm. *International Journal of Systems Science*, Informa UK Limited, v. 46, n. 5, p. 776–789, nov 2014. <http://dx.doi.org/10.1080/00207721.2014.981237>.

GUO, Y. et al. Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems. *Neurocomputing*, Elsevier BV, v. 173, p. 715–723, jan 2016. <http://dx.doi.org/10.1016/j.neucom.2015.08.022>.

HAFIZ, F.; SWAIN, A.; MENDES, E. Orthogonal floating search algorithms: From the perspective of nonlinear system identification. *Neurocomputing*, Elsevier BV, apr 2019. <http://dx.doi.org/10.1016/j.neucom.2019.03.069>.

HAFIZ, F.; SWAIN, A.; MENDES, E. M. Two-dimensional (2d) particle swarms for structure selection of nonlinear systems. *Neurocomputing*, Elsevier BV, v. 367, p. 114–129, nov 2019. <http://dx.doi.org/10.1016/j.neucom.2019.07.071>.

HIRCH, M.; RE, L. del. Iterative identification of polynomial NARX models for complex multi-input systems. *IFAC Proceedings Volumes*, Elsevier BV, v. 43, n. 14, p. 445–450, sep 2010. <http://dx.doi.org/10.3182/20100901-3-it-2016.00187>.

HO, A. P. H.; NGUYEN, N. T. Identification of mimo dynamic system using inverse mimo neural narx model. *Science and Technology Development Journal*, Viet Nam National University Ho Chi Minh City, v. 16, n. 2, p. 13–25, jun 2013. <http://dx.doi.org/10.32508/stdj.v16i2.1506>.

HRENYA, M. W. B. D. E. R. A. S. W. F. C. Multi-level pseudo-random signal design and "model-on-demand" estimation applied to nonlinear identification of a RTP wafer reactor. In: *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*. [S.l.]: IEEE, 1999. <http://dx.doi.org/10.1109/acc.1999.786090>.

KARAMI, K.; WESTWICK, D.; SCHOUKENS, J. Applying polynomial decoupling methods to the polynomial NARX model. *Mechanical Systems and Signal Processing*, Elsevier BV, v. 148, p. 107134, feb 2021. <http://dx.doi.org/10.1016/j.ymssp.2020.107134>.

KEESMAN, K. J. *System Identification*. [S.l.]: Springer London, 2011. <http://dx.doi.org/10.1007/978-0-85729-522-4>.

KORENBERG, M. et al. Orthogonal parameter estimation algorithm for non-linear stochastic systems. *International Journal of Control*, Informa UK Limited, v. 48, n. 1, p. 193–210, jul 1988. <http://dx.doi.org/10.1080/00207178808906169>.

KUKREJA, S. L.; KEARNEY, R. E.; GALIANA, H. L. A bootstrap method for narmax model order selection. *IFAC Proceedings Volumes*, Elsevier BV, v. 33, n. 3, p. 329–332, mar 2000. [http://dx.doi.org/10.1016/s1474-6670\(17\)35537-4](http://dx.doi.org/10.1016/s1474-6670(17)35537-4).

LEONTARITIS, I. J.; BILLINGS, S. A. Input-output parametric models for non-linear systems part i: deterministic non-linear systems. *International Journal of Control*, Informa UK Limited, v. 41, n. 2, p. 303–328, feb 1985. <http://dx.doi.org/10.1080/0020718508961129>.

LJUNG, L.; SODERSTROM, T. *Theory and Practice of Recursive Identification (Signal Processing, Optimization, and Control)*. [S.l.]: The MIT Press, 1983. ISBN 9780262120951.

MAO, K. Z.; BILLINGS, S. A. Algorithms for minimal model structure detection in nonlinear dynamic system identification. *International Journal of Control*, Informa UK Limited, v. 68, n. 2, p. 311–330, jan 1997. <http://dx.doi.org/10.1080/002071797223631>.

MARLIN, T. *Process Control: Designing Processes and Control Systems for Dynamic Performance*. [S.l.]: McGraw-Hill Science/Engineering/Math, 2000. ISBN 0070393621.

MONTGOMERY ELIZABETH PECK, G. V. D. Introduction to linear regression analysis, fifth edition. *International Statistical Review*, Wiley, v. 81, n. 2, p. 318–319, aug 2013. [http://dx.doi.org/10.1111/insr.12020\\_10](http://dx.doi.org/10.1111/insr.12020_10).

NELLES, O. *Nonlinear System Identification*. [S.l.]: Springer Berlin Heidelberg, 2001. <http://dx.doi.org/10.1007/978-3-662-04323-3>.

PENG, Z. et al. Feasibility study of structural damage detection using NARMAX modelling and nonlinear output frequency response function based analysis. *Mechanical Systems and Signal Processing*, Elsevier BV, v. 25, n. 3, p. 1045–1061, apr 2011. <http://dx.doi.org/10.1016/j.ymssp.2010.09.014>.

PIRODDI, L.; SPINELLI, W. An identification algorithm for polynomial NARX models based on simulation error minimization. *International Journal of Control*, Informa UK Limited, v. 76, n. 17, p. 1767–1781, nov 2003. <http://dx.doi.org/10.1080/00207170310001635419>.

PORFIRIO, C.; Almeida Neto, E.; ODLOAK, D. Multi-model predictive control of an industrial c3/c4 splitter. *Control Engineering Practice*, v. 11, n. 7, p. 765 – 779, 2003. ISSN 0967-0661. [Htpps://doi.org/10.1016/S0967-0661\(02\)00183-1](https://doi.org/10.1016/S0967-0661(02)00183-1).

RAJ, K.; CHOUDHARY, S. K.; MUTHUKUMAR, V. Robust finite-time sliding mode control of twin rotor MIMO system. *International Journal of Modelling, Identification and Control*, Inderscience Publishers, v. 35, n. 1, p. 1, 2020. <http://dx.doi.org/10.1504/ijmic.2020.113292>.

RETES, P. F. L.; AGUIRRE, L. A. NARMAX model identification using a randomised approach. *International Journal of Modelling, Identification and Control*, Inderscience Publishers, v. 31, n. 3, p. 205, 2019. <http://dx.doi.org/10.1504/ijmic.2019.098779>.

ROJAS, J. D.; MORILLA, F.; VILANOVA, R. Multivariable PI control for a boiler plant benchmark using the virtual reference feedback tuning. *IFAC Proceedings Volumes*, Elsevier BV, v. 45, n. 3, p. 376–381, 2012. <http://dx.doi.org/10.3182/20120328-3-it-3014.00064>.

SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. ISSN 00905364. Disponível em: <http://www.jstor.org/stable/2958889>.

SOLARES, J. R. A.; WEI, H.-L. Nonlinear model structure detection and parameter estimation using a novel bagging method based on distance correlation metric. *Nonlinear Dynamics*, Springer Nature, v. 82, n. 1-2, p. 201–215, may 2015. <http://dx.doi.org/10.1007/s11071-015-2149-3>.

TUDOROIU, N.; ZAHEERUDDIN, M.; TUDOROIU, R. E. Modelling, identification, implementation and MATLAB simulations of multi-input multi-output proportional integral-plus control strategies for a centrifugal chiller system. *International Journal of Modelling, Identification and Control*, Inderscience Publishers, v. 35, n. 1, p. 64, 2020. <http://dx.doi.org/10.1504/ijmic.2020.113290>.

WEI, H.; BILLINGS, S. Improved model identification for non-linear systems using a random subsampling and multifold modelling (RSMM) approach. *International Journal of Control*, Informa UK Limited, v. 82, n. 1, p. 27–42, oct 2008. <http://dx.doi.org/10.1080/00207170801955420>.

WEI, H. L.; BILLINGS, S. A. Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information. *International Journal of Modelling, Identification and Control*, Inderscience Publishers, v. 3, n. 4, p. 341, 2008. <http://dx.doi.org/10.1504/IJMIC.2008.020543>.

WEI, H. L.; BILLINGS, S. A.; LIU, J. Term and variable selection for non-linear system identification. *International Journal of Control*, Informa UK Limited, v. 77, n. 1, p. 86–110, jan 2004. <http://dx.doi.org/10.1080/00207170310001639640>.

WESTWICK, D. T. et al. Using decoupling methods to reduce polynomial NARX models. *IFAC-PapersOnLine*, Elsevier BV, v. 51, n. 15, p. 796–801, 2018. <http://dx.doi.org/10.1016/j.ifacol.2018.09.133>.