



IDENTIFICABILIDADE DE PARÂMETROS: ASPECTOS CONCEITUAIS E
DESENVOLVIMENTO DE NOVOS PROCEDIMENTOS

Kese Pontes Freitas Alberton

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

Orientadores: Argimiro Resende Secchi
María Soledad Díaz

Rio de Janeiro
Dezembro de 2013

IDENTIFICABILIDADE DE PARÂMETROS: ASPECTOS CONCEITUAIS E
DESENVOLVIMENTO DE NOVOS PROCEDIMENTOS

Kese Pontes Freitas Alberton

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

Examinada por:

Prof. Argimiro Resende Secchi, D.Sc.

Profa. María Soledad Díaz, D.Sc.

Prof. Evaristo Chalbaud Biscaia Jr., D.Sc.

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

Prof. André Luiz Hemerly Costa, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
DEZEMBRO DE 2013

Alberton, Kese Pontes Freitas

Identificabilidade de parâmetros: aspectos conceituais e desenvolvimento de novos procedimentos/Kese Pontes Freitas Alberton. – Rio de Janeiro: UFRJ/COPPE, 2013.

XII, 192 p.: il.; 29,7 cm.

Orientadores: Argimiro Resende Secchi

María Soledad Díaz

Tese (doutorado) – UFRJ/ COPPE/ Programa de Engenharia Química, 2013.

Referências Bibliográficas: p. 182-192.

1. Identificabilidade de parâmetros. 2. Análise de sensibilidade. 3. Dados experimentais escassos. I. Secchi, Argimiro Resende et al. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Química. III. Título.

Ao meu grande amigo Eduardo Lemos, Mr. Monster;
seu exemplo de fé, força e coragem nos mostra a grandeza de um homem de bem.

UPSIDE DOWN

Jack Johnson

Who's to say
What's impossible?
Well they forgot
This world keeps spinning
And with each new day
I can feel a change in everything
And as the surface breaks reflections fade
But in some ways they remain the same
And as my mind begins to spread its wings
There's no stopping in curiosity

I wanna turn the whole thing upside down
I'll find the things they say just can't be found
I'll share this love I find with everyone
We'll sing and dance to mother nature's songs
I don't want this feeling to go away

Who's to say I can't do everything?
Well I can try, and as I roll along I begin to find
Things aren't always just what they seem
I want to turn the whole thing upside down
I'll find the things they say just can't be found
I'll share this love I find with everyone
We'll sing and dance to mother nature's songs

This world keeps spinning
And there's no time to waste
Well it all keeps spinning spinning
Round and round and upside down

Who's to say what's impossible and can't be found?
I don't want this feeling to go away
Please don't go away
Is this how it's supposed to be?

AGRADECIMENTOS

A Deus, agradeço por tudo e por todos.

À minha querida mãe Sandra; seu lindo coração de amor e forte espírito de luta me fizeram ir sempre além, ultrapassando limites, rompendo barreiras, quebrando preconceitos e mantendo em mim a vontade e a coragem de buscar um mundo melhor. Obrigada por seu exemplo de fé e honestidade, você é um ser humano fantástico e a melhor mãe do multiverso, eu te amo infinita e incondicionalmente.

Ao meu generoso pai Alberto, o meu querido super pai, que com grande sapiência e simplicidade tornou a minha infância e toda a minha vida muito especial. Obrigada por seu exemplo de fé, paciência e humildade, que sempre guiarão os meus passos e certamente os dos meus filhos, netos, bisnetos... Papai eu te amo muito, sempre e para sempre!

À minha doce irmã Ketly, a brilhante estrela que ilumina meu coração e traz paz a minha vida. Obrigada por me ensinar o amor generoso, a fé nos sonhos e a paciência frente às adversidades.

À minha querida Bebel, grande mamãe do meu coração, e aos irmãos mais companheiros e amigos, em especial: o querido Carlos Alberto (Piu), a querida boneca Elisa (Lili), o querido Geraldo (Lalado) e o doce querido José do Socorro (Coks). Agradeço muito por todo apoio e amor, por me ensinarem as tradições e o valor do torrão familiar mineiro.

Aos queridos tios Abidon e Celso (*in memmorian*), agradeço por me ensinarem os verdadeiros valores e o aconchego do torrão familiar.

Aos queridos tios e grandes amigos José Carlos e Claudina, aos primos-irmãos Carlos e Rafael, agradeço por todo apoio e carinho desde a minha chegada ao Rio de Janeiro.

À minha querida família Alberton, que me acolheu com muito amor. Agradeço a querida Dona Ivete e aos queridos Aroldo, José Ricardo, João Luís e Carla, Mário e Alessandra, o aconchego e o apoio familiar. Aos pequenos, João, Gabriel, Arthur, Iara e Miguel, não tão pequenos, agradeço a felicidade de ser tia e toda diversão e carinho desta mágica função.

Ao amigo-irmão, grande cúmplice, João Baptista Severo Jr., agradeço pelo grande companheirismo ao longo do mestrado e boa parte do doutorado. Joãozinho, você faz parte do que há de melhor em mim e está para sempre em meu coração; sinto muito a sua falta.

Eu sou a caçula de uma geração de Trogloditas!; sou eternamente grata por todo o apoio, intelectual e emocional de toda a G-130/LMSCP, em especial dos grande amigos: André Luís Alberton (Deco), Diego Prata (Di), Eduardo Lemos (Monstro), Eduardo Lima (Edu), Elisa Barbosa Coutinho, Fabrício Vieira (Fafas Goderinho), Heloísa Sanches (Helôsinha), João Baptista Severo Jr. (Joãozinho), José da Paixão dos Santos (Grande Zuza), Júlio Dutra, Marcio Schwaab (Schwaabão) e Matheus Soares; em mim há um pouco de todos, que compõe o melhor do que eu sou! E aos mais recentes, que também já estão ficando velhos por lá: Bruno Oeschler (Brunão), Cauê Costa (Sr. Coruja), Paula Naomi (Paulinha), Pedro Alijó (Vampman); companheiros fraternos e generosos em convivência, respeito e consideração.

Aos queridos companheiros de outros laboratórios do PEQ, em especial: Carlos Chagas (NUCAT), Claudiney Fernandes de Melo (LABPOL), Prof. Fábio Toniolo (NUCAT), João Paulo Bassin (LABPOL), Paulo André Marinho (LECC), Simone Miyoshi (LADES) e Thaís Farias (LTFD).

Ao querido amigo Fabiano Castoldi, sempre um bom companheiro.

À querida amiga Andressa Nakao, agradeço por toda boa vontade em construir um mundo melhor; obrigada por fazer a diferença sempre!

Aos grandessíssimos companheiros da linda Argentina, conquistados em PLAPIQUI-CONICET/UNS, em especial: Cecília Paulo (Ceci Portuguesa Rubia), Jimena di Maggio (Jime de la Noche), Vanina Estrada (Vani), Sabrina Belén (Belénzita) e Mathias Récupero (Mate); amigos preciosos! Agradeço também à suas famílias que me acolheram com muito carinho e grande generosidade.

Ao casal Prof^a. Andréa Costa e Prof. Esly Costa Jr., eternos mestres; obrigada por terem me iniciado nesta jornada, que agora termina com muito de vocês.

À querida dinda Prof. Dora Baldanza, agradeço por ser um anjo bom na minha e em muitas vidas.

À Prof. Sole, minha querida orientadora, agradeço por todo o apoio e suporte durante minha estadia em Bahía Blanca, inclusive as muitas caronas radicais, as noites de *pancetas y huevos fritos* e todas as soluções latino-américa.

Ao Prof. Arge, meu precioso orientador; dentre as suas inúmeras qualidades, a generosidade em exímia orientação. Agradeço por todo companheirismo e dedicação, por toda a confiança e credibilidade que empenhaste sempre, sobretudo por toda paciência e sabedoria com que conduzieste nosso trabalho, indispensáveis em tantos momentos. Peço desculpas por todas as minhas dificuldades, muitas agora bem superadas, e agradeço por toda a formação adquirida ao seu lado como engenheira, pesquisadora e futura mestra; você é um grande exemplo a ser seguido! E eu que sou para sempre a caçula, quero ser como você quando eu crescer!!!

Ao estimado Andrés Luís Alberton, estimador de parâmetros de corações amigos e fraternos! Esta é com certeza mais uma tese extra oficialmente co-orientada por você, grande discípulo de mestre Schwaab; grande parte do sucesso deste trabalho se deve a sua grande benevolência em transferir com sabedoria e humildade o conhecimento em sua forma mais simples, bela e rica aos demais. Agradeço por todas as “loucas” e divertidas discussões sobre estimação de parâmetros, por todas as aulas de reatores e aplicações termodinâmicas, por todo o curso de formatação e apresentação de tese (inclusive *blofous!*)..., agradeço também por tudo que me ensinou e eu não pude absorver. Reagradeço por todas as pessoas que você vem assistindo com sua generosidade em sua grande e linda jornada de ensino e aprendizado; obrigada muito e sempre.

Ao modeleiro, estimador, reconciliador, ..., e sobretudo grande engenheiro químico “fortranzeiro”, Marcio Schwaab; grande amigo e o exemplo do lar dos Trogloditas. Schwaab, agradeço por toda grande boa influência em minha formação profissional e pessoal. Eu fui e sou uma pessoa privilegiada de compartilhar da sua amizade e ser assistida por sua elevada generosidade. Obrigada por toda paciência despendida em transformar esta engenheira sanitária e ambiental em uma “fortranzeira” com conhecimentos de modelagem e estatística. Reagradeço por todas as pessoas que você vem assistindo com sua generosidade em sua grande e linda jornada de ensino e aprendizado; obrigada muito e sempre.

A todos os professores do PEQ, agradeço e peço a benção, em especial:

- Aos meus sempre orientadores, mestres da vida e do coração: Prof. Enrique Lima, Prof. José Carlos Pinto e Profa. Márcia Dezotti; agradeço por toda a grande participação na minha formação profissional e pessoal, pelas muitas oportunidades concedidas e pela liberdade de escolha sempre; muito obrigada por acreditarem e investirem em mim, enquanto ser humano, acima de tudo.
- Ao grande mestre Prof. Evaristo Biscaia Jr., agradeço por toda grande contribuição a minha formação profissional e pessoal, sobretudo pelo carinho paternal que sempre dedicastes a nossa amizade.
- Ao querido Prof. Príamo Melo, grande companheiro de todas as horas! Obrigada por toda assistência e amizade prestados desde o início da minha chegada ao programa.
- Ao grande Prof. Schmal, agradeço por toda amizade e carinho, e pelas muitas tardes de conversas!
- Ao meu querido orientador acadêmico Prof. Tito Lívio Alves, agradeço pelo acompanhamento ao longo desta jornada.

À Banca de Seminário; agradeço pelas discussões e sugestões ao longo do desenvolvimento desta tese.

Aos queridos alunos da engenharia química da PUC-Rio, em especial: Amanda Brandão, Anita Simões, Paulo Mirilli e Thiago Mota.

Aos queridos alunos da UERJ, em especial: Daniel Pedrosa, Flávia Albuquerque, Isidro Correia, Guilherme Nóbrega, Marco Santos, Raquel Galvão e Larissa Pinheiro.

À Banca Examinadora, agradeço muito as colaborações prestadas a esta tese.

Aos funcionários do PEQ; grandes em amizade, eficientes em apoio e suporte. Completos! Agradeço em especial a querida secretária Vera Lúcia Silva da Cruz, por toda paciência, boa vontade e marcações/desmarcações para a defesa desta tese; obrigada por resolver praticamente tudo por mim, a sua ajuda foi essencial.

À minha adorável amiga Beatriz, a companhia mais agradável de todos os tempos! Sempre sinto saudades!

Agradeço a todos que se fizeram comigo, aos que mesmo ausentes se fizeram presentes, não na duração, mas na intensidade dos momentos.

Ao PEQ-COPPE/UFRJ e a PLAPIQUI-CONICET/UNS, agradeço a infra-estrutura e o suporte.

À Capes e ao CNPq, agradeço o apoio financeiro durante o doutorado.

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

IDENTIFICABILIDADE DE PARÂMETROS: ASPECTOS CONCEITUAIS E DESENVOLVIMENTO DE NOVOS PROCEDIMENTOS

Kese Pontes Freitas Alberton

Dezembro/2013

Orientadores: Argimiro Resende Secchi

María Soledad Díaz

Programa: Engenharia Química

O problema de identificabilidade de parâmetros afeta a maioria dos processos modernos, que empregam modelos complexos para predição. Usualmente, as medidas não são em quantidade e/ou qualidade suficientes à estimação de todos os parâmetros do modelo. Nesta tese, lacunas importantes da literatura quanto à classificação de parâmetros e à caracterização estatística de incertezas são abordadas, apontando a dependência da identificabilidade com a estimativa inicial dos parâmetros como um grande obstáculo. Demonstra-se que o procedimento de fixar um subconjunto de parâmetros e estimar os demais, pode levar a resultados enganosos e discute-se o uso apropriado dos resultados do procedimento de identificabilidade. Por fim, buscando contornar as dificuldades associadas à dependência com a estimativa inicial, dois novos procedimentos foram propostos, com diferentes focos: (i) a avaliação intensiva, cujo objetivo é a análise de todos os parâmetros para estudo de sensibilidade do modelo e planejamento de condições experimentais de interesse, e (ii) a avaliação acelerada, cujo objetivo é permitir a aplicação da identificabilidade em tempo real; ambos demonstraram sucesso em tratar o problema de identificabilidade em modelos complexos com elevado número de parâmetros e medidas experimentais infrequentes/ausentes.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

PARAMETERS IDENTIFIABILITY: CONCEPTUAL ASPECTS AND
DEVELOPMENT OF NEWS PROCEDURES

Kese Pontes Freitas Alberton

December/2013

Advisors: Argimiro Resende Secchi

María Soledad Díaz

Department: Chemical Engineering

The problem of parameters identifiability is present in the most of modern processes, which employ complex models as a key tool for predictions. Usually, as measurements are insufficient in quantity and/or quality enough for the estimation of all the parameters of the model. In this thesis, important aspects of the identifiability problem are discussed, like the statistical characterization of uncertainties, emphasizing the high dependence with initial estimates. It is shown that the estimation of only a subset of parameters can lead to misleading results; and the proper scenario for the use of identifiability is discussed. Last, to avoid the dependence regard to the initial estimates, two new procedures are proposed, with different motivations: (i) the intensive evaluation, whose purpose is the evaluation of all models parameters for model sensitivity investigation and design of experimental regions of interest, and (ii) the accelerated evaluation, which objective is to allow the application of identifiability in real time; both were succeed for treating the identifiability problem in complex models with a large number of parameters and few experimental data.

ÍNDICE

CAPÍTULO 1. INTRODUÇÃO	1
CAPÍTULO 2. REVISÃO BIBLIOGRÁFICA	6
2.1 ESTIMAÇÃO DE PARÂMETROS	7
2.1.1 Incerteza Experimental	8
2.1.2 Incerteza Paramétrica	10
2.1.3 Incerteza da Predição	15
2.2 IDENTIFICABILIDADE DE PARÂMETROS	16
2.2.1 Definições	17
2.2.2 O procedimento de identificabilidade de parâmetros	20
2.3 ANÁLISE DE SENSIBILIDADE	25
2.4 IDENTIFICABILIDADE DE PARÂMETROS VIA ANÁLISE DE SENSIBILIDADE	32
2.4.1 Identificabilidade de parâmetros via análise de sensibilidade local	38
2.4.2 Identificabilidade de parâmetros via análise de sensibilidade global	45
2.5 CONCLUSÕES PARCIAIS	49
CAPÍTULO 3. ESTUDO COMPARATIVO DE MÉTODOS PROPOSTOS NA LITERATURA	51
3.1 ESTRATÉGIAS PARA ABORDAR A IDENTIFICABILIDADE DE PARÂMETROS	52
3.2 ESTUDOS DE CASOS	53
3.2.1 Modelo hipotético 1	56
3.2.2 Modelo hipotético 2	69
3.2.3 Modelo hipotético 3	72
3.2.4 Modelo cinético de crescimento celular	78
3.3 CONCLUSÕES PARCIAIS	87
CAPÍTULO 4. ASPECTOS CONCEITUAIS DO PROBLEMA DE IDENTIFICABILIDADE DE PARÂMETROS	90
4.1 ASPECTOS CONCEITUAIS DA CLASSIFICAÇÃO DOS PARÂMETROS	91
4.2 ASPECTOS CONCEITUAIS DA CARACTERIZAÇÃO ESTATÍSTICA DE INCERTEZAS	101
4.2.1 Propagação da incerteza dos parâmetros não identificáveis para a predição	104
4.2.2 Quando o uso do procedimento de identificabilidade leva a predição confiável	107
4.2.3 Informação desprezada ao fixar o valor dos parâmetros	114

4.3	CONCLUSÕES PARCIAIS	130
CAPÍTULO 5. DESENVOLVIMENTO DE NOVOS PROCEDIMENTOS DE IDENTIFICABILIDADE DE PARÂMETROS		131
5.1	CLASSIFICAÇÃO ROBUSTA DOS PARÂMETROS	134
5.1.1	Procedimento numérico I: Avaliação intensiva dos parâmetros	136
5.1.2	Estudo de caso: modelo das rotas metabólica da E. coli K-12 W3110	141
5.2	ACELERANDO O PROCEDIMENTO DE IDENTIFICABILIDADE	157
5.3	BUSCA BINÁRIA	158
5.4	SELEÇÃO CONJUNTO A CONJUNTO DOS PARÂMETROS	160
5.4.1	Procedimento numérico II: Avaliação acelerada dos parâmetros	161
5.4.2	Estudo de caso 1: modelo hipotético 6	166
5.4.3	Estudo de caso 2: modelo das rotas metabólicas da E. coli K-12 W3110	170
5.5	CONCLUSÕES PARCIAIS	176
CAPÍTULO 6. CONCLUSÕES E SUGESTÕES		178
REFERÊNCIAS		183

CAPÍTULO 1. INTRODUÇÃO

A complexidade dos modernos processos de engenharia associada a dificuldades experimentais frequentemente levam a problemas de estimação de parâmetros mal condicionados. Tais problemas são intratáveis por meio da abordagem clássica da estimação, mas podem ser contornados ajustando o modelo com base apenas nos parâmetros que podem ser estimados a partir dos experimentos disponíveis. Objeto de estudo desta tese, a identificabilidade de parâmetros e sua importância para os processos no cenário atual, são brevemente explorados neste Capítulo 1, que visa à contextualização, à justificativa e à apresentação das principais linhas deste estudo.

Os processos modernos de engenharia são descritos por modelos complexos, cuja estrutura matemática usualmente apresenta número elevado de parâmetros. É sabido que a estimação de parâmetros consistente depende da confiabilidade das medidas experimentais disponíveis. Todavia, além dos erros inerentes ao procedimento de medição, também é muito comum a ausência ou a baixa frequência de informação. Portanto, um dos problemas mais comuns enfrentados na obtenção de modelos adequados é o fato de que as medidas experimentais disponíveis não são em quantidade e qualidade suficientes para permitir a estimação dos parâmetros.

Considerando as dificuldades de obtenção de informação sobre os sistemas, as etapas de coleta e de análise são elementos fundamentais. Um exemplo que representa bem este cenário são os estudos ambientais. Nestes estudos, geralmente o ponto de coleta é distante e o acesso, bem como a qualidade das amostras, podem ser fortemente dependentes das condições edafoclimáticas. Além disto, os objetos chaves de estudos ambientais podem levar horas ou até dias para serem analisados (e.g., a demanda bioquímica de oxigênio – DBO e a atividade de microorganismos). Tais fatos aumentam consideravelmente o tempo e o custo destes estudos.

É importante observar que, embora o mercado disponha de sensores de muitos tipos para inferência de propriedades físicas e químicas, o mesmo não ocorre para propriedades bioquímicas e para os chamados parâmetros intangíveis (e.g., propriedades organolépticas). Voltando ao exemplo dos estudos ambientais, percebe-se que ainda não estão disponíveis sensores capazes de inferir a composição do chorume de aterro sanitário ou do efluente final de uma estação de tratamento de esgoto, dada a complexidade destes compostos que necessitam de muitas análises para serem caracterizados.

Um ambiente prático para averiguação destes problemas são as plantas industriais. Embora a disponibilidade de dados aparentemente seja grande, geralmente não é possível variar as condições de tomada das observações, visto as rígidas restrições de custos e de qualidade dos produtos, bem como as metas de produção e as normas ambientais. Além disso, alguns sensores podem não estar disponíveis ou possuem tempo de análise superior ao tempo de amostragem da planta; neste caso, são empregados modelos inferenciais, que também possuem parâmetros, usuais para a obtenção de algumas variáveis decisivas ao controle de qualidade do produto e as estratégias de controle avançado e de otimização. Tais problemas afetam mesmo os processos plenamente estabelecidos, tais como: (i) o refino de petróleo - caracterização

da composição do diesel e outros componentes (FREITAS, 2009), (ii) a fabricação de celulose e papel - obtenção do grau de qualidade número *kappa* (FREITAS et al., 2006), e (iii) a produção de polímeros - determinação do índice de fluidez (PRATA, 2009).

Outro problema grave que afeta a obtenção de modelos adequados é a não-linearidade comum à maioria dos modelos, que pode implicar em dificuldades adicionais como a multiplicidade de soluções, a alta correlação entre os parâmetros e a presença de parâmetros não significativos (SCHWAAB, 2007).

Neste contexto, a estimação de todos os parâmetros do modelo pode levar a valores ruidosos, sem significado físico e/ou estatístico, ou até mesmo tornar-se inviável. Uma alternativa para contornar o problema de estimação mal condicionado é estimar os parâmetros mais influentes do modelo, mantendo os valores dos demais parâmetros na estimativa inicial; este procedimento é comumente tratado no contexto da identificabilidade de parâmetros.

Basicamente, a identificabilidade de parâmetros busca determinar o conjunto de parâmetros do modelo que podem ser consistentemente estimados a partir das observações experimentais disponíveis. Estes procedimentos levam em conta a influência dos parâmetros selecionados sobre a predição do modelo e/ou a correlação com os demais parâmetros. Desta forma, a identificabilidade apresenta duas características importantes (KRAVARIS et al., 2012): (i) a interpretabilidade - definição dos parâmetros que são mais importantes para o comportamento do modelo, e (ii) a simplificação - redução do número de variáveis de decisão no problema de otimização.

Várias abordagens de identificabilidade vêm sendo propostas na literatura. O senso comum é o uso da análise de sensibilidade, como uma escolha natural, por ser uma ferramenta valiosa para avaliar a influência dos parâmetros sobre a predição (SECCHI et al., 2006). Na última década, vários procedimentos foram desenvolvidos baseados na análise de sensibilidade local (BRUN et al., 2001; SANDINK et al., 2001; YAO et al., 2003; LI et al., 2004; SECCHI et al., 2006; SUN e HAHN, 2006; CHU e HAHN, 2007; LUND e FOSS, 2008), e na menos usual análise de sensibilidade global (CHU et al., 2011). Na maioria destas metodologias, os parâmetros são selecionados a partir do mais estimável para o menos estimável, com base na estrutura do modelo, nas medidas experimentais disponíveis e na estimativa inicial dos parâmetros (THOMPSON et al., 2009). Em se tratando de modelos complexos, tais análises não são uma tarefa trivial (YUE et al., 2006).

Especialmente com relação a modelos complexos, existem lacunas importantes a serem preenchidas no estudo da identificabilidade. O problema da dependência dos procedimentos de identificabilidade com a estimativa inicial dos parâmetros é pouco abordado na literatura (OMLIN e REICHERT, 1999; CHU et al., 2011). Também há o problema associado ao esforço computacional dos procedimentos, cuja seleção de parâmetros é baseada na avaliação um a um dos parâmetros. A ausência de procedimentos que avaliam todos os parâmetros do modelo ante à classificação também é uma questão relevante. Mais além, a classificação dos parâmetros quanto à identificabilidade ainda é um tema controverso na literatura, que apresenta variados tipos de definições ambíguas (LEE e HERGET, 1975; GREWAL e GLOVER, 1976; REID, 1979; COBELLI e DISTEFANO, 1980; RUSNAK et al., 1983; VAJDA et al., 1989; LJUNG e GLAD, 1994; AUDOLY et al., 2001; RAUE et al., 2009).

Assim, esta tese apresenta a identificabilidade de parâmetros como tema de estudo. O principal objetivo perseguido é o desenvolvimento de procedimentos de identificabilidade capazes de tratar as dificuldades presentes na estimação de parâmetros de modelos complexos, assegurando previsões estatisticamente confiáveis. Dentre as características desejadas destes novos procedimentos, incluem-se a robustez à estimativa inicial de parâmetros e a convergência rápida.

Neste sentido, esta tese foi elaborada ao longo das três principais linhas de ação:

1. Estudo e comparação de métodos de identificabilidade propostos na literatura.
2. Estudo dos aspectos conceituais do problema de identificabilidade de parâmetros.
3. Desenvolvimento de novos procedimentos de identificabilidade.

O tema apresenta pronta aplicação em problemas das engenharias química e ambiental, em que são muitos os desafios a se considerar quando se trata da identificabilidade de parâmetros, principalmente o desempenho dos procedimentos frente aos cenários reais, tipicamente compostos por modelos complexos contendo elevado número de parâmetros e experimentos escassos.

A presente tese está organizada em seis capítulos, incluindo esta introdução. O próximo capítulo, Capítulo 2, intitulado revisão bibliográfica, trata da apresentação e discussão de conceitos reportados pela literatura, a serem empregados nesta tese. O

Capítulo 3, intitulado estudo comparativo de métodos propostos na literatura, ilustra o princípio de funcionamento e analisa o desempenho de três métodos populares com distintas metodologias (índices de sensibilidade global, ortogonalização, análise dos componentes principais) frente a diferentes estudos de casos. O Capítulo 4, intitulado aspectos conceituais do problema de identificabilidade de parâmetros, esclarece a classificação de parâmetros à luz da estatística e aborda as dificuldades da caracterização estatística de incertezas, definindo o cenário em que o uso de tais procedimentos pode levar a modelos apropriados para a predição. No Capítulo 5, intitulado desenvolvimento de novos procedimentos de identificabilidade de parâmetros, dois novos algoritmos são propostos com diferentes focos, a avaliação de todos os parâmetros ante à classificação e a avaliação de parâmetros com reduzido esforço computacional. Finalmente, o Capítulo 6 apresenta as principais conclusões obtidas ao longo desta tese, e ainda outras possíveis propostas a explorar no tema. As referências são apresentadas ao final desta tese.

CAPÍTULO 2. REVISÃO BIBLIOGRÁFICA

A identificabilidade de parâmetros é um tema complexo, que envolve várias áreas do conhecimento, motivada na dificuldade de estimação e baseada direta ou indiretamente na análise de sensibilidade. Para uma melhor compreensão do tema tratado nesta tese, alguns conceitos teóricos são apresentados e uma investigação sobre o seu estado da arte é conduzida neste Capítulo 2.

2.1 ESTIMAÇÃO DE PARÂMETROS

Os modelos em geral são compostos basicamente por três elementos: as variáveis independentes $x \in X \subseteq \mathfrak{R}^{NX}$, as variáveis dependentes $y \in Y \subseteq \mathfrak{R}^{NY}$ e os parâmetros $\theta \in \Theta \subseteq \mathfrak{R}^{NP}$. A relação funcional entre as variáveis independentes e as variáveis dependentes compõe a estrutura do modelo. Já os parâmetros são certos valores ou constantes, presentes nestas relações funcionais (e.g., a energia de ativação de uma reação química).

Embora muitas vezes não possam ser obtidos via medição, os parâmetros podem ser inferidos por meio das observações experimentais. O procedimento de inferência dos parâmetros de um modelo é chamado de estimação de parâmetros e consiste em ajustar os valores dos parâmetros de tal forma que as previsões do modelo sejam as mais próximas possíveis dos valores medidos experimentalmente, respeitadas as incertezas características da medição. Este é tipicamente um problema de otimização, em que estão presentes a função objetivo e o modelo do processo, que podem estar sujeitos a restrições adicionais.

O interesse do procedimento de estimação de parâmetros se concentra nos valores dos parâmetros que minimizam a função objetivo. A função objetivo é uma métrica que quantifica a diferença entre o valor medido experimentalmente e a previsão do modelo matemático. SCHWAAB e PINTO (2007) apresentam as características matemáticas básicas necessárias para uma função ser denominada função objetivo. De acordo com SCHWAAB (2007), a função objetivo mais empregada é a “mínimos quadrados”, apesar de ser estatisticamente correta apenas para casos muito restritos. Uma forma estatisticamente consistente de abordar o problema é o uso do princípio de máxima verossimilhança, que pode ser aplicado a qualquer distribuição de probabilidade (CHERNOFF, 1953; SCHWAAB e PINTO, 2007).

O conceito de máxima verossimilhança está baseado em pressupostos relativamente simples e permite a análise rigorosa de qualquer problema experimental, desde que se conheça de forma apropriada como se comportam os erros de medição na região experimental (SCHWAAB e PINTO, 2007). Desta forma, admitindo que o modelo seja perfeito e que os experimentos são feitos de modo que seja plausível admitir que os resultados experimentais obtidos sejam os mais prováveis, parece razoável variar os parâmetros de forma a maximizar a probabilidade de se observar os resultados experimentais obtidos (BARD, 1974). Os pressupostos para o uso adequado

do princípio de máxima verossimilhança estão detalhados em BARD (1974) e SCHWAAB e PINTO (2007).

De acordo com o princípio de máxima verossimilhança, admitindo que as variáveis sigam a distribuição normal de probabilidade e são não correlacionadas e que as variáveis independentes estão livres de erros, a função objetivo pode ser descrita matematicamente pela expressão apresentada na Equação (E 2.1), usualmente conhecida como “mínimos quadrados ponderados” (BARD, 1974; SCHWAAB e PINTO, 2007).

$$(E 2.1) \quad FObj = \sum_{i=1}^N \sum_{j=1}^{NY} \frac{(y_{ij}^E - y_{ij}^C)^2}{v_{Y_j}}$$

em que $FObj$ representa a função objetivo, N o número de experimentos e v_{Y_j} a variância experimental (descrita no próximo item desta seção); os sobrescritos E e C na variável dependente y indicam, respectivamente, experimental e calculado a partir do modelo; os índices i e j estão relacionados ao número de experimento e ao número de variáveis independentes, respectivamente.

É preciso estar atento ao fato de que, qualquer que seja a função objetivo empregada, a confiabilidade dos parâmetros é influenciada pela acurácia e precisão das medidas experimentais, que inerentemente contêm erros. Desta forma, a definição adequada das incertezas experimentais constitui o primeiro passo para que o procedimento de estimação de parâmetros forneça resultados confiáveis (SCHWAAB, 2007).

2.1.1 Incerteza Experimental

As incertezas experimentais são fundamentais no estudo de sistemas, revelando importantes aspectos de comportamento (SCHWAAB, 2007). Por exemplo, a análise criteriosa destas incertezas pode ser útil para fornecer informações sobre mecanismos de reações (LARENTIS *et al.*, 2003) e para avaliação de catalisadores de craqueamento catalítico (CERQUEIRA *et al.*, 1999; RAWET *et al.*, 2001).

Considere o par de variáveis z_i e z_j pertencentes a um conjunto $Z \subseteq \mathfrak{R}^{NZ}$ que engloba as variáveis independentes e dependentes (X e Y). Dado um conjunto de NR réplicas de experimentos, a média e a covariância podem ser calculadas por meio das

Equações (E 2.2) e (E 2.3) (BARD, 1974; SCHWAAB e PINTO, 2007). As incertezas experimentais são dadas pela matriz de covariância experimental, $V_Z = [v_{z_i, z_j}]$, que caracteriza a variabilidade do conjunto experimental em torno do valor médio \bar{z} .

$$(E 2.2) \quad \bar{z} = \frac{1}{N} \sum_{k=1}^{NR} z_k ,$$

$$(E 2.3) \quad v_{z_i, z_j} = \frac{1}{NR - 1} \sum_{k=1}^{NR} (z_{i,k} - \bar{z}_i)(z_{j,k} - \bar{z}_j),$$

para $i = j$ denota a variância experimental.

Frequentemente, a matriz de covariância dos erros experimentais é total ou parcialmente desconhecida (ROMAGNOLI e SÁNCHEZ, 1999). Ainda que conhecida, mudanças nas condições experimentais podem causar alterações nesta matriz, visto que alguns dispositivos de medição podem apresentar valores de variância de acordo com a faixa experimental analisada (MORAD *et al.*, 1999; ROMAGNOLI e SÁNCHEZ, 1999). Um exemplo são as análises químicas (e.g., DBO – Demanda Bioquímica de Oxigênio e DQO – Demanda Química de Oxigênio), em que muitas vezes a incerteza é proporcional ao valor da medida experimental. Portanto, ainda que corretamente caracterizada, a covariância dos erros experimentais deve ser avaliada constantemente.

Para contornar as dificuldades de caracterizar a matriz de covariância dos erros experimentais, usualmente considera-se que as incertezas experimentais são independentes. Neste caso, esta matriz torna-se diagonal e seus elementos são dados pela matriz de variância dos erros experimentais. Todavia, considerações inadequadas sobre as covariâncias da incerteza experimental podem levar a conclusões equivocadas sobre os valores dos parâmetros e a adequabilidade do modelo (SANTOS e PINTO, 1998). Por outro lado, quando se considera que as incertezas experimentais são dependentes, uma dificuldade que pode estar presente é o mau condicionamento da matriz de covariância dos erros experimentais. O mau condicionamento de uma matriz está associado à proximidade da singularidade, situação em que a matriz torna-se não inversível. Isto é particularmente importante, quando se utiliza a inversa da matriz de covariância experimental para ponderar as variáveis na função objetivo. Além disso, a inversa desta matriz é empregada no cálculo da incerteza paramétrica, fundamental para avaliação dos parâmetros estimados. Portanto, o mau condicionamento desta matriz

pode inserir dificuldades adicionais à minimização da função objetivo, além de levar à estimação de parâmetros sem qualquer significado físico e/ou estatístico (SANTOS e PINTO, 1998).

2.1.2 Incerteza Paramétrica

Uma vez que o procedimento de estimação de parâmetros é baseado em observações experimentais sujeitas a incertezas, torna-se necessário avaliar a qualidade dos parâmetros estimados. A incerteza paramétrica é dada pela matriz de covariância dos parâmetros, de forma que a qualidade dos parâmetros será tão melhor quanto menor se apresentarem os termos da matriz V_{Θ} . O cálculo ideal desta grandeza estatística é baseado no cenário hipotético em que um conjunto de N experimentos pudesse ser realizado infinitas vezes, de forma que o parâmetro verdadeiro Θ_* fosse considerado valor médio dos parâmetros estimados $\bar{\Theta}$ a partir das infinitas repetições dos conjuntos de N experimentos; como ilustra a Figura 2.1.

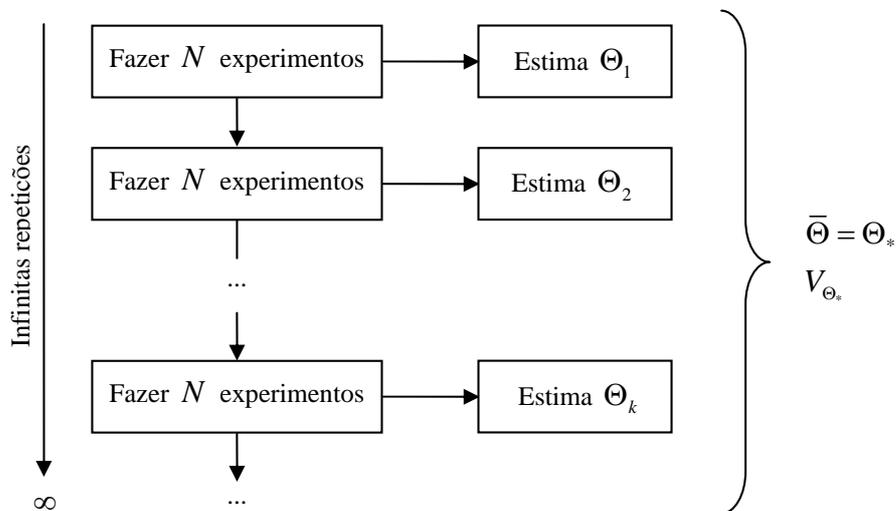


Figura 2.1 – Procedimento hipotético para determinação da média e da matriz de covariância paramétrica a partir de um conjunto de N experimentos.

Na prática, não é possível a realização de N experimentos infinitas vezes. Desta forma, a obtenção da matriz de covariância é baseada na suposição de que os valores estimados Θ' são os valores verdadeiros Θ_* dos parâmetros e que a incerteza paramétrica $\Delta\Theta$ é uma transformação linear da incerteza experimental ΔY ; como ilustra a Figura 2.2.

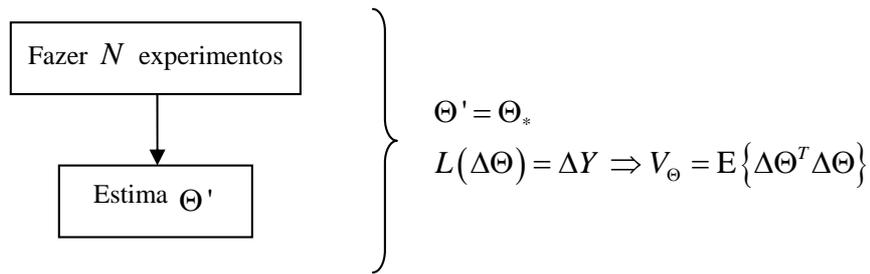


Figura 2.2 – Procedimento prático para inferência da média e da matriz de covariância dos parâmetros a partir de um conjunto de N experimentos.

Claramente, a matriz de covariância dos parâmetros não é conhecida com exatidão, já que é determinada empregando uma quantidade limitada de observações experimentais que representam apenas uma amostra de todos os possíveis experimentos (SCHWAAB, 2007).

O cálculo da matriz de covariância dos parâmetros V_{Θ} pode ser realizado por meio da Equação (E 2.4) (BARD, 1974; SCHWAAB e PINTO, 2007)

$$(E\ 2.4) \quad V_{\Theta} = (B^T V_Z^{-1} B)^{-1},$$

em que $B = \partial Z / \partial \Theta$ representa a matriz de sensibilidade, muito utilizada para avaliar a influência dos parâmetros sobre a predição do modelo matemático.

Considerando que não há erros nas variáveis de entrada X , a matriz de covariância dos parâmetros é reduzida para a forma da Equação (E 2.5) (BARD, 1974; SCHWAAB e PINTO, 2007):

$$(E\ 2.5) \quad V_{\Theta} = \left(\left(\frac{\partial Y}{\partial \Theta} \right)^T V_Y^{-1} \frac{\partial Y}{\partial \Theta} \right)^{-1}.$$

Frequentemente adota-se a inversa desta matriz, comumente conhecida como matriz de informação de Fisher, $FIM = (V_{\Theta})^{-1}$. Desta forma, quanto maiores são as métricas da matriz FIM , maior a informação do sistema. O uso direto da matriz de informação de Fisher tem como objetivo evitar o procedimento de sua inversão, que pode ser proibitivo quando esta matriz for mal condicionada. Nesta situação, a

observação dos termos da matriz *FIM* pode fornecer uma medida inadequada de informação dos parâmetros, pois neste caso possivelmente existe dependência linear entre os mesmos.

As matrizes de caracterização de incerteza paramétrica permitem obter informações relevantes dos parâmetros, tais como a definição de intervalos de confiança, o grau de significância e a indesejável correlação paramétrica.

Geralmente, a obtenção do intervalo de confiança admite que o erro paramétrico possua distribuição normal de probabilidades, considerando que os erros de medida também possuam esta distribuição. Embora seja garantida apenas para modelos lineares nos parâmetros, esta hipótese pode consistir em uma boa aproximação para modelos não-lineares nos parâmetros quando se trata de sistemas onde o erro experimental é pequeno. Desta forma, usualmente se expressa a faixa de valores possíveis para o parâmetro θ_i por meio da Equação (E 2.6) (BARD, 1974; SCHWAAB e PINTO, 2007):

$$(E\ 2.6) \quad \theta_i' - t \cdot \sigma_{\theta_i} < \theta_i < \theta_i' + t \cdot \sigma_{\theta_i},$$

em que θ_i' é o valor estimado do parâmetro no menor valor da função objetivo, t é o valor obtido a partir da distribuição normal (ou t de Student) de acordo com o grau de confiança especificado e o número de graus de liberdade.

Portanto, o intervalo de confiança pode ser usado para definir o grau de significância dos parâmetros, sendo que um parâmetro é dito não significativo se o intervalo admite valores positivos e negativos, podendo assumir o valor zero. Há uma variedade de expressões matemáticas empregadas para o cálculo do intervalo de confiança dos parâmetros (SCHWAAB e PINTO, 2007), sendo que a maioria destas expressões utiliza os desvios padrões dos parâmetros. Embora a matriz de covariância dos parâmetros seja sempre cheia, apenas a diagonal (a variância dos parâmetros) é empregada para este cálculo, de modo que os intervalos de confiança não possuem informação sobre a correlação paramétrica, o que pode levar a uma avaliação inadequada das incertezas para modelos multi-paramétricos (SCHWAAB, 2007).

De grande importância, a correlação paramétrica é uma grandeza estatística que expressa a interação entre os efeitos dos parâmetros, tornando a faixa válida para um parâmetro dependente dos valores dos demais parâmetros. Embora esteja quase

sempre presente, a correlação paramétrica é indesejável podendo causar problemas numéricos significativos durante o procedimento de estimação de parâmetros, tais como a ineficiência da minimização da função objetivo (ESPIE e MACCHIETTO, 1988) e a estimativa de parâmetros com baixa significância estatística (WATTS, 1994). A interação entre os parâmetros pode ser ocasionada por três fatores distintos: os experimentos mal planejados, a estrutura inadequada do modelo e a característica não-linear do modelo (SCHWAAB, 2007). Os primeiros fatores podem ser evitados, em ordem, utilizando-se técnicas de planejamento de experimentos e de re-parametrização do modelo. Comum à maioria dos modelos matemáticos da engenharia química, o terceiro fator também pode ser tratado com técnicas de re-parametrização, mas não pode ser totalmente evitado.

Uma vez que o intervalo de confiança dos parâmetros não leva em conta a correlação paramétrica, outra forma utilizada para avaliar o conjunto de parâmetros é a região de confiança. A região de confiança dos parâmetros pode ser definida como o conjunto de valores dos parâmetros que descrevem adequadamente as observações experimentais com certa precisão estatística. O conceito de região de confiança advém da consideração de que a estimação dos parâmetros é feita com base em observações experimentais finitas que inerentemente contêm erros, não sendo possível obter os valores corretos dos parâmetros, mas sim uma região de valores admissíveis. Definida pelo ponto de mínimo da função objetivo e pela matriz de covariância paramétrica, para modelos lineares, a região de confiança se apresenta simétrica ao valor do parâmetro estimado facilitando sua interpretação geométrica. Deste modo, considerando que a incerteza paramétrica segue a distribuição normal de probabilidades, admitindo válida aproximação linear do modelo, a região de confiança pode ser calculada por meio da seguinte expressão (BARD, 1974; SCHWAAB e PINTO, 2007):

$$(E\ 2.7) \quad (\Theta - \Theta')^T V_{\Theta}^{-1} (\Theta - \Theta') < \chi_{NP, \alpha}^2,$$

em que $\chi_{NP, \alpha}^2$ representa o limite superior da função chi-quadrado para NP graus de liberdade considerando nível de confiança α .

A forma geométrica da região de confiança obtida pela Equação (E 2.7) é um elipsóide, o que é válido somente para condições restritas, como quando os modelos são lineares ou quando os erros experimentais podem ser considerados pequenos; se esta

elipse é inclinada, os termos fora da diagonal principal de V_{Θ} são significativos para a representação apropriada da região de confiança, indicando a existência de correlação entre os parâmetros. Em se tratando de modelos não-lineares, a geometria da região de confiança pode variar bastante, por vezes assumindo formas bastante complexas (SCHWAAB *et al.*, 2008).

A matriz de covariância experimental é usualmente obtida a partir de uma quantidade resumida de experimentos, sendo bastante difícil a caracterização rigorosa da matriz de covariância dos parâmetros. Desta forma, para a obtenção de regiões de confiança estatisticamente mais rigorosas utiliza-se a Equação (E 2.8) (BARD, 1974; SCHWAAB e PINTO, 2007).

$$(E\ 2.8) \quad (\Theta - \Theta')^T V_{\Theta}^{-1} (\Theta - \Theta') < \frac{NP}{N - NP} F_{NP, N - NP}^{\alpha},$$

em que $F_{NP, N - NP}^{\alpha}$ é o limite superior da distribuição F de Fisher, fornecidos NP e $N - NP$ graus de liberdade considerando grau de confiança α .

Vale ressaltar que a validade das conclusões obtidas com bases nas regiões de confiança obtidas por meio das Equações (E 2.7) e (E 2.8) depende da qualidade da aproximação linear do modelo e, conseqüentemente, da aproximação quadrática da função objetivo. Estas regiões de confiança consideram a interação entre os parâmetros, informação indisponível no intervalo de confiança do parâmetro.

A Figura 2.3 ilustra algumas formas geométricas da incerteza paramétrica para um modelo linear hipotético.

De acordo com a Figura 2.3, a forma geométrica da região de confiança dos parâmetros está diretamente relacionada à correlação paramétrica. A correlação entre os parâmetros dificulta a análise dos resultados, uma vez que a faixa válida para um parâmetro é dependente dos valores dos demais parâmetros, sendo inadequada a representação dos valores dos parâmetros por meio do intervalo de confiança nestes casos. Em modelos lineares a região de confiança pode ser associada à geometria da esfera, na ausência de correlação de paramétrica, ou a geometria elíptica, na presença de correlação paramétrica. Mas quando a região de confiança é delimitada por meio do intervalo de confiança dos parâmetros, sua forma é retangular e, portanto não fornece informação sobre a correlação entre os parâmetros (SCHWAAB, 2007).

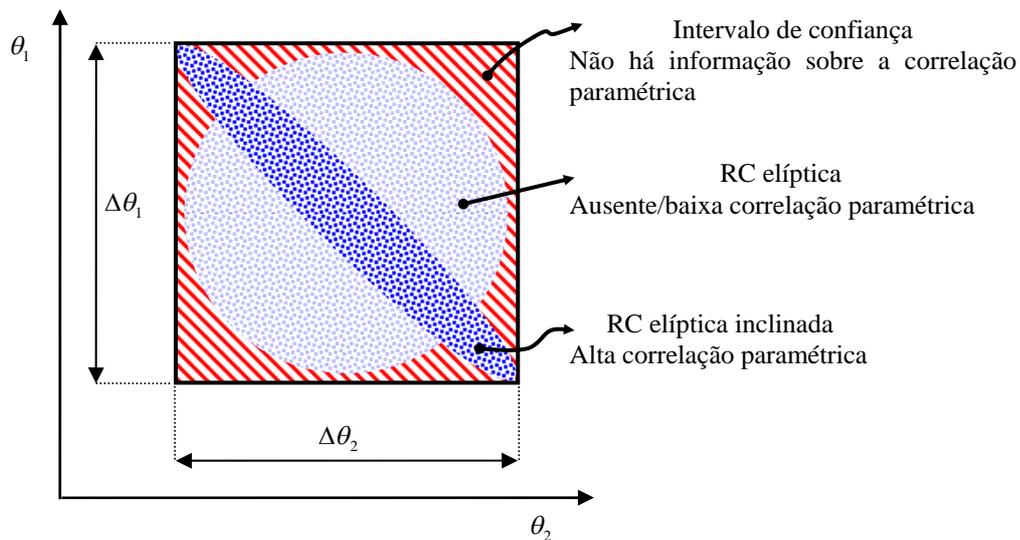


Figura 2.3 – Algumas formas geométricas possíveis da incerteza paramétrica considerando um modelo linear hipotético.

Definições e discussões sobre a utilização de intervalos e de regiões de confiança de parâmetros em maiores detalhes podem ser encontradas em DRAPER e SMITH (1994), SCHWAAB (2005) e SCHWAAB *et al.* (2008).

2.1.3 Incerteza da Predição

Admitindo o modelo matemático perfeito, a confiabilidade da predição depende da qualidade dos experimentos e dos parâmetros. Uma vez que os parâmetros são estimados com base em dados experimentais, a qualidade dos experimentos e a correta caracterização da matriz de covariância experimental são fatores fundamentais.

A caracterização dos erros de predição é dada pela matriz de covariância de predição, que pode ser obtida por meio das Equações (E 2.9) e (E 2.10) (BARD, 1974; SCHWAAB e PINTO, 2007).

$$(E\ 2.9) \quad V_{\hat{y}} = BV_{\theta}B^T + V_Y,$$

$$(E\ 2.10) \quad V_{\hat{y}} = B\left(B^T(V_Y)^{-1}B\right)^{-1}B^T + V_Y.$$

As Equações (E 2.9) e (E 2.10) descrevem matematicamente a propagação da incerteza experimental para a predição. Esta propagação é ilustrada pela Figura 2.4.

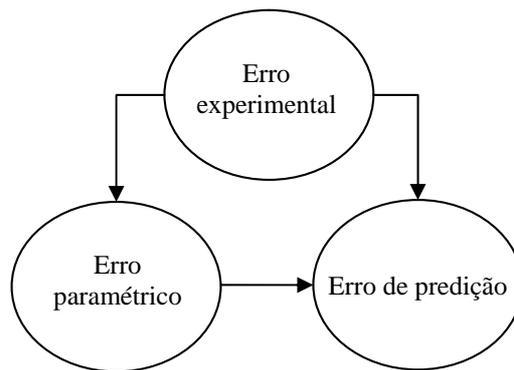


Figura 2.4 - Propagação da incerteza experimental.

Na Figura 2.4, pode-se observar que ao admitir o modelo perfeito, os experimentos passam a ser os elementos fontes de incerteza, de modo que a qualidade da informação experimental é essencial para a confiabilidade da predição. Sabe-se que observações experimentais inerentemente contêm erros, sendo natural admitir que estes erros sejam aleatórios, normalmente distribuídos, não correlacionados, com média nula e variância conhecida. Em um caso mais grave, estas colocações nem sempre estão de acordo com a realidade de operação do processo, implicando na presença de erros grosseiros e/ou que não seguem a distribuição normal, o que pode invalidar a base estatística admitida para as variáveis envolvidas, levando à obtenção de estimativas errôneas para as variáveis e os parâmetros e, conseqüentemente, prejudicando a estimativa da incerteza da predição.

2.2 IDENTIFICABILIDADE DE PARÂMETROS

O problema de estimação mal condicionada pode ser contornado ajustando o modelo com base apenas nos parâmetros mais significativos e menos correlacionados, determinados por um procedimento comumente conhecido como identificabilidade de parâmetros.

Basicamente, a identificabilidade de parâmetros consiste em um procedimento numérico empregado para a seleção dos parâmetros do modelo que apresentam maior potencial de estimação, também chamados parâmetros estimáveis ou mais comumente identificáveis, e tem como objetivo principal definir o conjunto de valores únicos dos parâmetros que pode melhor descrever o sistema.

O conceito de unicidade dos valores dos parâmetros é necessário, visto que pode ser possível inferir valores de parâmetros para descrição de observações experimentais a uma dada condição, mas sem validade para descrever o sistema; o que pode tornar o modelo pouco confiável para previsões em outras condições experimentais.

Em meados do século passado, a definição de identificabilidade de parâmetros e seus termos relativos presentes na literatura técnico-científica estavam repletos de contradições, equívocos e ambiguidades. Nas décadas de 70 e 80, dada a multiplicidade de termos e jargões, COBELLI e DISTEFANO (1980) e outros autores (NGUYEN e WOOD, 1982; JACQUEZ, 1987) buscavam esclarecer e definir formalmente o vocabulário desta área.

2.2.1 Definições

Inicialmente, é importante apresentar os conceitos de identificação de sistemas e identificabilidade de parâmetros. Antes, porém, propõem-se as seguintes definições básicas para este estudo:

- (i). Sistema. Define-se sistema a porção do universo delimitada por fronteiras, regido por leis naturais (físicas, químicas e/ou outras).
- (ii). Modelo. Denomina-se modelo a descrição lógico-matemática das variáveis do sistema, deduzido a partir de leis naturais (fenomenológico) ou obtido a partir de suposições arbitrárias (empírico ou empírico-fenomenológico).

A identificação de sistemas é a elaboração da estrutura lógico-matemática do modelo e a estimação dos parâmetros a partir dos dados experimentais (COBELLI e DISTEFANO, 1980). Já o termo identificabilidade de parâmetros trata da seleção dos parâmetros do modelo que podem ser estimados, classificada em dois tipos básicos (COBELLI e DISTEFANO, 1980): (i) estrutural - procedimentos baseados apenas na estrutura do modelo e (ii) prática - procedimentos baseados nas observações experimentais disponíveis.

Uma vez que os procedimentos de identificabilidade estrutural investigam a consistência da estrutura lógica-matemática com relação aos parâmetros, as dificuldades de aplicação para tais procedimentos se eleva bastante e pode até tornar-se inviável de acordo com o aumento da complexidade do modelo. Embora seja necessária, a

identificabilidade estrutural não é uma condição suficiente para a estimação, visto que o potencial de estimação dos parâmetros também pode ser afetado por outros fatores como a quantidade e qualidade dos dados experimentais disponíveis. Desta forma, o procedimento de identificabilidade estrutural pode levar à seleção de parâmetros que podem ser estimados apenas em condições experimentais muito específicas e/ou irrealizáveis, resultando em um problema de estimação mal condicionado.

A proposta da identificabilidade prática é selecionar os parâmetros de acordo com a observação experimental disponível, como apresentado pela Figura 2.5. Entretanto, além das informações experimentais, a identificabilidade prática requer informações a respeito dos valores dos parâmetros e/ou suas características, frequentemente indisponíveis. Infelizmente, boas estimativas iniciais para os valores de parâmetros são raramente conhecidas, sendo usual o uso de valores obtidos de sistemas similares reportados na literatura ou mais comumente adotados arbitrariamente (KOU et al., 2005a; KOU et al., 2005b). Assim, embora seja uma discussão negligenciada na literatura, é importante enfatizar que os procedimentos de identificabilidade prática também podem resultar em problemas de estimação de parâmetros mal condicionados.

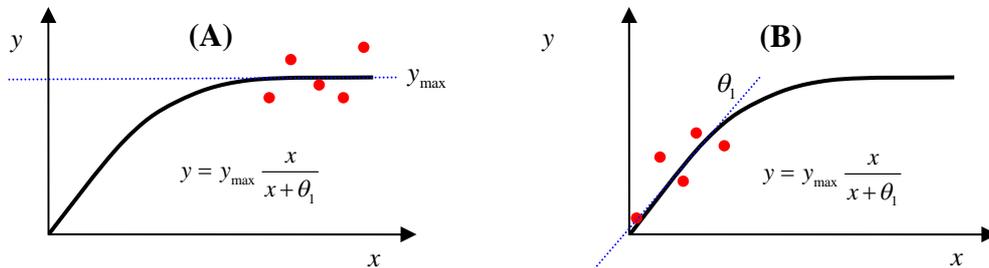


Figura 2.5 - Identificabilidade prática de parâmetros.

A Figura 2.5 apresenta a idéia da identificabilidade prática de parâmetros quando o conjunto de experimentos disponíveis permite a estimação consistente de apenas um dentre os dois parâmetros do modelo. De acordo com Figura 2.5A, mesmo quando o grau de liberdade do sistema permite a estimação, o conjunto de dados experimentais não possui informação para estimação de todos os parâmetros, apenas y_{\max} é identificável; o contrário ocorre na Figura 2.5B, onde apenas θ_1 é identificável. Fica claro que apesar de quantitativamente se dispor de informação para a estimação de todos os parâmetros, qualitativamente esta informação pode não levar a estimação de todos os parâmetros.

Nas últimas décadas, a maioria dos trabalhos reportados na literatura está relacionada à identificabilidade prática (e.g.: WEIJERS e VANROLLEGHEM, 1997; SANDINK et al., 2001; BRUN et al., 2001; YAO et al., 2003; LI et al., 2004; SECCHI et al., 2006; SUN e HAHN, 2006; CHU e HAHN, 2007; LUND e FOSS, 2008; CHU et al., 2011; e outros) devido à elevada complexidade dos modelos atuais que dificulta a análise da identificabilidade estrutural e às dificuldades de se obter a informação experimental com conteúdo para a estimação. Todavia, é pouco comum o uso dos termos identificabilidade prática ou identificabilidade estrutural, que fica restrito à definição da abordagem de identificabilidade empregada, sendo mais adotada a forma genérica de identificabilidade de parâmetros.

As definições a seguir estão presentes na maior parte dos estudos de identificabilidade de parâmetros, de forma bastante similar (COBELLI e DISTEFANO, 1980; AUDOLY *et al.*, 2001; SACCOMANI *et al.*, 2003). Admitindo as hipóteses do modelo perfeito e dos experimentos bem feitos, os parâmetros θ_i podem ser classificados em três tipos quanto à identificabilidade (COBELLI e DISTEFANO, 1980):

- (i). Identificáveis. O parâmetro θ_i é dito identificável se existe um número finito de soluções do modelo para θ_i .
- (ii). Unicamente identificáveis. O parâmetro θ_i é dito unicamente identificável se existe uma única solução do modelo para θ_i .
- (iii). Não-identificáveis. O parâmetro θ_i é dito não-identificável se existe um número infinito de soluções do modelo para θ_i .

Outra classificação bastante empregada na literatura é a identificabilidade global (COBELLI e DISTEFANO, 1980; RUSNAK et al., 1983; AUDOLY *et al.*, 2001; SACCOMANI *et al.*, 2010). De acordo com esta classificação, análoga à anterior, os parâmetros podem ser classificados como localmente identificáveis, globalmente identificáveis e não-identificáveis, respectivamente.

Sob outra perspectiva, GREWAL e GLOVER (1976) apresentaram a classificação de identificabilidade dos parâmetros em termos da distinguibilidade da variável dependente. Nesta proposta de classificação, observam-se as características

preditivas do modelo em certa região experimental ao invés da estimabilidade de parâmetros a partir de observações experimentais.

Equivalente ao conceito proposto por GREWAL e GLOVER (1976), REID (1979) introduziu a noção de identificabilidade em termos da matriz de sensibilidade avaliada em um valor nominal Θ_0 . Para este autor, uma vez que os valores nominais dos parâmetros estão disponíveis, a questão da identificabilidade pode ser abordada realizando pequenas mudanças nestes valores e observando as mudanças que podem ser distinguidas no valor nominal da predição.

Além das classificações citadas, várias outras definições são dadas na literatura (e.g.: SACCOMANI *et al.*, 1998; LJUNG e GLAD, 1994; VAJDA *et al.*, 1989; LEE e HERGET, 1975, RAUE *et al.*, 2009), sem plena concordância com as demais.

2.2.2 O procedimento de identificabilidade de parâmetros

Em meados do século XX, a identificação de sistemas tratava-se de um tema corrente nos periódicos de engenharia, sobretudo nas áreas relacionadas à automação e ao controle de processos. Neste período, o avanço científico e tecnológico encontrava como barreira as limitadas ferramentas computacionais da época, seja para coleta e armazenamento de dados e/ou para realização de cálculos. Frente aos novos modelos, complexos e contendo um elevado número de parâmetros, o problema de estimação de parâmetros, frequentemente tornava-se intratável. Como alternativa, as observações experimentais disponíveis foram utilizadas para selecionar os parâmetros do modelo passíveis de identificação e realizar a estimação destes parâmetros. Foram elaborados procedimentos que, a partir de experimentos disponíveis, tornava possível identificar os parâmetros mais estimáveis do modelo. Ao procedimento de identificar parâmetros, deu-se o nome de identificabilidade de parâmetros.

Nas três últimas décadas do século XX, a chegada da tecnologia digital gradativamente revolucionou os sistemas de informação e automação permitindo, simultaneamente, a tomada, o armazenamento e o rápido acesso de várias propriedades dos processos, facilitando a operação em tempo real. Este período representa um importante marco na evolução dos processos modernos de engenharia, com a implementação de estratégias de controle avançado e de otimização, bem como a introdução de elementos computacionais importantes como os sensores virtuais (do

termo em inglês “*soft sensors*”), criando um novo conceito de operação de processos (FREITAS, 2009).

Apesar de todo o progresso científico e tecnológico vivenciado nos dias atuais, a literatura técnico-científica demonstra que a identificabilidade de parâmetros cada vez mais aumenta sua importância. Os modernos processos de engenharia são representados por modelos definidos por estruturas lógico-matemática de elevada complexidade contendo grande número de parâmetros, de modo que a estimação de parâmetros estatisticamente consistentes para tais modelos não é uma tarefa trivial, mesmo quando os experimentos são bem conduzidos e se dispõem de um grande número de pontos experimentais. Adicionalmente, com frequência, a disponibilidade de informação dos processos está limitada a uma estreita faixa experimental, dadas as exigências do controle de qualidade dos produtos, as restrições da segurança operacional e os padrões da legislação ambiental, que gradativamente tornam-se mais rigorosos.

Desde os estudos iniciais, a identificabilidade de parâmetros mantém seu caráter de seleção e de ordenamento dos parâmetros do modelo quanto ao potencial de estimação. Desta forma, baseado no modelo e na informação experimental disponível, este procedimento tem como objetivo determinar o potencial de estimação do parâmetro θ_i a partir da informação experimental disponível. Para tal, é necessário conhecer a influência que os parâmetros exercem sobre o comportamento do modelo, a qualidade da predição e quais variáveis são as mais adequadas para conduzir os experimentos.

Na Figura 2.6, é possível observar a importância da adequada descrição do sistema (o modelo) e da acurácia das medidas experimentais. Para o procedimento de identificabilidade utilizam-se as hipóteses do modelo perfeito e do experimento bem feito. De acordo com estas considerações, a incerteza experimental não deve apresentar qualquer tipo de tendência ou polarização, com o erro experimental flutuando em torno do valor zero, de modo que o valor médio esperado para o experimento é o próprio valor calculado com o modelo (SCHWAAB e PINTO, 2007). Desta forma, embora não seja considerada uma etapa, a caracterização da matriz de covariância experimental é fundamental para o sucesso do procedimento de identificabilidade de parâmetros.

A Figura 2.6 apresenta o esquema geral do procedimento de identificabilidade de parâmetros.

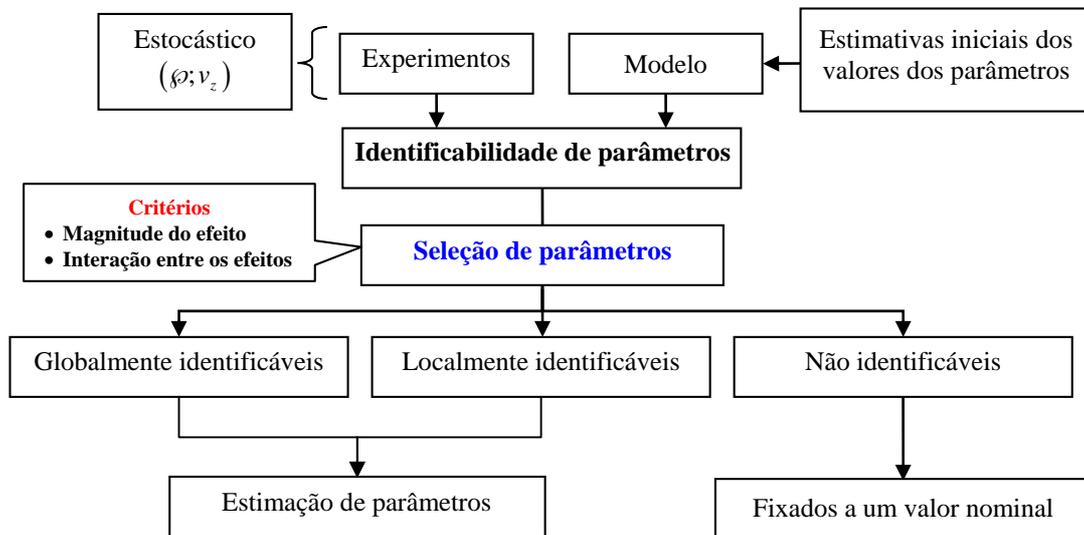


Figura 2.6 - Esquema geral do procedimento de identificabilidade de parâmetros.

De modo geral, os procedimentos de identificabilidade apresentam duas etapas básicas: a seleção e o ordenamento dos parâmetros, que naturalmente são realizadas de forma simultânea na maior parte dos algoritmos. Ambas as etapas requerem critérios que permitam julgar qualitativamente e/ou quantitativamente a importância dos parâmetros para o comportamento do modelo, visando à boa qualidade da predição. A definição destes critérios pode ser específica para cada metodologia desenvolvida, mas usualmente medidas de magnitude dos efeitos dos parâmetros sobre a predição são adotadas. Além da medida de magnitude, também podem ser adotadas medidas de interação dos efeitos dos parâmetros sobre a predição.

A etapa de seleção de parâmetros tem a finalidade de identificar quais os parâmetros do conjunto de parâmetros do modelo podem ser estimados a partir dos experimentos realizados. Nos procedimentos convencionais os parâmetros são avaliados um a um quanto a sua identificabilidade de forma que a ordem de seleção é uma informação facilmente armazenada. Os parâmetros selecionados como identificáveis devem ser estimados. Na abordagem convencional, a estimação dos parâmetros selecionados é uma etapa posterior ou exclusiva ao procedimento de identificabilidade.

Frente ao problema de parâmetros não identificáveis, uma solução alternativa deve ser buscada. Em estudos envolvendo modelos fenomenológicos a escolha convencional é manter o valor dos parâmetros não identificáveis como a estimativa

inicial (que pode ser obtido da literatura, da experiência do profissional ou arbitrado), realizando o procedimento de estimação dos demais parâmetros. Um exemplo típico para este caso são os modelos de polimerização (YAO et al., 2003; KOU et al., 2005a; KOU et al., 2005b; THOMPSON et al., 2008; TOJAL et al., 2013), geralmente formados por estruturas dinâmicas e não-lineares contendo um elevado número de parâmetros; nestes sistemas, a função objetivo pode apresentar múltiplos mínimos globais, como apresentado por SCHWAAB (2005). Já quando se trata de modelos empírico-fenomenológicos ou empíricos, a identificabilidade de parâmetros pode ser útil para simplificação e/ou redução destes modelos. Nesta classe, podem ser encontrados modelos para predição de propriedades termodinâmicas, inferidores de propriedade bioquímicas ou de composição de produtos, entre outros. Um bom exemplo são os modelos UNIFAC, que devido à grande correlação paramétrica, podem levar à estimação de parâmetros sem significado estatístico e/ou físico (COSTA, 2011).

Assim, a recente popularidade da identificabilidade se deve a duas vantagens associadas a estes procedimentos (KRAVARIS et al., 2012):

- (i). Interpretabilidade. A seleção de parâmetros não é meramente uma solução matemática para problemas de estimação mal condicionados. Trata-se também de uma ferramenta poderosa para análise do modelo, e.g., os resultados indicam quais parâmetros são importantes para o comportamento de um modelo bem como fornece informações a respeito da correlação dos efeitos que os parâmetros apresentam sobre a predição.
- (ii). Simplificação. A seleção de parâmetros reduz o número de variáveis de decisão no problema de otimização, visto que os parâmetros não selecionados são fixados na estimativa inicial. Esta característica é muito útil para a estimação de parâmetros em modelos complexos em que o esforço associado à otimização pode ser um problema; especialmente este é o caso de estratégias *on-line* de controle avançado (e.g., Controle Preditivo baseado no Modelo, do termo em inglês *Model Predictive Control* - MPC) e filtragem (e.g., Horizonte de Estimação Móvel, do termo em inglês *Moving Horizon Estimation* - MHE).

Essas características levaram a uma grande mudança no foco dos estudos de identificabilidade nas últimas décadas, antes fortemente voltada a parte estrutural do modelo – a identificabilidade estrutural, passou a se concentrar na informação

experimental – a identificabilidade prática. Por este motivo, a diferenciação entre a identificabilidade estrutural e a identificabilidade prática está mais presente na literatura clássica, em que os procedimentos são distintos em duas classes respectivas: (i) os métodos baseado em observabilidade e (ii) os métodos baseados em análise de sensibilidade. Uma vez que os métodos de observabilidade indiretamente fazem uso da análise de sensibilidade, esta divisão torna-se pouco distinguível na teoria e na prática.

A primeira classe de procedimentos, os métodos de observabilidade, representa os métodos clássicos de identificabilidade de parâmetros, muito populares no século passado. Estes métodos retomam a questão da identificabilidade como uma observação, alargando o espaço de estado para incluir parâmetros. A abrangência inicial destes métodos pode ser constatada pela grande variedade de metodologias disponíveis para verificação da identificabilidade estrutural, tais como: (i) expansão em séries de potências (POHJANPALO, 1978), (ii) abordagem Volterra e geração de séries de potência, (iii) abordagem da transformada de similaridade (VAJDA et al., 1989) e, (iv) métodos algébrico-diferenciais (LJUNG e GLAD, 1994). Infelizmente, tais métodos tornam-se inviáveis com o aumento do tamanho e da complexidade do modelo (MARGARIA et al., 2001; WHITE et al., 2001).

Assim, atualmente, os métodos de observabilidade são conhecidos pela dificuldade de aplicação prática, visto que necessitam resolver sistemas algébricos lineares de grandes dimensões. Além disso, estes métodos exigem condições de observabilidade e controlabilidade, que podem não ser satisfeitas para a maioria dos modelos (e.g., para os modelos biológicos) (AUDOLY *et al.*, 2009). Portanto, esta é uma ferramenta atualmente pouco empregada pelas engenharias química e ambiental, visto que estes métodos são usualmente desenvolvidos considerando modelos lineares nos parâmetros e/ou estacionários. Dentre os procedimentos de identificabilidade estrutural atuais, destaca-se o pacote computacional DAISY (do termo em inglês *Differential Algebra for Identification of SYstems*), desenvolvido por BELLU *et al.* (2007); o algoritmo base do pacote DAISY está implementado em linguagem Reduce, um software utilizado para computação algébrica geral (e.g., o Mathematica) que apresenta dificuldades de solução para problemas complexos (AUDOLY *et al.*, 2009).

Baseada na informação experimental, a segunda classe de métodos, os métodos de sensibilidade, consiste na verificação das relações funcionais entre os parâmetros e as variáveis de interesse através de simulação e de estimação de parâmetros do modelo. A abordagem baseada em análise de sensibilidade vem sendo explorada cada vez mais nos

últimos anos. Tal fato se deve a variedade de tipos destas análises e a facilidade de interpretação dos seus resultados, embora a maioria dos procedimentos esteja concentrada na matriz de sensibilidade local B . Naturalmente, o sucesso desta abordagem está baseado na adequabilidade da análise de sensibilidade ao problema, questão pouco abordada pelos usuários destas análises (SALTELLI et al., 2004).

Alternativa às metodologias tradicionais, baseadas em conceitos de observabilidade e controlabilidade de sistemas, a análise de sensibilidade requer importantes informações dos parâmetros (e.g., estimativas iniciais e grandezas estatísticas como a distribuições de probabilidades dos parâmetros, médias, variâncias) geralmente indisponíveis. A ausência de informações a respeito dos parâmetros pode levar a resultados enganosos, sendo comum o uso de informações reportadas na literatura para sistemas similares, baseada em opiniões de profissionais e/ou derivada de experimentos. De qualquer modo, estas informações não são conhecidas em absoluto sendo usualmente adotadas de forma arbitrária (KOU et al., 2005a; KOU et al., 2005b). Esta é uma questão que afeta fortemente os métodos de sensibilidade, especialmente aqueles que fazem uso de análises de sensibilidade baseadas em derivadas, também inclusa nos métodos de observabilidade.

Comparando ambas as abordagens de identificabilidade de parâmetros, análise de sensibilidade e métodos de observabilidade, a análise de sensibilidade está à frente pelo grande potencial de aplicação a modelos complexos, tendo como vantagem principal a versatilidade, devido à grande diversidade de métodos e as suas diferentes características.

2.3 ANÁLISE DE SENSIBILIDADE

Historicamente, a análise de sensibilidade está vinculada aos estudos de modelagem. Utilizadas para fins diversos, a análise de sensibilidade é considerada uma ferramenta essencial para avaliação da credibilidade de modelos.

Nas últimas décadas do século passado, a ciência vivenciou problemas de legitimidade e de relevância dos estudos de modelagem (SALTELLI *et al.*, 2008), sobretudo quanto à construção e ao uso dos modelos ambientais, que utilizados para fins de prognósticos a longo tempo, foram questionados por filósofos pós-modernos, intelectuais e cientistas de diversas tendências políticas. O despertar desta “era” parece ter sido provocado por RABITZ (1989). Com importantes considerações sobre a

generalidade dos estudos em escalas macroscópicas e microscópicas, este autor reforça a necessidade de uma abordagem mais sistemática para orientar os esforços teóricos e experimentais, chamando a atenção para a importância dos estudos de modelagem, dando ênfase à análise de sensibilidade. Desde então, houve uma explosão de conteúdos relacionados a análise de sensibilidade.

De modo geral, as análises de sensibilidade investigam a influência dos fatores, os elementos do modelo tais como parâmetros e variáveis independentes, sobre a resposta do sistema, as variáveis dependentes. Os fatores mais influentes são escolhas naturais para priorização, sendo excelentes elementos para a tomada de informações e a realização de interferências no sistema. Desta forma, analogamente, a compreensão de como o modelo se comporta em resposta às alterações nas entradas é de fundamental importância para sua utilização correta.

Admitindo o modelo perfeito, definem-se como fatores as variáveis independentes X e os parâmetros Θ , cuja medida de importância é usualmente fornecida pela análise de sensibilidade. Nesta contextualização, a análise de sensibilidade (AS) pode ser entendida como a ferramenta de estudo (determinação e quantificação) das fontes de incerteza da predição.

Devido à popularidade destas ferramentas que permeiam por diversas áreas do conhecimento técnico e científico (e.g., ciências sociais, econômicas e/ou naturais), não é surpreendente o número de significados encontrados para descrever o termo análise de sensibilidade. Entretanto, a definição que parece ser a mais empregada é dada por SALTELLI *et al.* (2004), uma vez que esta definição abrange a análise de sensibilidade em sua amplitude sem restrições a um ou outro contexto. De acordo com estes autores, a análise de sensibilidade apresenta o seguinte significado: "*Análise de sensibilidade é o estudo de como a incerteza presente na resposta do modelo (variáveis dependentes – Y) pode ser ocasionada por diferentes fontes de incerteza presentes nos fatores (variáveis independentes – X e parâmetros – Θ)*".

A Figura 2.7 apresenta a propagação da incerteza dos fatores para a incerteza da predição e suas metodologias de estudos: (i) a análise de incerteza - AI e (ii) a análise de sensibilidade - AS. A propagação de incerteza oriunda de diferentes fontes. Tanto a análise de incerteza quanto a análise de sensibilidade se ocupam do estudo de incerteza da predição, sendo que a primeira se ocupa em determinar esta incerteza enquanto a segunda se ocupa em definir as suas causas. Esta característica metodológica torna a AS

uma ferramenta muito atrativa aos estudos de modelagem, permitindo avaliar a credibilidade do modelo frente as informações disponíveis.

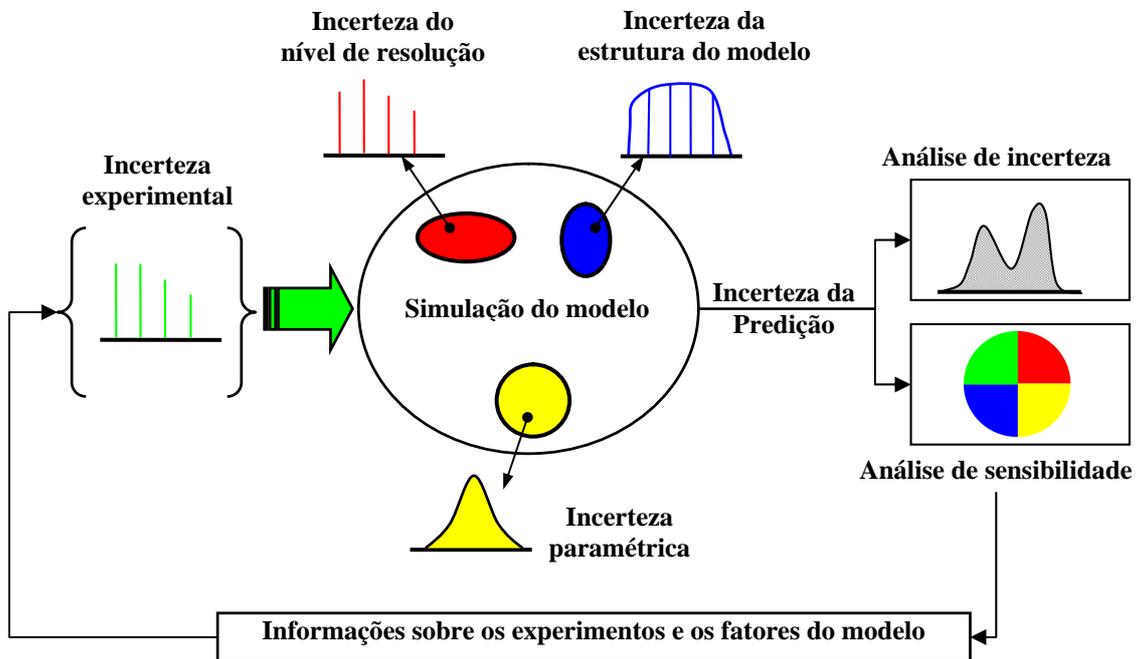


Figura 2.7 - Estudo da propagação de incerteza dos fatores para a incerteza da predição (Adaptado de SALTELLI *et al.*, 2004).

De amplo uso, as análises de sensibilidade encontram como principal utilidade a economia das etapas que envolvem o estudo da modelagem. São várias as aplicações desta ferramenta, dentre as quais se destacam:

- (i). Planejamento de experimentos. Definição de quais são os fatores mais importantes que requerem uma análise mais aprofundada.
- (ii). Construção de modelos. Determinação dos fatores que descrevem o sistema.
- (iii). Validação de modelos. Avaliação da robustez da predição do modelo e da consistência dos pressupostos em que estão baseados.
- (iv). Identificação de zonas críticas ou de outras regiões interessantes no espaço dos fatores de entrada. A identificação dos fatores que interagem e que podem, assim, gerar valores extremos.
- (v). Identificabilidade de parâmetros. Seleção prévia dos parâmetros que apresentam maior sensibilidade à predição de acordo com os experimentos disponíveis, para a estimação.

- (vi). Simplificação ou redução de modelos. Fatores que apresentam baixa ou nenhuma influência podem ser fixados a um valor nominal ou retirados da estrutura do modelo.

Dada a sua importância, a literatura técnico-científica reporta uma variedade de procedimentos empregados para realizar a AS, geralmente divididos em três classes: (i) a análise de fator por varredura (ASV), (ii) a análise de sensibilidade local (ASL) e (iii) a análise de sensibilidade global (ASG).

A primeira classe de métodos desempenha uma análise qualitativa, usualmente empregada em conjunto com uma das outras classes de AS ou como alternativa quando o tamanho e a complexidade do modelo tornam inviável a análise quantitativa. De modo geral, a ASV consiste em um procedimento de seleção de fatores em ordem decrescente de importância, sendo mais adequada quando o número de fatores investigados no modelo é pequeno comparado com o número total de fatores; segue-se a lei de Pareto, com alguns fatores de grande influência e a maioria não influentes (SALTELLI *et al.*, 2004). Embora sejam computacionalmente eficientes, os métodos de varredura tendem a abordar apenas um ponto específico ou região local no espaço de parâmetros (YU e HARRIS, 2009). Nesta classe de métodos destacam-se pela popularidade o planejamento fatorial fracionário e pela eficiência computacional o método de efeitos elementares (MORRIS, 1991).

É importante ressaltar que os métodos de varredura podem apresentar caráter local ou global. Assim, pode-se dizer que a classe de métodos ASV está dividida em 2 subclasses: (i) ASV local (ASVL) - varia o valor de um fator e fixa o valor dos demais e, (ii) ASV global (ASVG) - varia o valor de todos os fatores.

Para análises quantitativas, são utilizadas as classes ASL e ASG. Basicamente, a estratégia da ASL consiste em manter os valores dos parâmetros fixos variando as condições experimentais. Já a classe ASG, considera que os valores dos parâmetros e as condições experimentais variam simultaneamente. Dentre estas classes, a ASL parece ser a abordagem mais presente nos estudos de identificabilidade de parâmetros. Esta abordagem é baseada em derivadas dos fatores de entrada em relação à resposta do modelo (RABITZ, 1989); para estes métodos considera-se que apenas um ponto do espaço de parâmetros é explorado, sendo os fatores alterados um por vez. Todavia, há argumentos (SALTELLI *et al.*, 2004; SALTELLI *et al.*, 2008) sobre a possibilidade desta medida de sensibilidade ser válida somente para o ponto de base em que fora

computada e não oferecer uma ampla exploração do espaço dos fatores de entrada; sendo adequada em casos restritos. Embora este fato seja pouco importante para modelos lineares, em que a propriedade em um ponto de distância da linha de base pode ser calculada rapidamente por extrapolação linear por meio de derivadas de primeira ordem, para modelos não-lineares pode ter grande repercussão. Outro problema desta abordagem é que o cálculo de derivadas analíticas nem sempre ser possível, devido à complexidade da estrutura presente na maioria dos modelos das engenharias química e ambiental. Neste caso, as derivadas são computadas via procedimentos numéricos, sendo necessária a resolução do modelo repetidas vezes, o que implica em alto custo computacional (embora ainda inferior ao requerido pela ASG). Assim, pode-se dizer que uma limitação de abordagens locais baseadas em derivadas é que seu uso fica restrito aos casos em que a linearidade do modelo é conhecida. Exemplos de métodos presentes nesta classe são as análises baseadas em valor nominal ou diferencial tais como a matriz de sensibilidade local (SALTELLI *et al.*, 2004).

Em contrapartida, a abordagem da ASG, não apresenta muitas restrições às variadas aplicações, mas pode apresentar elevada exigência computacional. Nesta classe de métodos os fatores variam simultaneamente de acordo com suas escalas e sua função de distribuição de probabilidade; ou seja, quanto aos fatores, toda a gama de valores é considerada e todos eles são variados ao mesmo tempo. Desta forma, estes métodos apresentam uma maior chance de serem resistentes aos erros tipo II*, visto que a possibilidade de fatores importantes serem ignorados ou das combinações críticas entre estes fatores serem negligenciadas diminuem com o nível de exploração do espaço dos fatores (SALTELLI *et al.*, 2008). Esta classe engloba vários métodos de amostragem dentre os quais o robusto Monte Carlo (SOBOL, 1994) e os recentes métodos baseados em decomposição da variância.

A Figura 2.8 apresenta de forma ilustrativa a idéia básica das diferentes classes de análise de sensibilidade. A análise de sensibilidade local se caracteriza pela avaliação da importância dos fatores em valores fixados para os mesmos. Por outro lado, a análise de sensibilidade global avalia a ordem de importância dos fatores para diferentes valores dos fatores investigados; neste sentido, o resultado será obtido em termos de um valor médio da importância dos fatores, uma vez que eles são avaliados em faixas. O que diferencia a análise de sensibilidade de varredura das demais análises de sensibilidade é

* Na Estatística em teste de hipóteses, um erro de tipo II consiste em considerar como inválida uma hipótese que na verdade é válida. Erros tipo II também são conhecidos como Falso Negativo.

que esta análise se caracteriza pela avaliação das diferentes faixas experimentais. Esta análise pode ser local em relação aos parâmetros (valores de parâmetros fixados) ou global em relação aos parâmetros (valores de parâmetros variando). Embora estes termos sejam normalmente empregados na literatura, na prática não existe real diferença entre as análises de varredura e as demais análises de sensibilidade.

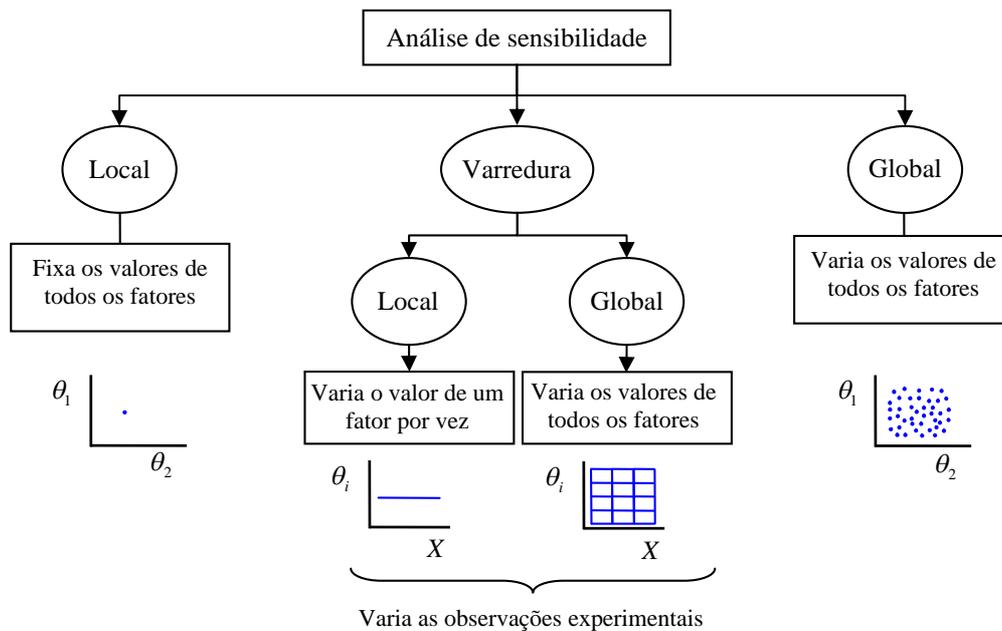


Figura 2.8 - Classes de análises de sensibilidade e suas metodologias.

A grande diversidade de métodos pode tornar difícil a escolha adequada da análise de sensibilidade ao problema. Embora a seleção da metodologia de análise de sensibilidade possa variar de problema para problema, algumas propriedades são desejáveis e recomendadas tais como (SALTELLI *et al.*, 2004):

- (i). A capacidade de lidar com a influência da dimensão e da forma. A influência dos fatores de entrada deve incorporar o efeito do intervalo de sua variação e a forma de sua função densidade de probabilidade (fdp).
- (ii). Inclusão de uma média multidimensional. O método deve ser capaz de avaliar o efeito de um fator enquanto todos os outros também são variados.
- (iii). Independência do modelo. Os efeitos individuais e de interação dos fatores devem ser caracterizados, independentemente das características do modelo.

- (iv). Tratamento de fatores agrupados como se fossem fatores únicos. Esta propriedade de síntese é essencial para a agilidade da interpretação dos resultados.

Para exemplificar a identificação destas propriedades em métodos de AS, uma comparação pode ser feita entre duas análises populares, que são: a medida de sensibilidade baseada em derivadas e a filtragem Monte Carlo (SOBOL, 1994). Dentre as quatro propriedades citadas acima, a sensibilidade baseada em derivadas possibilita apenas o agrupamento de fatores, sendo esta a única dentre as propriedades que a filtragem Monte Carlo não dispõe. Desta forma, pode-se dizer que estas análises apresentam características opostas, que refletem em resultados diferentes dependendo das características do modelo (e.g., a aditividade e a linearidade). Parece que estas propriedades são satisfeitas apenas por outra classe de métodos baseados em decomposição da variância (SALTELLI *et al.*, 2008).

Não obstante, a seleção de análise de sensibilidade adequada ao problema é um tema pouco explorado e bastante relegado (SALTELLI, 1999). A literatura técnico-científica reporta uma variedade de estudos em que diferentes testes de sensibilidade são aplicados ao mesmo problema. Esta prática produz uma variedade de resultados, frequentemente confusos e/ou não conclusivos (SALTELLI, 1999). De fato, este é tipicamente um erro tipo III[†]. John Tukey, brilhante estatístico do século 20, afirmava que “*uma resposta aproximada da questão certa é mais valiosa do que uma resposta certa de um problema aproximado*” (SALTELLI *et al.*, 2008). A grande dificuldade encontrada é que, em se tratando de erros tipo III, a análise de sensibilidade não pode ajudar (SALTELLI *et al.*, 2008).

Uma vez que o uso de diferentes AS para um mesmo problema leva a várias medidas para a importância dos fatores, surge naturalmente a seguinte questão: qual o método adotar? A resposta para esta pergunta é também a primeira e uma das mais decisivas etapas do estudo de sensibilidade de fatores e consiste em definir qual o objetivo da análise de sensibilidade. Devido à dificuldade em delimitar este objetivo, baseado nos estudos de análise de sensibilidade de SALTELLI *et al.* (2004), apresenta-se uma descrição geral das etapas a serem seguidas para utilização de análise de sensibilidade, independentemente do método utilizado:

[†] Erro tipo III corresponde aos erros de estruturação do problema, em que são buscadas respostas certas para problemas mal elaborados.

- (i). Estabelecer qual é o objetivo da análise e definir a forma de uma função que permita avaliar este objetivo.
- (ii). Selecionar quais os fatores a serem considerados e suas características (e.g., o valor do fator e/ou a distribuição de probabilidade).
- (iii). Selecionar o método de análise de sensibilidade.
- (iv). Aplicar a sequência de cálculos do método adotado.
- (v). Analisar os resultados da análise e avaliar o critério de parada, verificando a necessidade de repetir o procedimento.

É importante estar atento à etapa quatro que requer um conhecimento prévio do problema em questão e dos métodos de análise de sensibilidade. Nesta etapa é preciso estar ciente do tipo de problema que está sendo tratado e a solução que se deseja alcançar. É necessário definir o tipo de informação requerida para o tratamento do problema, qualitativa ou quantitativa, e a partir de então buscar métodos que estejam de acordo com o custo e o tempo computacional que podem ser despendidos. Além disto, a escolha da análise de sensibilidade depende também do número de fatores de entrada, das características do modelo (e.g., a linearidade e a monotonicidade) e da interação entre os fatores do modelo.

2.4 IDENTIFICABILIDADE DE PARÂMETROS VIA ANÁLISE DE SENSIBILIDADE

As dificuldades experimentais e a crescente complexidade dos processos modernos ocasionaram uma explosão dos procedimentos de identificabilidade baseados em análise de sensibilidade na virada do século passado. Atualmente, tais procedimentos são divididos em duas categorias (KRAVARIS *et al.*, 2012): (i) métodos baseados em otimização e (ii) métodos heurísticos.

Os métodos baseados em otimização utilizam critérios de planejamento de experimento ótimo para seleção de parâmetros (KRAVARIS *et al.*, 2012). A maioria dos critérios de planejamento visa reduzir a incerteza paramétrica (e.g., A-ótimo, C-ótimo, D-ótimo e E-ótimo), estando geralmente relacionados aos valores singulares da matriz de covariância da predição $V_{\hat{\theta}}$ e, em problemas de estimação mal condicionados, da matriz FIM , que carregam informação sobre a região de confiança dos parâmetros do modelo; logo, utilizam a matriz de sensibilidade local B para avaliar

quantitativamente a acurácia da estimação. Todavia, diversos trabalhos apresentam formas simplificadas de cálculo da matriz FIM , principalmente devido ao desconhecimento de termos importantes, como a variância experimental V_Y e as dificuldades associadas ao cálculo e a inversão da própria FIM . Devido ao esforço experimental requerido para caracterizar adequadamente a incerteza experimental, a simplificação mais adotada é considerar que o valor desta incerteza é constante e os experimentos não são correlacionados, o que pode levar a graves distorções do planejamento (KIEFER, 1959; KIEFER, 1975). A Figura 2.9 apresenta a relação entre a forma da região de confiança de um modelo linear, a hiper-elipse, e os valores singulares da matriz FIM .

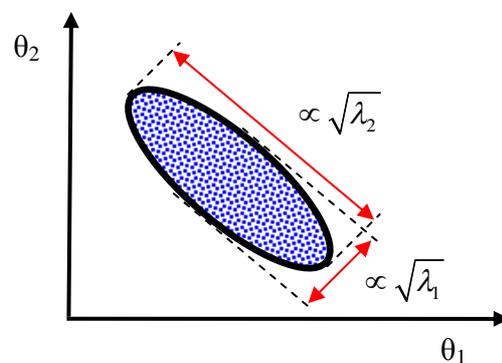


Figura 2.9 - Relação entre os valores característicos da matriz FIM e a hiper-elipse (ALBERTON, 2010; SCHWAAB, 2007).

Outros critérios (e.g., G-ótimo e I-ótimo) visam reduzir a incerteza da predição, o que conseqüentemente podem levar a parâmetros bem estimados, conforme demonstrado a partir do Teorema da Equivalência Geral (ATKINSON et al., 2007; KIEFER e WOLFOWITZ, 1960).

Chamados alfabeticamente, os principais critérios de planejamento ótimo para redução da incerteza paramétrica reportados pela literatura são apresentados pela Tabela 2.1. A literatura reporta uma ampla discussão a respeito do uso e adequabilidade destes critérios ao problema de estimação de parâmetros. No problema de identificabilidade de parâmetros a escolha comum vem sendo o uso do critério D-ótimo (critério do volume) associado ao critério E-ótimo (critério da forma) (WEIJERS e VANROLLEGHEM, 1997; BRUN et al., 2001; MACHADO et al., 2009).

O critério D-ótimo é um dos mais empregados na literatura de planejamento de experimentos (ALBERTON, 2010). Uma vantagem associada a este critério é a

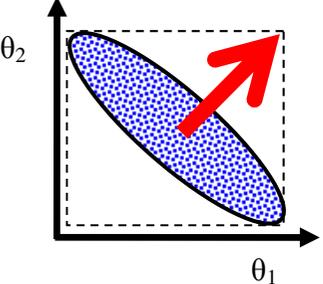
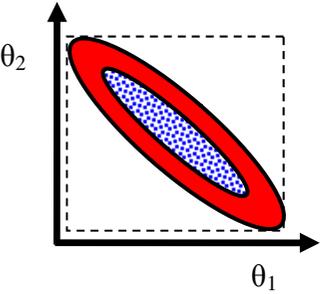
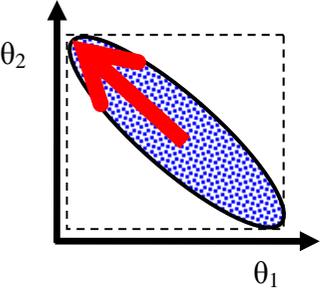
invariância com respeito a algumas transformações aplicadas nos parâmetros, como o re-escalonamento (PINTO et al., 1990). Além disso, este critério tende a dar grande importância para parâmetros que apresentam maior sensibilidade sobre a predição do modelo (SCHWAAB, 2007; ALBERTON, 2010), empregando a informação experimental para a seleção dos parâmetros estatisticamente mais significativos. Todavia, o uso do critério de planejamento D-ótimo, que consiste na minimização do volume da região de confiança, pode levar a parâmetros altamente correlacionados. Neste sentido aplica-se o critério E-ótimo (critério da forma) que objetiva minimizar a correlação entre os parâmetros. A principal desvantagem do critério E é concentrar experimentos na redução da incerteza de um único parâmetro, cujo erro pode ser difícil de ser minimizado (PINTO et al., 1990).

Empregando qualquer dentre os critérios de planejamento ótimo, a literatura aborda o problema de otimização para seleção de parâmetros por meio da formulação (KRAVARIS et al., 2012):

$$\begin{aligned}
 \Omega^* &= \arg \max_{\Omega} \phi(B_S^T B_S) \\
 \text{(E 2.11)} \quad &\text{sujeito a: } B_S = B I_S \\
 &B = [b_{\theta_1}, b_{\theta_2}, \dots, b_{\theta_{NP}}] \\
 &I_S = [l_{i_1}, l_{i_2}, \dots, l_{i_{NPS}}], \text{ com } i_j \text{ que } \Omega_{i_j} = 1
 \end{aligned}$$

a matriz de sensibilidade B contém os vetores de sensibilidade de todos os NP parâmetros do modelo, $B = [b_{\theta_1}, b_{\theta_2}, \dots, b_{\theta_{NP}}]$ e a matriz identidade relacionada aos parâmetros selecionados $I_S = [l_{i_1}, l_{i_2}, \dots, l_{i_{NPS}}]$, com i_j que $\Omega_{i_j} = 1$, sendo que l_i corresponde a i -ésima coluna da matriz identidade.

Tabela 2.1 – Principais critérios de planejamento de experimentos: (i) A-ótimo, (ii) D-ótimo e, (iii) E-ótimo.

CONCEITO	DEFINIÇÃO	INTERPRETAÇÃO GEOMÉTRICA
A-ótimo Critério do traço	Maximização da soma da diagonal principal da matriz FIM , o que representa reduzir a diagonal do hiper-cubo que circunscribe a hiper-elipse.	$\phi_A(B^T B) = -\text{tr}(B^T B) = \sum_{i=1}^{NP} \lambda_i$ 
D-ótimo Critério do volume	Maximização do determinante da matriz FIM , o que representa reduzir o volume da hiper-elipse.	$\phi_D(B^T B) = -\det(B^T B) = \prod_{i=1}^{NP} \lambda_i$ 
E-ótimo Critério da forma	Maximização do maior valor característico de FIM , o que representa reduzir o maior eixo da hiper-elipse.	$\phi_E(B^T B) = \lambda_{\max}(B^T B) = \max\{\lambda_i\}$ 

A identificabilidade de parâmetros via otimização leva a um problema de programação não linear inteira mista, em que um vetor de decisão binário Ω , $\Omega_i \in \{0,1\}$, é utilizado para denotar quais são os parâmetros selecionados; sendo que $\Omega_i = 1$ indica a seleção do parâmetro θ_i . O número de parâmetros selecionados NPS é igual ao posto numérico da matriz de sensibilidade dos parâmetros selecionados B_S . Deste modo, a soma dos valores contidos no vetor Ω_i deve ser igual ao número de parâmetros selecionados, como:

$$(E 2.12) \quad \sum_{i=1}^{NP} \Omega_i = NPS.$$

O cálculo da matriz de parâmetros selecionados B_S dos métodos baseados em otimização é ilustrado pela Figura 2.10.

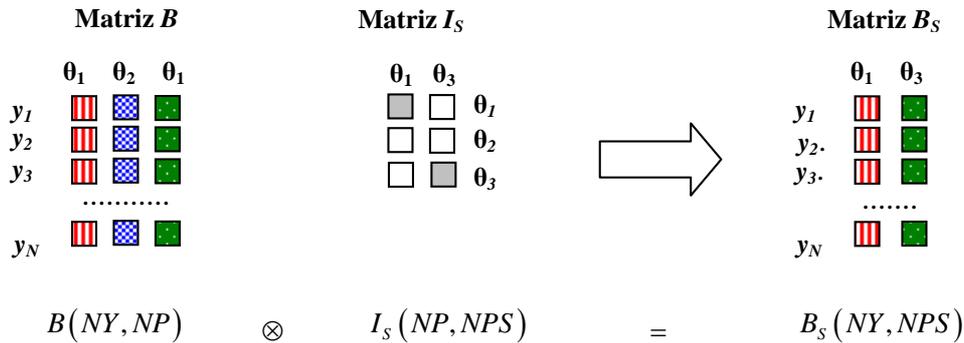


Figura 2.10 - Definição da matriz de parâmetros selecionados via métodos baseados em otimização.

Por se tratar de um problema combinatorial, a resolução dos métodos baseados em otimização não é uma tarefa trivial, em especial, devido ao número total de combinações para selecionar NPS parâmetros de um conjunto de NP parâmetros que é $NP! / (NPS!(NP - NPS)!)$ (KRAVARIS *et al.*, 2012). A solução ótima pode ser difícil ou até mesmo impossível de se obter, se um elevado número de parâmetros precisa ser selecionado a partir de um grande conjunto de parâmetros possíveis. Deste modo, formulações alternativas do problema de programação não linear inteira mista são utilizadas para determinar uma solução sub-ótima buscando menor esforço

computacional tais como (KRAVARIS *et al.*, 2012): (i) técnicas que incluem o procedimento de seleção sequencial (e.g., seleção de parâmetros um a um (BLANCHET *et al.*, 2008)), (ii) busca estocástica (e.g., algoritmo genético como ferramenta de seleção de subconjunto de parâmetros viáveis para a estimação (CHU e HAHN, 2007)) e, (iii) abordagens heurísticas de redução (e.g., *cluster* para agrupamentos de parâmetros de acordo com a faixa de valor do efeito do parâmetro sobre a predição (CHU e HAHN, 2009)).

Mais populares, os métodos heurísticos tendem a ser mais simples que os métodos baseados em otimização e são baseados diretamente no efeito que as variações dos parâmetros apresentam sobre a predição do modelo (KRAVARIS *et al.*, 2012). Muito provavelmente por este motivo, estes métodos representam a categoria de métodos em maior desenvolvimento e aplicação atualmente. KRAVARIS *et al.* (2012) apresentam dois casos para ilustrar a ideia básica dos métodos heurísticos:

- (i). Caso 1 – Parâmetros pouco significativos. O parâmetro apresenta apenas um efeito marginal sobre a predição do modelo, o que torna a estimação deste parâmetro uma tarefa difícil, visto que mesmo uma baixa incerteza experimental afetará significativamente o valor estimado do parâmetro.
- (ii). Caso 2 – Parâmetros correlacionados. O efeito que a variação de um parâmetro apresenta sobre a predição está correlacionado com os efeitos que os demais parâmetros apresentam, de modo que não é possível determinar unicamente os valores destes parâmetros sem que o procedimento de estimação se torne mal condicionado.

Em cada um destes casos, que frequentemente ocorrem simultaneamente em um problema de estimação, o procedimento de identificabilidade permite tratar o problema de estimação ignorando-se os parâmetros que não resultam em único efeito sobre a predição, o que é requerido para um bom ajuste do modelo. Assim, desde que a estimação de todos os parâmetros não é possível, a estimação de um subconjunto de parâmetros pode regularizar o problema de estimação mal condicionado e simplificar o problema de otimização associado (KRAVARIS *et al.*, 2012).

Nos métodos heurísticos, critérios de identificabilidade baseados em análise de sensibilidade são formulados para avaliar os parâmetros. Em tais critérios, a escolha mais comum é o uso análise de sensibilidade local, devido à facilidade de obtenção da matriz B numericamente. Consequentemente, a maioria destes procedimentos também

faz o uso da matriz *FIM* para avaliar o condicionamento do procedimento sendo usual adotar como critério de parada nestes procedimentos o ponto em que esta matriz torna-se singular. Apesar das muitas propriedades interessantes, o uso de análise de sensibilidade global é uma prática raramente empregada nestes procedimentos. Acredita-se que a justificativa maior seja o esforço computacional requerido pelas análises de sensibilidade global, que é muito superior quando comparado ao esforço para obtenção da matriz *B*.

É importante observar a ausência de trabalhos na literatura técnico-científica empenhados em comparar qualitativa e quantitativamente o desempenho de diferentes métodos de identificabilidade de parâmetros. Adicionalmente, verifica-se que a maioria dos trabalhos durante a etapa de validação destes procedimentos utiliza apenas um único sistema, por vezes com características bastante específicas, não fornecendo informações sobre a abrangência de aplicação destes métodos.

A seguir são descritas algumas metodologias de identificabilidade de parâmetros que utilizam diferentes abordagens de análise de sensibilidade: (i) local e (ii) global. Claramente, o sucesso destas metodologias requer, além de conhecimento sobre as mesmas, o entendimento da natureza do sistema, de forma que a ferramenta esteja adequada ao problema.

2.4.1 Identificabilidade de parâmetros via análise de sensibilidade local

Atualmente, a maioria dos procedimentos de identificabilidade de parâmetros utiliza a análise de sensibilidade local, sejam os métodos baseados em heurísticas ou os métodos baseados em otimização. Análise de sensibilidade local tem demonstrado ser uma ferramenta valiosa para identificar parâmetros mais significativos para a predição e pouco correlacionados (SECCHI et al., 2006). O uso desta ferramenta para identificabilidade levou a dois critérios de seleção de parâmetros (KRAVARIS et al., 2012):

- (i). Magnitude do efeito. A norma do vetor de sensibilidade do parâmetro selecionado deve ter um valor significativo quando comparada às normas dos vetores de sensibilidade dos demais parâmetros.
- (ii). Efeitos correlacionados. Os vetores de sensibilidade dos parâmetros selecionados não podem ser fortemente correlacionados.

Alguns métodos heurísticos empregam ambos os critérios de seleção, que são satisfeitos apenas para os parâmetros de elevada magnitude e baixa correlação. Para o uso de ambos os critérios, alguns métodos utilizam uma abordagem *ad hoc* para balancear.

A Tabela 2.2 apresenta alguns procedimentos de identificabilidade de parâmetros que empregam a análise de sensibilidade local como heurística de seleção (KRAVARIS et al., 2012). De acordo com esta tabela verifica-se que, nesta última década, vários procedimentos foram baseados na análise de sensibilidade local *B*. Muito populares, estas propostas de identificabilidade que utilizam a matriz de sensibilidade local quase sempre fazem uso da matriz de Informação de Fisher (*FIM*); assim como os métodos baseado em otimização. O número condicional (razão entre o máximo e o mínimo valor característico) da matriz *FIM* indica o condicionamento do problema de estimação de parâmetros, sendo a singularidade desta matriz usualmente adotada como critério de parada da seleção. Vale ressaltar que estes procedimentos vem sendo aplicados a modelos não lineares contendo elevado número de parâmetros com relativo sucesso.

Tabela 2.2 – Alguns procedimentos de identificabilidade de parâmetros que empregam a análise de sensibilidade local como heurística de seleção (baseado em KRAVARIS et al., 2012).

Metodologia	Referências
Ortogonalização	YAO et al. (2003)
	LUND e FOSS (2008)
Índice de colinearidade	BRUN et al. (2001)
Matriz de ganhos relativos	SANDINK et al. (2001)
	BOTELHO (2013)
Valores singulares de Hankel	SUN e HAHN (2006)
PCA e Ângulo de independência linear	LI et al. (2004)
	SECCHI et al. (2006)
D-ótimo	BRUN et al. (2002)
	CHU e HAHN (2007)
D-ótimo combinado ao E-ótimo modificado	WEIJERS e VANROLLEGHEM (1997)
	MACHADO et al. (2009)

As metodologias de OMLIN *et al.* (2001) e BRUN *et al.* (2001) aplicam como critério de seleção, a importância relativa dos parâmetros obtida por meio da matriz *B* e

o índice de colinearidade que trata da correlação paramétrica obtida por meio da matriz *FIM*. Estas metodologias mostram grande similaridade com a proposta de identificabilidade de WEIJERS e VAROLLEGHEM (1997), cujo critério está relacionado a forma da região de confiança dos parâmetros. Na metodologia proposta por LI *et al.* (2004) a importância dos parâmetros é obtida por meio da decomposição em valores singulares (SVD) da matriz *FIM*, sendo a correlação avaliada a partir do ângulo de independência linear entre os vetores de sensibilidade dos parâmetros projetados sobre a matriz de covariância da predição considerando apenas os parâmetros selecionados. SECCHI *et al.* (2006) apresentaram extensões deste método, com destaque para a inserção da etapa de estimação simultânea a seleção, sendo que LUZIVETTO *et al.* (2006) adicionaram uma estratégia para redução de modelos com base nos parâmetros considerados não identificáveis. Abordagens baseadas na Matriz de Ganhos Relativos também são proposta na literatura, por SANDINK *et al.* (2001) e BOTELHO (2012), empregando diretamente a matriz *FIM*, sendo mais efetiva em identificar a correlação paramétrica do que a influência do parâmetro sobre a predição. CHU e HAHN (2007) propuseram uma metodologia baseada em técnicas de ortogonalização e critério de planejamento, que busca determinar quais parâmetros não podem ser estimados conjuntamente, obtendo o grau da possível variação dos parâmetros não selecionados em relação as estimativas iniciais. LUND e FOSS (2008) propuseram uma metodologia que também é baseada em técnicas de ortogonalização e o algoritmo genético, com resultados muito similares ao procedimento de YAO *et al.* (2003), mas com elevado custo computacional. YAO *et al.* (2003), cuja proposta foi diretamente baseada na análise de sensibilidade local, avalia os parâmetros via ortogonalização de Gram-Schmidt, apresentando resultados de fácil análise e interpretação com baixo custo computacional quando comparado às demais metodologias. Estas características fizeram o procedimento de YAO *et al.* (2003) bastante popular, mas originalmente tal metodologia adotava limites de cortes pré-estabelecidos para o número de parâmetros, o que foi resolvido com o critério de parada estabelecido por THOMPSON *et al.* (2009). Um método que determina quantos parâmetros devem ser estimados é o apresentado em CHU *et al.* (2011) e demonstra uma predição mais acurada que os resultados obtidos por meio dos métodos de ortogonalização (e.g., YAO *et al.*, 2003; LUND e FOSS, 2008). VELEZ-REYES e VERGHESE (1995) e HISKENS (2001) também apresentaram procedimentos de identificabilidade baseados na análise de sensibilidade local e ortogonalização, fazendo

o uso das transformadas Householder. Ambos os métodos de ortogonalização, Gram-Schmidt e transformada Householder produzem resultados idênticos, mas a transformada Householder é mais estável numericamente (GOLUB e van LOAN, 1996) e o Gram-Schmidt é mais fácil de interpretar (KRAVARIS et al., 2012).

Dentre as metodologias baseadas em análise de sensibilidade local, a abordagem de YAO et al. (2003) destaca-se pela sua simplicidade de interpretação dos resultados e facilidade de implementação. A abordagem de SECCHI et al. (2006) também se destaca por apresentar a etapa de estimação simultaneamente à seleção dos parâmetros. Estas metodologias são descritas a seguir.

2.4.1.1 YAO et al. (2003)

Proposto originalmente por YAO et al. (2003), este procedimento consiste em uma análise de estimabilidade dos parâmetros baseada na matriz de sensibilidade dos parâmetros em relação à predição do modelo normalizada $\left(\left(\partial y_i / \partial \theta_j \right) \cdot \left(\theta_j / y_i \right) \right)$ por meio do método de ortogonalização de Gram-Schmidt. Devido a sua simplicidade e a fácil interpretação dos seus resultados (KRAVARIS et al., 2012), este é um dos procedimentos mais populares (KOU et al., 2005a; KOU et al., 2005b; WU et al., 2007; THOMPSON et al., 2008; WU et al., 2011; McLEAN et al., 2012). No procedimento original a seleção é interrompida com base em um valor de corte específico definido a priori que delimita o tamanho do conjunto de parâmetro selecionado definido a priori. Infelizmente, pode ser muito difícil escolher um valor de corte adequado, sem que o número de parâmetro selecionados se torne arbitrário. Para contornar esta dificuldade, THOMPSON et al. (2009) adotaram como critério de parada o ponto em que a matriz *FIM* torna-se singular.

A Figura 2.11 ilustra o procedimento de identificabilidade de YAO et al. (2003) para 3 parâmetros, com o critério de parada modificado por THOMPSON et al. (2009).

O algoritmo da estratégia de identificabilidade de YAO et al. (2003) com o critério de parada modificado por THOMPSON et al. (2009) é dado pelas etapas a seguir:

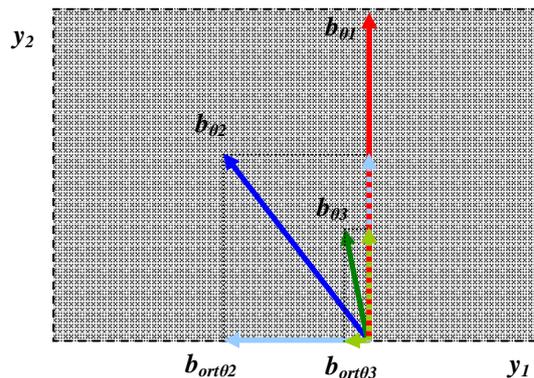
- (i). Calcular a magnitude (a soma dos quadrados dos elementos) de cada coluna b da matriz de sensibilidade local normalizada $B: \|b_0\|$.
- (ii). Selecionar o parâmetro cuja coluna da matriz B apresenta maior magnitude como o primeiro parâmetro identificável.

- (iii). Montar a matriz de sensibilidade local para os parâmetros selecionados B_{Sel} . Quando o primeiro parâmetro é selecionado, $k = 1$, a matriz contém apenas uma coluna. Quando parâmetros subsequentes forem selecionados, a matriz B_{Sel} contém k colunas.
- (iv). Calcular B_{Orth} , a correspondente matriz ortogonal da completa matriz de sensibilidade local B : $B_{Orth} = B_{Sel} (B_{Sel}^T B_{Sel})^{-1} B_{Sel}^T B$, em que $B_{Sel}^T B_{Sel}$ corresponde à matriz de Informação de Fisher FIM dos parâmetros selecionados.
- (v). Calcular a matriz residual R : $R = B - B_{Orth}$.
- (vi). Calcular a magnitude (a soma dos quadrados dos valores dos elementos) de cada coluna da matriz R : $\|r\|$. A coluna com a maior magnitude corresponde ao próximo parâmetro identificável.
- (vii). Selecionar, na matriz B , a correspondente coluna obtida na etapa (vi) e adicioná-la à matriz B_{Sel} .
- (viii). Avançar o contador da iteração em um e repetir as etapas (iii) a (vii) até que todos os parâmetros sejam selecionados ou problemas de singularidade sejam encontrados no procedimento de inversão da matriz FIM .

O primeiro parâmetro é selecionado de acordo com a magnitude de seu efeito. O parâmetro de maior magnitude corresponde ao parâmetro cujo vetor de sensibilidade apresenta maior comprimento, $\|b_{\theta_i}\|$, neste caso o parâmetro 1 ($b_{\theta_1} > b_{\theta_2} > b_{\theta_3}$).

Matriz B

θ_1	θ_2	θ_3	
			y_1
			y_2
↓	↓	↓	
b_{θ_1}	b_{θ_2}	b_{θ_3}	



Para que a estimação seja viável, o próximo parâmetro selecionado deve ser o parâmetro que apresenta a menor correlação com o parâmetro 1. Para tal, faz-se a projeção de b_{θ_2} e b_{θ_3} sobre o b_{θ_1} , tendo como resultado os vetores ortogonais b_{orth02} e b_{orth03} . O parâmetro que apresenta menor correlação é o parâmetro cujo vetor de sensibilidade ortogonal ao parâmetro selecionado apresenta maior comprimento, $\|b_{orth\theta_i}\|$, neste caso o parâmetro 2 ($b_{orth02} > b_{orth03}$).

Uma vez que o b_{θ_3} , está fortemente correlacionado com b_{θ_1} , o parâmetro 3 não pode ser selecionado sem que a matriz FIM se torne singular. Então, o procedimento é interrompido, sendo identificáveis neste caso apenas os parâmetros 1 e 2.

Figura 2.11 - Procedimento de YAO et al. (2003) ilustrado para 3 parâmetros.

2.4.1.2 SECCHI et al. (2006) – Seleção e estimação de parâmetros (SELEST)

Baseado originalmente no método de LI *et al.* (2004), SECCHI *et al.* (2006) propuseram um algoritmo para seleção e estimação de parâmetros baseado na matriz de sensibilidade dos parâmetros em relação à predição do modelo, denominado SELEST (do termo em inglês *Selection and Estimation*).

Basicamente, este método consiste em definir o conjunto de parâmetros identificáveis, ordenando-os de acordo com sua importância para a predição e o seu grau de independência linear. A influência de cada parâmetro sobre a predição é quantificada pela aplicação da técnica de análise dos componentes principais (do termo em inglês *Principal Component Analysis - PCA*) à matriz de sensibilidade local (B). A dependência linear dos parâmetros é determinada pela métrica de independência linear (LI *et al.*, 2004) que relaciona o vetor de sensibilidade de um parâmetro candidato à seleção com os vetores de sensibilidade dos parâmetros já selecionados para a estimação. Estas informações, computadas simultaneamente, são utilizadas para geração do índice de identificabilidade de parâmetros. De acordo com esta metodologia, os parâmetros são selecionados e ordenados decrescentemente de acordo com os maiores valores obtidos para este índice, sendo os primeiros parâmetros desta ordem considerados identificáveis e aptos à re-estimação. Embora a metodologia de identificabilidade de parâmetros proposta por LI *et al.* (2004) seja embasada em grandezas estatísticas bastante conhecidas e utilizadas para avaliação da qualidade de parâmetros, esta estratégia não fornece um critério para determinar o número ótimo de parâmetros a serem selecionados para a realização do procedimento de estimação. Tal fato pode dificultar e até inviabilizar o procedimento de estimação de parâmetros. Para contornar esta dificuldade, SECCHI *et al.* (2006) introduziram dois critérios de seleção e ordenamento dos parâmetros: o índice de degradação da predição e o índice de degradação da correlação. Desta forma, o algoritmo é interrompido quando não há melhoria na predição do modelo ou quando a matriz de correlação dos parâmetros torna-se mal condicionada. Outra proposta apresentada por SECCHI *et al.* (2006) foi à extensão do método para aplicações a modelos dinâmicos e a sistemas com múltiplos pontos de operação. Portanto, a estratégia desempenhada pelo SELEST permite estimar parâmetros de modelos complexos mesmo quando dados experimentais são escassos, se uma boa estimativa inicial para os valores dos parâmetros é fornecida.

A Figura 2.12 ilustra o algoritmo da estratégia de identificabilidade de parâmetros de SECCHI *et al.* (2006).

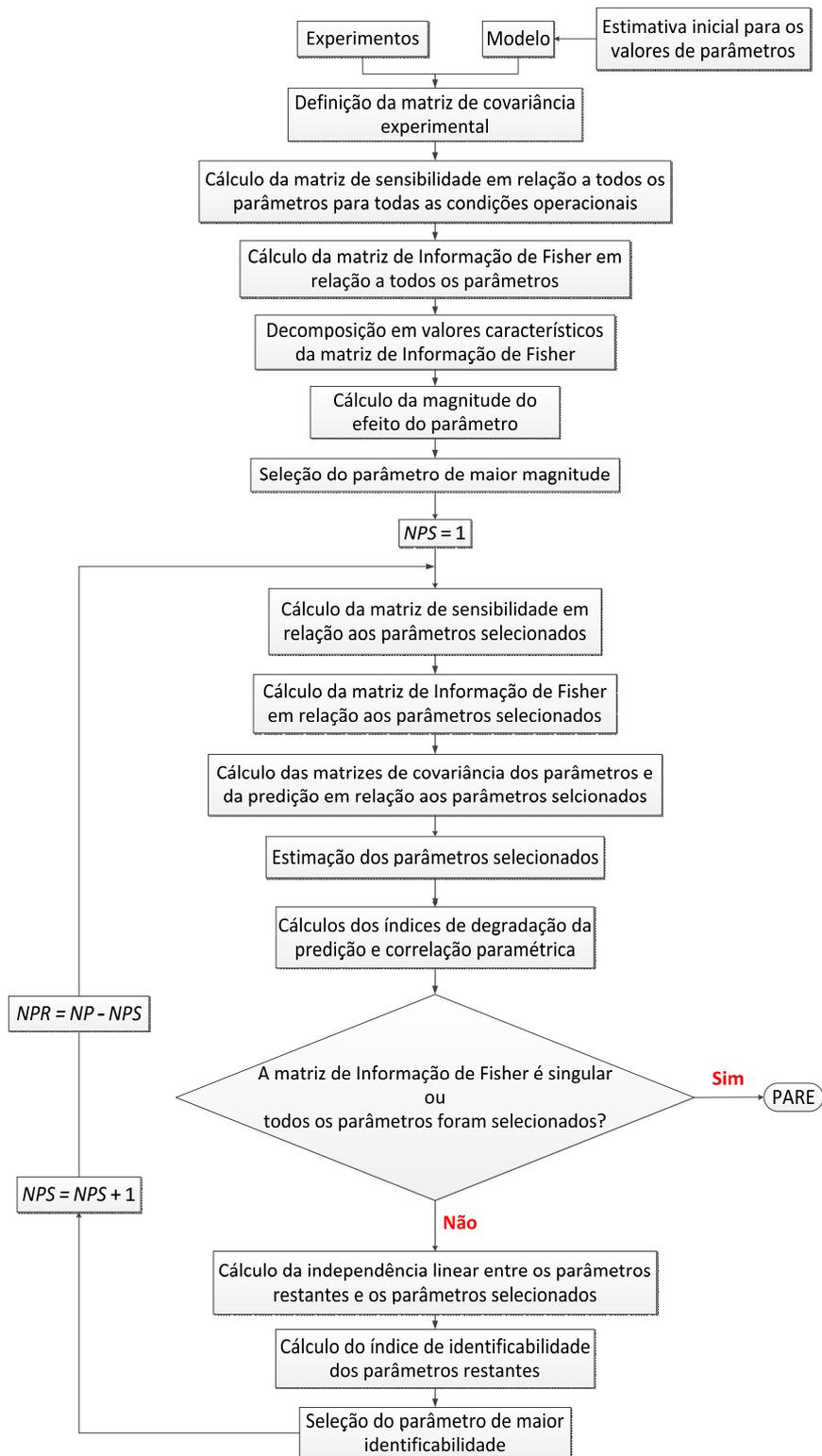


Figura 2.12 – Algoritmo de identificabilidade de parâmetros proposto por SECCHI *et al.* (2006) – SELEST (do termo seleção e estimação); NP , NPS e NPR , representam respectivamente o número total de parâmetros do modelo, o número de parâmetros selecionados e o número de parâmetros restantes.

2.4.2 Identificabilidade de parâmetros via análise de sensibilidade global

Atualmente, a classe de análise de sensibilidade global tem recebido maior atenção pela literatura (ESTRADA *et al.*, 2009; Di MAGGIO *et al.*, 2010; CHU *et al.*, 2011). Tal fato se deve ao amplo panorama de aplicação destas ferramentas (SALTELLI *et al.*, 2008). No contexto de análise de sensibilidade, as três últimas décadas do século passado foram marcadas pelo desenvolvimento dos métodos de análise de sensibilidade global baseados em métodos de decomposição da variância. Estes métodos verificam a distribuição das saídas e as suas mudanças, em média, quando um fator é fixado.

Os métodos de análise de sensibilidade global se preocupam com uma questão importante para a análise de sensibilidade: a quantificação da incerteza da predição. O cálculo da variância é usualmente recomendado como um resumo da medida desta incerteza. Desta forma, estes métodos apresentam como grande vantagem a possibilidade da variância ser decomposta em componentes do sistema, permitindo identificar os fatores que mais contribuem para a análise.

Nesta classe de métodos de análise de sensibilidade pode-se citar o pioneiro trabalho de CUKIER *et al.* (1973), com desenvolvimento do método FAST (do termo em inglês “*Fourier Amplitude Sensitivity Test*”), o trabalho de IMAN e HORA (1990), com importantes contribuições para análise da incerteza da predição, e os índices de sensibilidade global desenvolvidos pelo matemático russo SOBOL (1991).

Inspirado no trabalho de CUKIER *et al.* (1973), SOBOL (1990) apresentou uma metodologia baseada no método Monte Carlo, capaz de computar medidas de sensibilidade para grupos arbitrários de fatores. Esta metodologia, mais tarde estendida por SALTELLI (2002), é atualmente o procedimento que apresenta melhor eficiência computacional para computar índices de sensibilidade baseados puramente em avaliações do modelo. Considerando a eficiência computacional do método apresentado por SALTELLI (2002), os índices de sensibilidade global foram empregados como estratégia de identificabilidade de parâmetros. A seguir descreve-se a metodologia empregada por SALTELLI (2002) para a geração de índices de sensibilidade global.

2.4.2.1 SALTELLI (2002) – Índices de sensibilidade global

A autoria deste método é atribuída a SALTELLI (2002) e representa uma extensão das abordagens desenvolvidas por SOBOL (1990) e HOMMA e SALTELLI (1996). Do ponto de vista de identificabilidade de parâmetros, esta metodologia

apresenta características muito interessantes, tais como a independência do modelo, a captura da influência da faixa de variação dos fatores e a avaliação dos efeitos e das interações entre os fatores (SALTELLI et al., 2008).

Nesta análise, a medida de sensibilidade é dada por meio dos chamados “índices de sensibilidade” (IS), cujo cálculo consiste em um procedimento numérico baseado em técnicas de amostragem Monte Carlo. O IS total (ST_i) reflete a influência do fator (neste caso, o parâmetro θ_i) na predição Y , enquanto o IS de primeira ordem (S_i) indica o quanto, em média, se pode reduzir a variância da predição se tal fator (θ_i) for fixado; logo, S_i é a medida do efeito principal.

$$(E 2.13) \quad S_i = \frac{V[E(Y/\theta_i)]}{V(Y)},$$

$$(E 2.14) \quad ST_i = 1 - \frac{V[E(Y/\theta_{\sim i})]}{V(Y)},$$

em que $V[E(Y/\theta_i)]$ é a variância condicional de Y para um dado θ_i , $V[E(Y/\theta_{\sim i})]$ é a variância condicional de Y exceto um dado θ_i e $V(Y)$ representa a variância incondicional (total) de Y .

Os efeitos de interação entre os parâmetros, que refletem o grau de correlação do fator, são fornecidos pelo IS de interação (SI_i):

$$(E 2.15) \quad SI_i = ST_i - S_i.$$

Considerando que o índice ST_i é a medida de efeito total (a influência sobre a predição do modelo e a interação com os demais parâmetros) do parâmetro θ_i enquanto o índice S_i é a medida do efeito isolado do parâmetro θ_i ; por definição, quando o parâmetro θ_i está correlacionado com outro parâmetro do modelo, ST_i deve ser maior do que S_i , caso contrário ST_i deve ser igual a S_i . Desta forma, a diferença $ST_i - S_i$ pode ser utilizada como medida do quanto o parâmetro θ_i está correlacionado com os demais parâmetros do modelo. Caso o ST_i seja igual a 0, o parâmetro θ_i é dito não

significativo para a predição e pode ser fixado em qualquer ponto de sua distribuição sem afetar a variância da predição.

Vale ressaltar que a soma de todos os valores de ST_i é sempre maior ou igual a 1. No caso de modelos aditivos, em que não há interações entre os fatores dos modelos, esta soma equivale a 1 e a soma de todos os S_i também equivale a 1. Deste modo, a diferença $1 - \sum_i S_i$ também pode ser utilizada como um indicador de presença de interações entre os fatores do modelo.

De acordo com SALTELLI *et al.* (2008), estas estimativas $\{SI_i, ST_i, i = 1, 2, \dots, k\}$ apresentam um custo computacional de $C = N(k + 2)$, em que N é o número de realizações e k o número de parâmetros. Vale ressaltar, que atualmente, este procedimento é o que apresenta melhor eficiência computacional para computar índices de sensibilidade baseados puramente em avaliações do modelo.

Uma dificuldade que pode ser encontrada é a definição da distribuição de probabilidade dos parâmetros, bem como da média e da variância. Tal informação pode ser obtida de várias formas: tomada da literatura, derivada de observações experimentais pelo ajuste da função de uma distribuição empírica, baseada em opiniões de profissionais, definida como uma distribuição normal truncada (em que o truncamento é utilizado para evitar *outliers*) ou determinada de uma estrutura de correlação entre os parâmetros (caso exista). Geralmente esta informação é desconhecida, sendo usualmente adotada de forma arbitrária (KOU et al., 2005a; KOU et al., 2005b).

As etapas a seguir descrevem o algoritmo de SALTELLI (2002) empregado para o cálculo de índices de sensibilidade global.

(i). Geração das matrizes de valores dos parâmetros:

$$\begin{aligned} M_{j,k} &= \bar{\theta}_k + \text{Aleatório} \cdot \sigma_{\theta_k} & j &= 1..N & k &= 1..2NP \\ A_{j,i} &= M_{j,i} & j &= 1..N & i &= 1..NP \\ B_{j,i} &= M_{j,i+NP} & j &= 1..N & i &= 1..NP \end{aligned}$$

(ii). Definição da matriz de valores dos parâmetros considerando o parâmetro θ_m fixo (\mathbf{B}_{A_m}); esta matriz contém todos os elementos da matriz \mathbf{B} , exceto a m-ésima coluna que é tomada da matriz \mathbf{A} :

$$\begin{aligned} B_{A_j,i} &= B_{j,i} & j &= 1..N & i &= 1..NP & i \neq m \\ B_{A_j,m} &= A_{j,m} & j &= 1..N \end{aligned}$$

- (iii). Cálculo das repostas do modelo a partir das matrizes de experimentos simulados: \mathbf{y}_A , \mathbf{y}_B , \mathbf{y}_{BA_i} .
- (iv). Cômputo dos índices de sensibilidade para cada fator:

$$S_m = \frac{(1/N) \sum_j y_{A(j)} \cdot y_{BA_m(j)} - fo^2}{(1/N) \sum_j (y_{A(j)})^2 - fo^2},$$

$$ST_m = \frac{(1/N) \sum_j y_{B(j)} \cdot y_{BA_m(j)} - fo^2}{(1/N) \sum_j (y_{A(j)})^2 - fo^2};$$

$$\text{em que: } fo^2 = \left[(1/N) \sum_{j=1}^N y_{A(j)} \right]^2$$

A ideia básica do procedimento está ilustrada pela Figura 2.13. O procedimento consiste na geração de uma matriz M , a ser dividida em duas matrizes A e B . Na matriz M , cada coluna corresponde a um parâmetro e cada linha os valores dos parâmetros gerados aleatoriamente de acordo com uma distribuição de probabilidade previamente especificada.

A Figura 2.13 ilustra o cálculo dos índices de sensibilidade para o parâmetro 2 (θ_2). O índice de sensibilidade de efeito total do parâmetro sobre a predição é obtido fixando-se o parâmetro avaliado e variando os demais parâmetros, o que é feito ao se comparar a variância da matriz B_{A_2} com a variância da matriz A . De forma similar, o índice de sensibilidade de interação do parâmetro com os demais parâmetros é obtido variando-se o parâmetro avaliado e fixando os demais parâmetros, o que é feito ao se comparar a variância da matriz B_{A_2} com a variância da matriz B . Desta forma, é possível avaliar a sensibilidade do parâmetro.

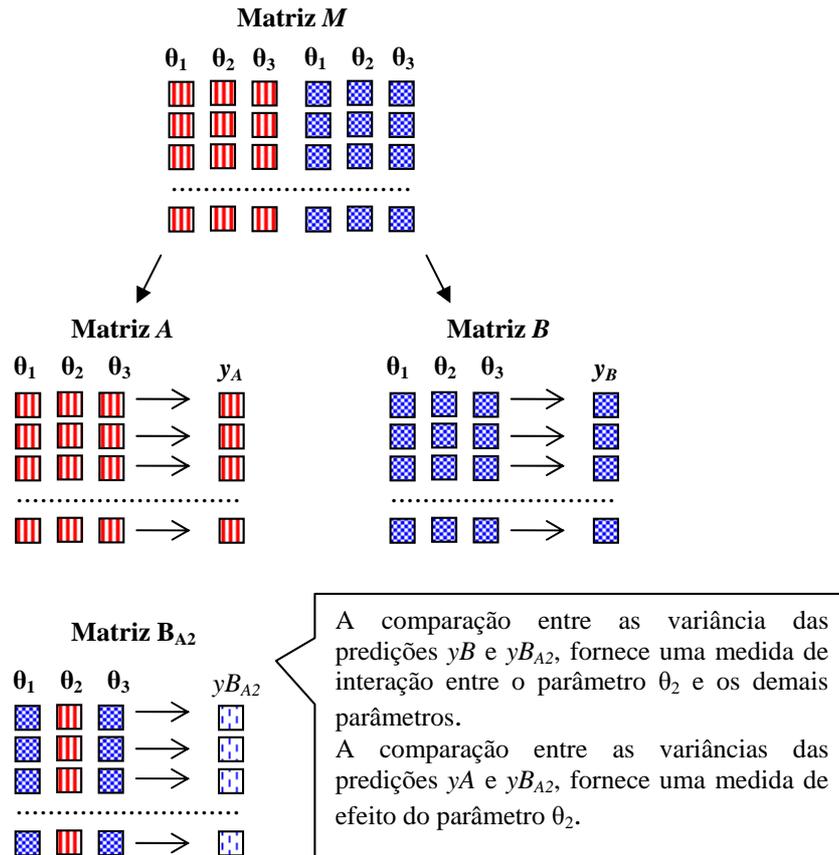


Figura 2.13 – Procedimento de identificabilidade de parâmetros via algoritmo de SALTELLI (2002).

2.5 CONCLUSÕES PARCIAIS

Baseado na revisão da literatura técnico-científica, algumas importantes conclusões foram obtidas e são apresentadas a seguir:

- As dificuldades do problema de estimação de parâmetro (e.g., problemas originados da escassez de observações experimentais, da estrutura inadequada do modelo e/ou de experimentos mal planejados) vêm sendo abordadas por meio de procedimentos de identificabilidade de parâmetros; como resultado, verifica-se a economia de recursos, tempo e esforço na realização de experimentos difíceis e/ou de alto custo que podem não ser suficientemente informativos para este procedimento.
- Os procedimentos de identificabilidade de parâmetros podem ser muito úteis quando observações experimentais são ausentes ou infrequentes. Nestes casos, em que conseqüentemente torna-se ainda mais difícil a caracterização de algumas grandezas estatísticas (e.g., a matriz de

covariância experimental), o procedimento de estimação de parâmetro pode apresentar problemas numéricos (e.g., matrizes singulares) dificultando ou até inviabilizando a obtenção da solução.

- Parece não haver uma metodologia padrão para avaliação do desempenho dos procedimentos de identificabilidade de parâmetros.
- Observa-se uma grande carência de informações a respeito da abrangência, bem como das peculiaridades dos métodos atuais de identificabilidade de parâmetros.
- Grande parte das metodologias não fornece um critério para determinar o número ótimo de parâmetros a serem selecionados para a realização do procedimento de estimação, sendo este critério usualmente adotado de forma arbitrária.
- Os parâmetros considerados não identificáveis terão seus valores mantidos fixos após os procedimentos de identificabilidade e re-estimação, ainda que seja apenas uma estimativa inicial fornecida arbitrariamente pelo usuário.

Mediante estas observações, vislumbra-se um amplo panorama para o estudo de procedimentos de identificabilidade de parâmetros, bem como para o desenvolvimento de novas metodologias.

CAPÍTULO 3. ESTUDO COMPARATIVO DE MÉTODOS PROPOSTOS NA LITERATURA

A literatura reporta vários procedimentos para seleção de parâmetros frente aos problemas de estimação mal condicionados. Dada a complexidade do tema, não há uma metodologia padrão para avaliação do desempenho destes procedimentos. Desta forma, observa-se uma grande carência de informações a respeito da abrangência, bem como das peculiaridades dos procedimentos atuais, que não raramente são aplicáveis a casos restritos. A fim de compreender melhor o funcionamento de tais procedimentos, neste Capítulo 3 foram investigadas três métodos reportados pela literatura, com diferentes propriedades de interesse, cujo desempenho foi avaliado frente a distintos problemas de identificabilidade.

3.1 ESTRATÉGIAS PARA ABORDAR A IDENTIFICABILIDADE DE PARÂMETROS

Baseado na literatura técnico-científica foram selecionadas três metodologias visando à investigação do potencial de identificabilidade dos parâmetros em diferentes cenários: a análise de sensibilidade global apresentada por SALTELLI (2002), e os procedimentos de identificabilidade propostos por YAO *et al.* (2003) e SECCHI *et al.* (2006); tais métodos estão descritos no Capítulo 2. Neste contexto, cada metodologia representa uma estratégia para abordar o problema de identificabilidade.

A primeira estratégia, SALTELLI (2002), consiste na análise de sensibilidade global baseada em métodos de decomposição de variâncias, apresentando como característica principal a amplitude de análise da influência dos parâmetros para a predição, analisando os mesmos em faixas de valores. Como resultado, a estratégia de SALTELLI (2002) fornece os parâmetros ordenados de acordo com a sua influência para a predição do modelo. Para o usuário desta análise, a facilidade de implementação é a principal vantagem; em contrapartida, apresenta um esforço computacional elevado quando comparado aos métodos de identificabilidade baseados em análise de sensibilidade local.

As demais estratégias, YAO *et al.* (2003) e SECCHI *et al.* (2006), são procedimentos de identificabilidade que utilizam critérios baseados na matriz de sensibilidade local B . Estas estratégias utilizam como critérios para a seleção e ordenamento de parâmetros, medidas de influência e interação sobre a predição. Tanto a estratégia de YAO *et al.* (2003) quanto a estratégia de SECCHI *et al.* (2006) apresentam um limite de corte para o número máximo de parâmetros a ser selecionados, sendo este limite dado pelo mau condicionamento da Matriz de Informação de Fisher (FIM). Desta forma, ambas as estratégias fornecem como resultados o número de parâmetros selecionados e a ordem de seleção destes parâmetros. Adicionalmente, a estratégia de SECCHI *et al.* (2006) realiza a estimação do conjunto de parâmetros selecionados a cada nova seleção, de forma que o usuário obtém os novos valores para os parâmetros selecionados ao final do procedimento de identificabilidade. Apesar de algumas semelhanças, as estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006) são conceitualmente diferentes, sendo que esta segunda apresenta maior complexidade de implementação.

3.2 ESTUDOS DE CASOS

Para fins de investigação do desempenho das estratégias de identificabilidade empregadas neste estudo, foram utilizados alguns estudos de casos. Os estudos de casos são compostos de modelos matemáticos de diferentes complexidade, visando apresentar as propriedades e ilustrar o funcionamento destas estratégias.

De maior simplicidade, os estudos de casos 1 e 2 apresentam modelos hipotéticos, para os quais foram empregados experimentos obtidos por planejamento fatorial ortogonal gerado pelo *software* Statistica versão 6. Abrangendo problemas mais complexos, os estudos de casos 3 e 4 apresentam os modelos empregados por SECCHI *et al.* (2006) em estudos anteriores de identificabilidade de parâmetros.

Para todos os estudos de caso, as estimativas iniciais fornecida aos parâmetros apresentam valores diferentes dos valores originais (exatos) dos mesmos, simulando o desconhecimento destes valores pelo usuário, o que na prática é muito comum. É importante ressaltar que para todos estes estudos de caso, com exceção do estudo de caso 4 em que os experimentos foram obtidos via experimental, as réplicas foram geradas sem considerar erros nas variáveis de entrada, usando os parâmetros exatos e adicionando às saídas ruídos com distribuição normal, com média zero e com desvio-padrão de 5%.

O cômputo dos índices de sensibilidade global via a estratégia de SALTELLI (2002) para os estudos de casos apresentados foi feito por meio da geração de perturbações aleatórias admitindo que o parâmetro apresenta distribuição de probabilidade normal, com desvio padrão (σ_{θ}) igual a 15% do valor médio do parâmetro (μ_{θ}), sendo este valor médio dado como a estimativa inicial do parâmetro. Para todos os estudos de casos, o número de realizações (N) utilizado foi 25000.

É importante observar que a estratégia de SALTELLI (2002), por se tratar de uma análise de sensibilidade, não apresenta um limite de corte para o número máximo de parâmetros selecionados, atuando apenas no ordenamento do parâmetro como seleção. Desta forma, pode ser possível que a introdução de todos os parâmetros selecionados resulte em procedimento de estimação de parâmetros mal condicionado. Buscando contornar esta dificuldade, nesta tese, esta estratégia foi adaptada ao problema de identificabilidade de parâmetros. Tal adaptação consiste em utilizar todos os parâmetros selecionados, retirando automaticamente um por vez os parâmetros que apresentam menor potencial de estimação, até que o procedimento de estimação de

parâmetros se torne possível, eliminando o mau condicionamento; a cada parâmetro retirado do conjunto de parâmetros selecionados avalia-se a inversibilidade da matriz *FIM*, de forma a serem mantidos neste conjunto apenas os parâmetros para os quais esta matriz é inversível.

Para fins de avaliação e de comparação do desempenho das estratégias de identificabilidade de parâmetros abordadas, os respectivos conjuntos de parâmetros selecionados por estas estratégias foram estimados. Desta forma, cada estratégia de identificabilidade de parâmetros apresenta novos valores para os parâmetros selecionados com os quais é possível verificar a qualidade do ajuste do modelo às observações experimentais; considera-se como mais adequada ao caso estudado, a estratégia de identificabilidade cujo conjunto de parâmetros selecionados após a estimação fornece o menor desvio entre o valor calculado com o modelo e o valor medido experimentalmente.

O procedimento de estimação de parâmetros foi realizado por meio do pacote computacional ESTIMA&PLANEJA (SCHWAAB et al., 2011), baseado no método Gauss-Newton. A estimação de parâmetros utilizando o método Gauss-Newton é um procedimento que inclui: (i) avaliação do modelo, (ii) *linsearch* (busca em linha) para cálculo do passo com verificação de faixa válida de valores dos parâmetros; (iii) avaliação da matriz *FIM* e sua inversa. Como cada uma destas etapas pode falhar interrompendo a estimação, nesta tese, algumas modificações foram introduzidas neste código para melhoria da convergência, a fim de tratar problemas de identificabilidade que apresentam modelos complexos e dados experimentais escassos.

Por simplicidade, a métrica proposta para avaliar a eficiência das metodologias identificadas é o desvio quadrático entre valores preditos e os valores medidos experimentalmente, normalizado pelos números de experimentos e de variáveis de saída, aqui chamado resíduo quadrático médio (*Res*) que são calculados conforme a equação a seguir:

$$(E\ 3.1) \quad Res = \sum_{i=1}^N \sum_{j=1}^{NY} \frac{(y_{i,j}^E - y_{i,j}^C)^2}{N \cdot NY}.$$

Outra proposta para a avaliação e comparação das estratégias é o ângulo de correlação (ANTON e RORRES, 2010). Propõe-se que os ângulos formados entre os

vetores de sensibilidade dos parâmetros sejam calculados como forma de averiguar a correlação paramétrica. Baseado na matriz de sensibilidade, o procedimento empregado para cálculo dos valores dos ângulos é ilustrado na Figura 3.1.

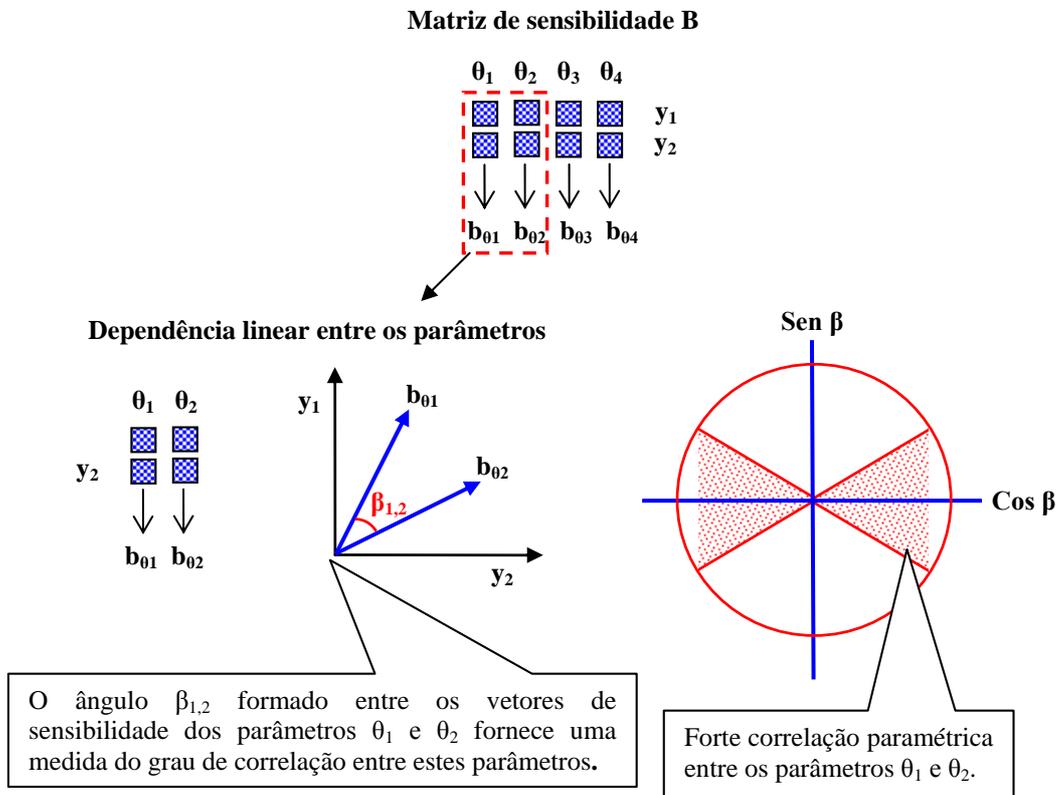


Figura 3.1 – Procedimento para cálculo dos ângulos entre os vetores de sensibilidade dos parâmetros.

Conforme indicado na Figura 3.1, os ângulos entre os vetores de sensibilidade que apresentam valores muito próximos a 0 ou a 180 graus correspondem a parâmetros com forte dependência linear. O ângulo entre dois vetores b_{θ_i} e b_{θ_j} pode ser calculado a partir da lei dos cossenos, como a seguir (ANTON e RORRES, 2010):

$$(E 3.2) \quad \beta_{i,j} = \arccos \left\{ \frac{\langle b_{\theta_i} \cdot b_{\theta_j} \rangle}{\|b_{\theta_i}\| \cdot \|b_{\theta_j}\|} \right\},$$

em que $\beta_{i,j}$ é o valor do ângulo formado entre o par de parâmetros avaliados indicados pelos índices i e j , e b_{θ_i} e b_{θ_j} são vetores que correspondem respectivamente às colunas dos parâmetros i e j da matriz de sensibilidade dos parâmetros.

3.2.1 Modelo hipotético 1

Buscando uma visualização inicial do problema de identificabilidade e do funcionamento das estratégias estudadas propôs-se o uso do modelo hipotético 1 como exemplo inicial, dada a sua simplicidade estrutural.

O modelo hipotético 1 apresenta estrutura linear, apresentada pela expressão a seguir:

$$(E\ 3.3) \quad y = \sum_{i=1}^7 \theta_i x_i$$

em que $y \in \mathcal{R}^1$, $x \in \mathcal{R}^7$ e $\theta \in \mathcal{R}^7$.

Uma vez que as metodologias de identificabilidade estão baseadas na influência do parâmetro sobre a predição, a identificabilidade dos parâmetros de modelos da natureza do modelo hipotético 1 foi avaliada para dois casos diferentes:

- (i). Caso 1: Os parâmetros apresentam valores diferentes e as faixas experimentais são idênticas para todas as variáveis.
- (ii). Caso 2: Os valores dos parâmetros são idênticos e as faixas experimentais para as variáveis variam.

Este exemplo ilustra diferentes situações de identificabilidade dos parâmetros. Para o caso 1 espera-se que os parâmetros selecionados sejam os parâmetros que apresentam maior valor, visto que estes parâmetros exercem maior influência sobre a predição. No caso 2 espera-se que os parâmetros selecionados sejam os parâmetros ligados às variáveis que apresentam a faixa mais ampla de variação, já que estas variáveis apresentam maior contribuição para a predição.

Caso 1

O conjunto experimental empregado neste primeiro caso é dado pelo planejamento fatorial com duas repetições apresentado na Tabela 3.1. Os valores de parâmetros empregados são apresentados na Tabela 3.2.

É importante observar que estes experimentos, obtidos pelo planejamento fatorial ortogonal, permitem a estimação de todos os parâmetros do modelo hipotético 1.

Tabela 3.1 – Experimentos empregados pelo modelo hipotético 1 para o caso 1.

N	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1
1	1	1	1	-1	-1	-1	1	-13,736
	1	1	1	-1	-1	-1	1	-14,436
2	1	1	-1	-1	1	1	-1	-3,941
	1	1	-1	-1	1	1	-1	-4,160
3	1	-1	1	1	-1	1	-1	-3,705
	1	-1	1	1	-1	1	-1	-4,188
4	1	-1	-1	1	1	-1	1	25,696
	1	-1	-1	1	1	-1	1	24,710
5	-1	1	1	1	1	-1	-1	-6,035
	-1	1	1	1	1	-1	-1	-6,063
6	-1	1	-1	1	-1	1	1	44,692
	-1	1	-1	1	-1	1	1	44,523
7	-1	-1	1	-1	1	1	1	68,619
	-1	-1	1	-1	1	1	1	60,775
8	-1	-1	-1	-1	-1	-1	-1	-104,140
	-1	-1	-1	-1	-1	-1	-1	-102,067

As três metodologias de identificabilidade de parâmetros selecionadas, SALTELLI (2002), YAO *et al.* (2003) e SECCHI *et al.* (2006) foram aplicadas para o modelo hipotético 1 para avaliação neste caso 1. Os resultados obtidos podem ser visualizados pela Figura 3.2 e Tabela 3.2.

Na Figura 3.2, as barras representam o potencial de estimação de cada parâmetro para a metodologia de SALTELLI (2002), de YAO *et al.* (2003) e de SECCHI *et al.* (2006), respectivamente. Nota-se que as três metodologias abordadas apresentaram os mesmos resultados quanto à seleção e ao ordenamento dos parâmetros do modelo hipotético 1 para o caso 1. Como esperado, observa-se que os parâmetros selecionados são os parâmetros que apresentam maior valor numérico, uma vez que os parâmetros e os experimentos não estão correlacionados e todas as variáveis variam em faixas iguais (com os valores mínimos -1 e máximos +1). Neste cenário ilustrado, o

valor dos parâmetros é o principal fator que influencia a predição, logo será o principal fator responsável pelo ordenamento dos parâmetros.

Na Tabela 3.2, a primeira linha apresenta o valor original dos parâmetros do modelo hipotético 1, o qual se pretende identificar empregando as estratégias de SALTELLI (2002), de YAO *et al.* (2003) e de SECCHI *et al.* (2006). Os valores iniciais dos parâmetros do modelo hipotético, com os quais foram realizados os procedimentos de estimação de parâmetros, são apresentados na segunda linha da Tabela 3.2. Ainda nesta tabela, os resultados idênticos para as três estratégias de identificabilidade de parâmetros são apresentados na última linha; nesta linha, os números que aparecem sobrescritos indicam a ordem de seleção de acordo com a significância dos parâmetros. Para este caso, dado ao número de graus de liberdade e a ortogonalidade do plano experimental, o procedimento de estimação apresentou bom condicionamento sendo a estimação de todos os parâmetros possível. As estratégias de identificabilidade demonstram ser capazes de obter o adequado ordenamento dos parâmetros quanto a sua influência sobre a predição, selecionando os parâmetros que apresentam maior valor numérico. Uma vez que todos os parâmetros foram selecionados e estimados, o resíduo quadrático médio apresentou valor baixo, indicando o sucesso do procedimento de estimação.

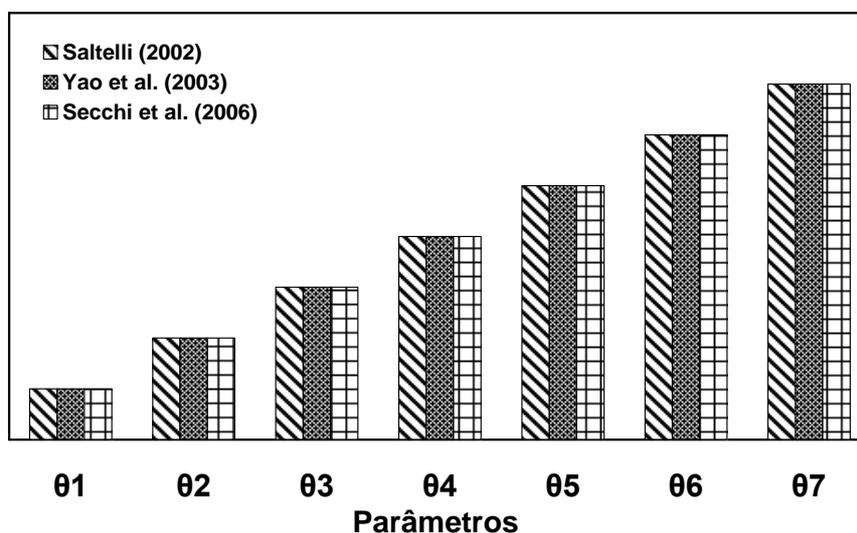


Figura 3.2 – Identificabilidade de parâmetros do modelo hipotético 1 para o caso 1.

Tabela 3.2 – Valores dos parâmetros utilizados e resultados obtidos para o modelo hipotético 1 para o caso 1.

Conjunto de parâmetros – Θ								Res
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	
Originais	1,0	5,0	10,0	15,0	20,0	25,0	30,0	0,229
Estimativas	1,0	2,0	3,0	4,0	5,0	6,0	10,0	1175,922
Novos*	0,971 ⁽⁷⁾	5,089 ⁽⁶⁾	10,190 ⁽⁵⁾	14,967 ⁽⁴⁾	20,060 ⁽³⁾	24,989 ⁽²⁾	30,171 ⁽¹⁾	0,150

* – Novos valores de parâmetros obtidos para as três estratégias de identificabilidade de parâmetros.

⁽¹⁾ – Posição do parâmetro quanto a sua identificabilidade.

A fim de avaliar a influência da estimativa inicial sobre o desempenho das estratégias de identificabilidade, também foi feita a análise deste caso 1 fornecendo outros valores como estimativas iniciais para os parâmetros do modelo hipotético 1. A dependência das estratégias de identificabilidade com relação às estimativas iniciais é uma característica importante, que pode auxiliar na identificação dos casos em que o uso destas metodologias é apropriado. Os resultados desta investigação são apresentados na Tabela 3.3 e na Figura 3.3.

Tabela 3.3 – Valores dos parâmetros utilizados e resultados obtidos para o modelo hipotético 1 para o caso 1.

Conjunto de parâmetros – Θ								Res
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	
Originais	1,0	5,0	10,0	15,0	20,0	25,0	30,0	0,229
Estimativas	10,0	10,0	10,0	10,0	10,0	10,0	10,0	863,240
Saltelli (2002)	0,971 ⁽⁷⁾	5,089 ⁽⁶⁾	10,190 ⁽⁵⁾	14,967 ⁽⁴⁾	20,060 ⁽³⁾	24,989 ⁽²⁾	30,171 ⁽¹⁾	0,150
Yao et al. (2003)	0,971 ⁽¹⁾	5,089 ⁽²⁾	10,190 ⁽³⁾	14,967 ⁽⁴⁾	20,060 ⁽⁵⁾	24,989 ⁽⁶⁾	30,171 ⁽⁷⁾	0,150
Secchi et al. (2006)	0,971 ⁽⁶⁾	5,089 ⁽⁵⁾	10,190 ⁽¹⁾	14,967 ⁽⁷⁾	20,060 ⁽³⁾	24,989 ⁽⁴⁾	30,171 ⁽²⁾	0,150

⁽¹⁾ – Posição do parâmetro quanto a sua identificabilidade.

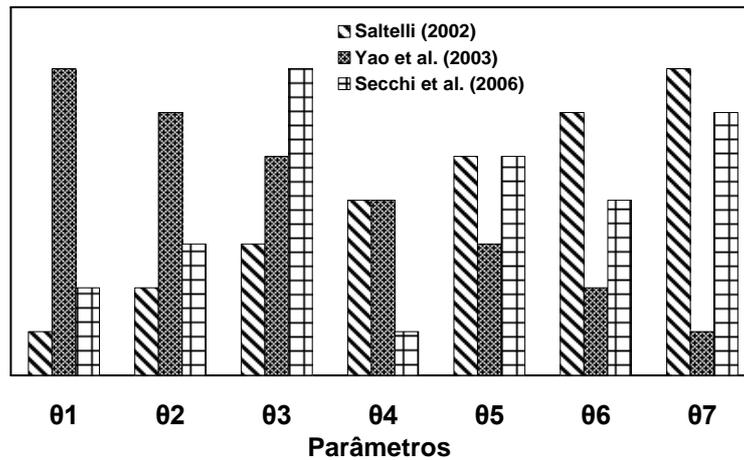


Figura 3.3 – Identificabilidade de parâmetros do modelo hipotético 1 para o caso 1, empregando novas estimativas iniciais.

Para essa investigação é fornecido um mesmo valor para a estimativa inicial de todos os parâmetros, diferentemente do que ocorre para o comportamento dos valores originais dos parâmetros que aumentam em ordem crescente do primeiro para o último parâmetro. Como resultado, as estratégias de identificabilidade forneceram diferentes ordens de seleção para os parâmetros.

Nota-se na Figura 3.3 e na Tabela 3.3 que apenas a estratégia de SALTELLI (2002) é capaz de identificar os parâmetros na ordem adequada de influência dos parâmetros sobre a predição. Este resultado está coerente com a correta identificabilidade dos parâmetros; esta observação apresenta uma importante característica da estratégia de SALTELLI (2002), que devido ao caráter global da análise de sensibilidade empregada, seu desempenho demonstra não ser afetado pela qualidade da estimativa inicial fornecida para os parâmetros, quando a correlação paramétrica é inexistente.

As estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006) utilizam a matriz de sensibilidade local normalizada como medida de influência dos parâmetros sobre a predição. Esta matriz por ser baseada no valor nominal dos parâmetros utiliza diretamente os valores das estimativas iniciais, que não sendo adequadas podem conduzir a conclusões incorretas sobre a identificabilidade dos parâmetros. O impacto das estimativas iniciais é bastante elevado sobre a estratégia de YAO *et al.* (2003). Para este estudo de caso, especificamente para o caso 1 em que todas as variáveis apresentam mesma faixa de valores, a estratégia de YAO *et al.* (2003) conclui que os parâmetros apresentam mesmo potencial de identificabilidade selecionando o número de

parâmetros de acordo com o grau de liberdade do problema e ordenando estes parâmetros sequencialmente do primeiro para o último parâmetro. Esta conclusão pode ser verificada pelos os resultados apresentados para a estratégia de YAO *et al.* (2003) na penúltima linha da Tabela 3.3. A estratégia de SECCHI *et al.* (2006) também demonstra forte dependência com as estimativas iniciais dos parâmetros, em especial com relação aos primeiros parâmetros selecionados. A medida que avança na seleção de parâmetros, a estratégia de SECCHI *et al.* (2006) demonstra ser capaz de identificar que os últimos parâmetros apresentam elevado potencial de estimação. Isto porque, a estratégia de SECCHI *et al.* (2006) realiza a estimação dos valores dos parâmetros selecionados a cada nova seleção, o que lhe confere maior robustez.

A investigação da identificabilidade dos parâmetros do modelo hipotético 1, para este caso, empregando valores diferentes como estimativa inicial para os parâmetros, revela que o sucesso das estratégias de identificabilidade empregadas baseadas em análise de sensibilidade local está fortemente vinculado à qualidade da informação disponível sobre os parâmetros. Vale ressaltar que para os procedimentos de identificabilidade que envolvem a estimação dos parâmetros selecionados a cada nova seleção, como a estratégia de SECCHI *et al.* (2006), a robustez do método de otimização também é de grande importância para a seleção adequada dos parâmetros.

Uma reformulação do caso 1 foi proposta considerando agora um número de experimentos menor que o número de parâmetros, sendo a Tabela 3.1 empregada até o quinto experimento ($N=5$). O objetivo é avaliar o desempenho das estratégias de identificabilidade de parâmetros selecionadas frente a cenário de informações escassas para a solução do problema. Para esta reformulação do caso 1 também foi investigada a influência das estimativas iniciais. Os resultados desta investigação são apresentados pela Figura 3.4 e pela Tabela 3.4.

Na Figura 3.4 percebe-se que alguns parâmetros não foram selecionados, dado ao grau de liberdade do problema que apresenta número de experimentos menor que o número de parâmetros a serem estimados. Nota-se também que a seleção e o ordenamento dos parâmetros identificáveis do modelo hipotético 1 alteram-se com os valores das estimativas iniciais. Parâmetros que são considerados não identificáveis para um conjunto de estimativas iniciais podem ser ditos identificáveis quando outras estimativas iniciais são empregadas, de modo que a qualidade da informação sobre os parâmetros tem forte influência no resultado do procedimento identificabilidade.

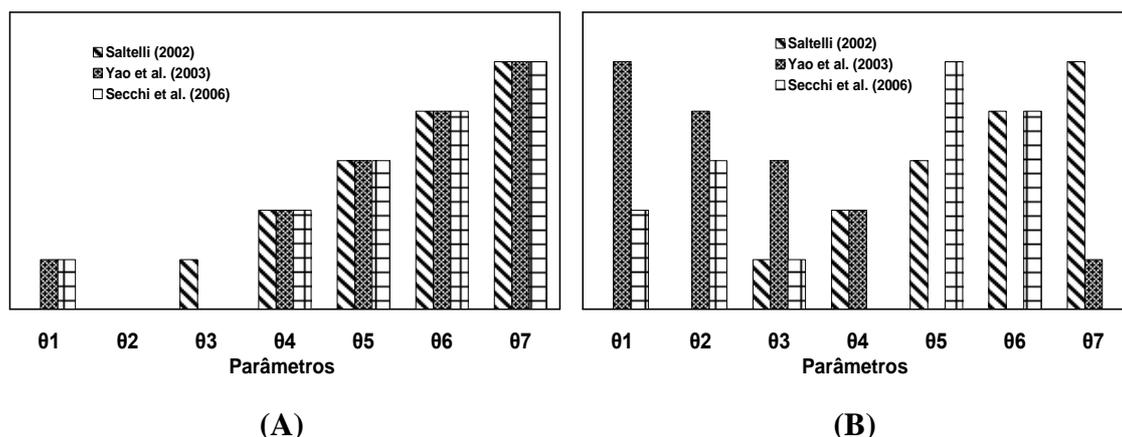


Figura 3.4 – Identificabilidade de parâmetros do modelo hipotético 1 para o caso 1 empregando diferentes estimativas iniciais (A: valores crescentes, B: valores uniformes), quando o número de parâmetros é maior que o número de experimentos.

Tabela 3.4 – Valores dos parâmetros utilizados pelo modelo hipotético 1 para o caso 1, quando o número de parâmetros é maior que o número de experimentos.

Conjunto de parâmetro – θ								
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	Res
Originais	1,0	5,0	10,0	15,0	20,0	25,0	30,0	0,223
Estimativas	1,0	2,0	3,0	4,0	5,0	6,0	10,0	30,041
Saltelli (2002)	<u>10,0</u> ⁽⁻⁾	<u>10,0</u> ⁽⁻⁾	0,269 ⁽⁵⁾	11,848 ⁽⁴⁾	10,065 ⁽³⁾	12,227 ⁽²⁾	17,005 ⁽¹⁾	0,038
Yao et al. (2003)	0,780 ⁽⁵⁾	<u>10,0</u> ⁽⁻⁾	<u>10,0</u> ⁽⁻⁾	11,848 ⁽⁴⁾	12,796 ⁽³⁾	15,068 ⁽²⁾	19,846 ⁽¹⁾	3×10^{-10}
Secchi et al. (2006)	0,780 ⁽⁵⁾	<u>10,0</u> ⁽⁻⁾	<u>10,0</u> ⁽⁻⁾	11,848 ⁽⁴⁾	12,796 ⁽³⁾	15,068 ⁽²⁾	19,846 ⁽¹⁾	3×10^{-10}
Estimativas	10,0	10,0	10,0	10,0	10,0	10,0	10,0	183,662
Saltelli (2002)	<u>10,0</u> ⁽⁻⁾	<u>10,0</u> ⁽⁻⁾	0,269 ⁽⁵⁾	11,848 ⁽⁴⁾	10,065 ⁽³⁾	12,227 ⁽²⁾	17,005 ⁽¹⁾	0,038
Yao et al. (2003)	0,780 ⁽¹⁾	0,152 ⁽²⁾	0,203 ⁽³⁾	10,423 ⁽⁴⁾	<u>10,0</u> ⁽⁻⁾	<u>10,0</u> ⁽⁻⁾	15,201 ⁽⁵⁾	2×10^{-9}
Secchi et al. (2006)	0,780 ⁽⁴⁾	0,152 ⁽³⁾	-4,998 ⁽⁵⁾	<u>10,0</u> ⁽⁻⁾	4,798 ⁽¹⁾	5,221 ⁽²⁾	<u>10,0</u> ⁽⁻⁾	3×10^{-9}

⁽¹⁾ – Posição do parâmetro quanto a sua identificabilidade.

⁽⁻⁾ – Parâmetro não selecionado.

Em particular para o modelo hipotético 1, quando as estimativas iniciais apresentam informações consistentes sobre os valores originais dos parâmetros (e.g.,

ordem de magnitude e/ou ordem de grandeza), mesmo apresentando valores bastante distantes dos valores originais dos parâmetros como apresentado na Tabela 3.4, os resultados da seleção e do ordenamento dos parâmetros se mostram semelhantes para as estratégias de identificabilidade.

Considerando que as estimativas iniciais empregadas para os resultados de identificabilidade apresentados no gráfico A da Figura 3.5 sejam mais adequadas que o outro conjunto de estimativas empregadas pode-se dizer que a eliminação dos três últimos experimentos do conjunto experimental dado pela Tabela 3.1 não permite a identificabilidade do parâmetro 2. Ainda considerando este gráfico, a não identificabilidade do parâmetro 3 pelas estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006) sugere que o novo conjunto de experimentos contendo 5 experimentos não ortogonais possivelmente introduz correlação deste parâmetro com outro(s) parâmetro(s) já selecionado(s); daí este parâmetro não ser selecionado por estas estratégias e ser selecionado pela estratégia de SALTELLI (2002), visto que este parâmetro apresenta maior influência que o parâmetro 1 sobre a predição.

É importante ressaltar que ao eliminar os três últimos experimentos do conjunto original dado pela Tabela 3.1, além de reduzir o grau de liberdade do procedimento de estimação tornando-o mal condicionado, o novo conjunto experimental perde a característica da ortogonalidade podendo apresentar variáveis cujos efeitos apresentam-se fundidos, o que pode ocasionar correlação entre alguns parâmetros.

Como foi observado anteriormente, o resultado da comparação dos gráficos A e B da Figura 3.4 mostra que a estratégia de SALTELLI (2002) parece não sofrer grande influência das estimativas iniciais, mantendo os mesmos resultados de identificabilidade dos parâmetros do modelo hipotético 1 quando diferentes estimativas são usadas, enquanto as estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006) se mostram fortemente dependentes destes valores. Portanto, uma vez que não se tenha informações confiáveis sobre os parâmetros pode ser preferível utilizar a estratégia de SALTELLI (2002), dado o seu caráter de análise global.

Foi realizada uma avaliação dos ângulos de dependência linear dos parâmetros do modelo hipotético 1, tanto mantendo o número original de experimentos ($N=8$) quanto para quando o número de experimentos é menor que o número de parâmetros ($N=5$). Como o cálculo da matriz de sensibilidade (não normalizada) para modelos lineares independe dos valores dos parâmetros, os valores das estimativas iniciais não

influenciam no cálculo dos ângulos de dependência linear dos parâmetros destes modelos. Os resultados são apresentados na Tabela 3.5.

Como indicado na Tabela 3.5, para o caso 1 em que o modelo é linear e o plano experimental é ortogonal, os parâmetros não apresentam dependência linear obtendo-se valores de 90° para os ângulos, o que indica que todos os parâmetros são passíveis de estimação. Ao eliminar três experimentos, o plano experimental perde a ortogonalidade introduzindo correlação paramétrica entre os parâmetros.

Tabela 3.5 - Ângulos (graus) de dependência linear dos parâmetros do modelo hipotético 1 para o caso 1, para as diferentes planos experimentais.

Caso 1 - Número de experimentos original (N=8)						
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
θ_1	90	90	90	90	90	90
θ_2		90	90	90	90	90
θ_3			90	90	90	90
θ_4				90	90	90
θ_5					90	90
θ_6						90
Caso 1 - Número de experimentos menor que o número de parâmetros (N=5)						
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
θ_1	102	102	102	102	78,5	78,5
θ_2		78,5	127	78,5	102	102
θ_3			78,5	127	102	102
θ_4				78,5	102	102
θ_5					102	102
θ_6						127

Caso 2

Para o caso 2, o conjunto de observações experimentais empregados pelo modelo hipotético 1 é apresentado pela Tabela 3.6. Nota-se que novamente foram empregados experimentos obtidos via planejamento fatorial ortogonal, sendo todos os parâmetros do modelo hipotético 1 passíveis de estimação.

Aplicaram-se as metodologias de identificabilidade de parâmetros ao caso 2 para o modelo hipotético 1. A Figura 3.5 e a Tabela 3.7 apresentam os resultados obtidos, onde as barras representam o potencial de estimação de cada parâmetro de acordo com a metodologia indicada pela legenda. Como esperado, o potencial de estimação dos parâmetros, fixos a um mesmo valor, está relacionado aos parâmetros ligados às variáveis que apresentam a faixa mais ampla de variação, visto que estas são as variáveis que mais contribuem para a predição. Como neste caso a faixa das variáveis cresce com o índice da variável, os parâmetros foram selecionados e ordenados decrescentemente quanto à identificabilidade, do último para o primeiro, como pode ser visualizado na Figura 3.5.

Tabela 3.6 – Experimentos empregados pelo Modelo hipotético 1 para o caso 2.

N	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1
1	1	2	-3	4	-5	-6	7	-0,753
	1	2	-3	4	-5	-6	7	0,177
2	1	2	-3	-4	5	6	-7	-0,187
	1	2	-3	-4	5	6	-7	-0,745
3	1	-2	3	4	-5	6	-7	-0,083
	1	-2	3	4	-5	6	-7	-0,815
4	1	-2	3	-4	5	-6	7	3,772
	1	-2	3	-4	5	-6	7	4,427
5	-1	2	3	4	5	-6	-7	-0,322
	-1	2	3	4	5	-6	-7	0,177
6	-1	2	3	-4	-5	6	7	8,473
	-1	2	3	-4	-5	6	7	7,915
7	-1	-2	-3	4	5	6	7	15,512
	-1	-2	-3	4	5	6	7	15,867
8	-1	-2	-3	-4	-5	-6	-7	-28,770
	-1	-2	-3	-4	-5	-6	-7	-28,597

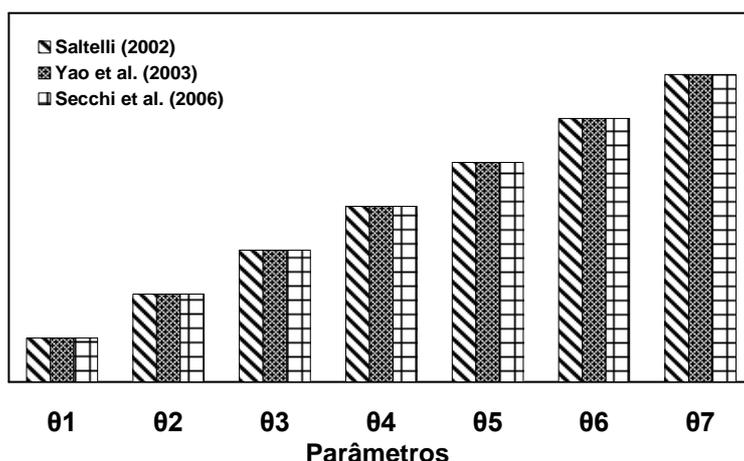


Figura 3.5 – Identificabilidade de parâmetros do modelo hipotético 1 para o caso 2.

Na Tabela 3.7, são apresentados os valores originais, as estimativas iniciais e os novos valores dos parâmetros. A sequência de ordenamento dos parâmetros, apresentada em sobrescrito, é a mesma para os três procedimentos de identificabilidade estudados. O resíduo quadrático normalizado calculado para o chute inicial é bastante elevado, porém como todos os parâmetros são estimáveis, os resíduos obtidos após a estimação dos parâmetros é da ordem dos resíduos obtidos para os parâmetros originais.

Tabela 3.7 – Valores dos parâmetros utilizados pelo modelo hipotético 1 para o caso 2.

Conjunto de parâmetro – θ								Res
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	
Originais	1,0	1,0	1,0	1,0	1,0	1,0	1,0	0,229
Estimativas	10,0	10,0	10,0	10,0	10,0	10,0	10,0	11303,896
Novos*	0,971 ⁽⁷⁾	1,044 ⁽⁶⁾	1,063 ⁽⁵⁾	0,992 ⁽⁴⁾	1,012 ⁽³⁾	0,998 ⁽²⁾	1,024 ⁽¹⁾	0,150

* – Novos valores de parâmetros obtidos para as três estratégias de identificabilidade de parâmetros.

⁽¹⁾ – Posição do parâmetro quanto a sua identificabilidade.

Como realizado para o caso 1 deste estudo de caso, uma reformulação do caso 2 foi proposta considerando agora um número de experimentos menor que o número de parâmetros, sendo a Tabela 3.6 empregada até o quarto experimento ($N=4$). O objetivo é avaliar o desempenho das estratégias de identificabilidade de parâmetros selecionadas frente a um cenário de informações escassas para a solução do problema. Os resultados para esta reformulação do caso 2 são apresentados pela Figura 3.6 e Tabela 3.8.

Com o número de experimentos menor que o número de parâmetros, o grau de liberdade é negativo e alguns parâmetros não podem ser estimados pelo procedimento convencional de estimação de parâmetros. Deste modo, as estratégias de identificabilidade selecionaram apenas o número de parâmetros para qual a estimação é possível, sendo descartados os parâmetros que apresentam menor potencial de identificabilidade.

Na Figura 3.6, observa-se que apenas os parâmetro 4 e 1 foram selecionado pelas três estratégias de identificabilidade. Embora estejam ligados às variáveis que apresentam menor faixa de variação, a estratégia de SALTELLI (2002) e YAO *et al.* (2003) selecionaram os parâmetros 1, 2, 3 e 4 como parâmetros identificáveis; a seleção contrária foi feita por SECCHI *et al.* (2006), que considera identificáveis os parâmetros 7, 5, 4, e 1. Acredita-se que para este caso, o resultado das estratégias de seleção dos parâmetros está ligada ao uso da informação experimental feito por estas estratégias, bem como à influência das estimativas iniciais. Ambas as estratégias de identificabilidade de SALTELLI (2002) e de YAO *et al.* (2003) obtêm a seleção dos parâmetros em ordem crescente, possivelmente devido ao fato dos parâmetros apresentarem valores iguais como estimativa inicial.

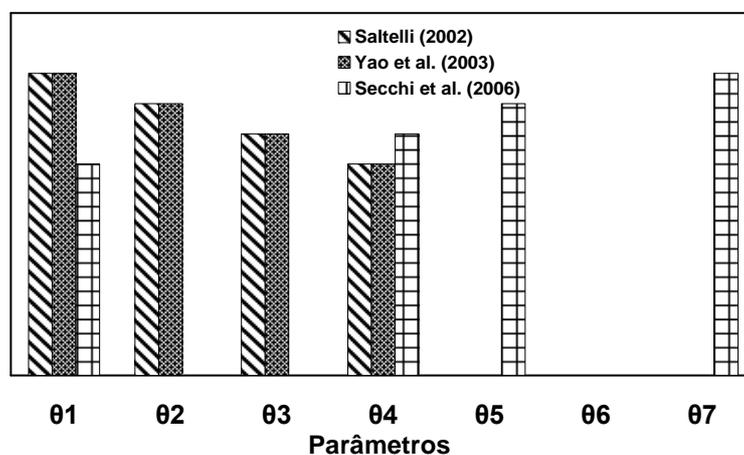


Figura 3.6 – Identificabilidade de parâmetros do modelo hipotético 1 para o caso 2 quando o número de experimentos é menor que o número de parâmetros.

A análise de resíduos, apresentada na Tabela 3.8, sugere que a seleção de parâmetros obtida pelas três estratégias de identificabilidade são equiparáveis, muito embora os parâmetros selecionados por SALTELLI (2002) e de YAO *et al.* (2003) não sejam os mais significativos. Para esta reformulação do caso 2, os parâmetros selecionados também foram estimados obtendo um novo conjunto de valores de

parâmetros, onde os valores dos parâmetros não selecionados, apresentados em sublinhado na Tabela 3.8, são mantidos fixos na estimativa inicial. De acordo com os valores dos resíduos apresentados na Tabela 3.8, comparando os valores obtidos para as estimativas iniciais com os valores obtidos após a estimação dos parâmetros, verifica-se que o uso das estratégias de identificabilidade permite melhorar o ajuste do modelo ao viabilizar a estimação de um subconjunto de parâmetros. Após o procedimento de estimação de parâmetros os resíduos ainda permanecem altos, devido à diferença entre as estimativas iniciais dos parâmetros não selecionados e os valores originais destes parâmetros que permanecem fixos na estimativa fornecida.

Tabela 3.8 – Valores dos parâmetros utilizados pelo modelo hipotético 1 para o caso 2, quando o número de parâmetros é maior que o número de experimentos.

	Conjunto de parâmetro – θ							Res
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	
Originais	1,0	1,0	1,0	1,0	1,0	1,0	1,0	0,262
Estimativas	10,0	10,0	10,0	10,0	10,0	10,0	10,0	322,467
Saltelli (2002)	0,724 ⁽¹⁾	7,661 ⁽²⁾	5,474 ⁽³⁾	12,227 ⁽⁴⁾	<u>10,0⁽⁻⁾</u>	<u>10,0⁽⁻⁾</u>	<u>10,0⁽⁻⁾</u>	77,898
Yao et al. (2003)	0,724 ⁽¹⁾	7,661 ⁽²⁾	5,474 ⁽³⁾	12,227 ⁽⁴⁾	<u>10,0⁽⁻⁾</u>	<u>10,0⁽⁻⁾</u>	<u>10,0⁽⁻⁾</u>	77,898
Secchi et al. (2006)	0,724 ⁽⁴⁾	<u>10,0⁽⁻⁾</u>	<u>10,0⁽⁻⁾</u>	6,361 ⁽³⁾	5,308 ⁽²⁾	<u>10,0⁽⁻⁾</u>	8,740 ⁽¹⁾	79,324

⁽¹⁾ – Posição dos parâmetros quanto a sua identificabilidade.

⁽⁻⁾ – Parâmetros não selecionados.

Para este caso também foram avaliados os ângulos de dependência linear dos parâmetros do modelo hipotético 1, tanto mantendo o número original de experimentos ($N=8$) quanto para quando o número de experimentos é menor que o número de parâmetros ($N=4$). Os resultados são apresentados na Tabela 3.9.

Como indicado na Tabela 3.9, para o caso 1 em que o modelo é linear e o plano experimental é ortogonal, os parâmetros não apresentam dependência linear obtendo-se valores de 90° para os ângulos, o que indica que todos os parâmetros são passíveis de estimação. Ao eliminar 4 experimentos, o plano experimental perde a ortogonalidade introduzindo forte correlação paramétrica entre os pares de parâmetros {2,4}, {5,3} e {7,6}. Tais resultados estão de acordo com a seleção de parâmetros

realizada por SECCHI et al. (2006), demonstrando que esta estratégia obteve sucesso na identificabilidade dos parâmetros mais significativos e menos correlacionados.

Tabela 3.9 - Ângulos (graus) de dependência linear dos parâmetros do modelo hipotético 1 para o caso 2, para diferentes planos experimentais.

Caso 2 - Número de experimentos original ($N=8$)						
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
θ_1	90	90	90	90	90	90
θ_2		90	90	90	90	90
θ_3			90	90	90	90
θ_4				90	90	90
θ_5					90	90
θ_6						90

Caso 2 - Número de experimentos menor que o número de parâmetros ($N=4$)						
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
θ_1	90	90	90	90	90	90
θ_2	0	90	<u>180</u>	90	90	90
θ_3	0	0	90	<u>180</u>	90	90
θ_4	0	0	0	90	90	90
θ_5	0	0	0	0	90	90
θ_6	0	0	0	0	0	<u>180</u>

3.2.2 Modelo hipotético 2

O modelo hipotético 2 é representado por um modelo linear nos parâmetros, cujo planejamento experimental não permite a estimação dos parâmetros de forma independente. A Equação (E 3.4) apresenta a estrutura matemática do modelo hipotético 2.

$$(E\ 3.4) \quad y = \sum_{i=1}^3 \theta_i x_i + \theta_4 x_1 x_2$$

em que $y \in \mathcal{R}^1$, $x \in \mathcal{R}^3$ e $\theta \in \mathcal{R}^4$.

O conjunto de observações experimentais utilizado para este estudo de caso é composto por 5 experimentos (N), sendo um destes no ponto central, cada qual contendo 2 réplicas, é dado pela Tabela 3.10.

Tabela 3.10 – Experimentos empregados pelo modelo hipotético 2.

N	x_1	x_2	x_3	y_1
1	1	1	1	89,247
	1	1	1	90,177
2	1	-1	-1	-30,187
	1	-1	-1	-30,745
3	0	0	0	-0,083
	0	0	0	-0,815
4	-1	1	-1	-40,223
	-1	1	-1	-39,573
5	-1	-1	1	-20,322
	-1	-1	1	-19,823

Tabela 3.11 – Valores dos parâmetros utilizados pelo modelo hipotético 2.

Conjunto de parâmetro – Θ					
	θ_1	θ_2	θ_3	θ_4	<i>Res</i>
Originais	30,0	25,0	20,0	15,0	0,223
Estimativas	4,0	3,0	2,0	1,0	1742,450
Saltelli (2002)	29,804 ⁽¹⁾	<u>3,0</u> ^(c)	18,018 ⁽²⁾	16,982 ⁽³⁾	390
Yao et al. (2003)	29,804 ⁽¹⁾	25,088 ⁽²⁾	34,001 ⁽³⁾	<u>1,0</u> ^(c)	0,067
Secchi et al. (2006)	29,804 ⁽¹⁾	25,088 ⁽²⁾	34,001 ⁽³⁾	<u>1,0</u> ^(c)	0,067

⁽¹⁾ – Posição do parâmetro quanto a sua identificabilidade.

^(c) – Parâmetro considerado não identificável.

O conjunto de experimentos apresentados pela Tabela 3.10 introduz uma dependência linear entre os parâmetros 3 e 4 durante o procedimento de estimação; a variável x_3 é sempre igual ao resultado da multiplicação das variáveis x_1 e x_2 . Desta forma, o modelo poderia ser reescrito como:

$$(E 3.5) \quad y = \theta_1 x_1 + \theta_2 x_2 + (\theta_4 + \theta_3) x_3,$$

indicando que os parâmetros θ_3 e θ_4 não podem ser estimados independentemente. Portanto, para este estudo de caso, espera-se que a fusão de efeitos das variáveis x_1 e x_2 dificulte a identificabilidade dos parâmetros, sobretudo do quarto parâmetro.

A identificabilidade dos parâmetros do modelo hipotético 2 foi estudada empregando as estratégias de SALTELLI (2002), YAO *et al.* (2003) e SECCHI *et al.* (2006). Os resultados são apresentados pelas Tabela 3.11 e Figura 3.7.

Na Figura 3.7 observa-se que o resultado para a identificabilidade de parâmetros do modelo hipotético 2 é o mesmo para as estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006), que caracterizam o parâmetro 4 como parâmetro não identificável. Estas estratégias são capazes de identificar a correlação paramétrica introduzida pelo plano experimental e a estimação dos parâmetros considerados identificáveis por elas resulta em um bom ajuste do modelo às observações experimentais, como observado na Tabela 3.11. Para a estratégia de SALTELLI (2002) os parâmetros 3 e 4 são considerados parâmetros identificáveis, dada a grande influência destes parâmetros sobre a predição. A seleção realizada por esta estratégia resulta em um procedimento mal condicionado de estimação de parâmetros, onde os novos valores dos parâmetros obtidos não fornecem um ajuste adequado do modelo às observações experimentais, o que pode ser verificado pelo valor do resíduo apresentado na Tabela 3.11.

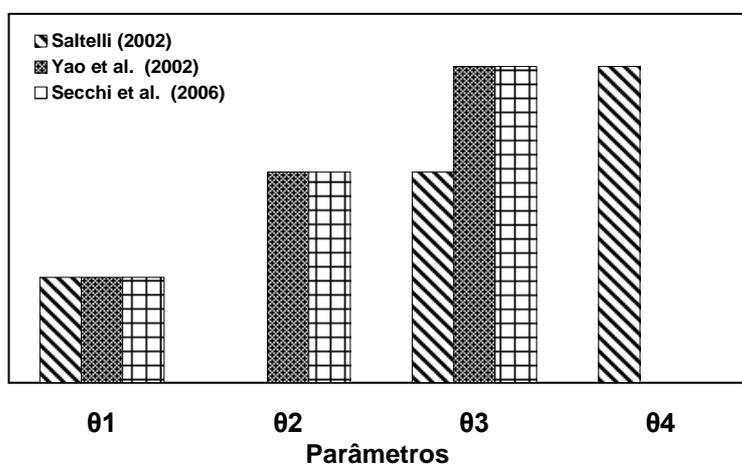


Figura 3.7 – Identificabilidade de parâmetros do modelo hipotético 2.

A estratégia de SALTELLI (2002) consiste em uma análise de sensibilidade bastante empregada para averiguação da influência dos fatores sobre a predição e quando aplicada à identificabilidade de parâmetros não leva em conta informações sobre o condicionamento do procedimento de estimação dos parâmetros selecionados, podendo levar à estimação de parâmetros ruidosos sem significado estatístico e físico ou tornando este procedimento inviável. As estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006) finalizam o procedimento de identificabilidade de parâmetros quando o problema de estimação torna-se mal condicionado, selecionando apenas os parâmetros para os quais a estimação é numericamente possível.

Na Tabela 3.12 são apresentados os ângulos de dependência linear entre os parâmetros do modelo hipotético 2. Como indicado nesta tabela, a medida do grau de dependência linear entre os parâmetros é capaz de identificar a forte correlação entre os parâmetros 3 e 4, sendo que estes parâmetros não podem ser estimados simultaneamente neste caso.

Tabela 3.12 – Ângulo (em graus) de dependência linear entre os parâmetros do modelo hipotético 2.

	θ_2	θ_3	θ_4
θ_1	90	90	90
θ_2		90	90
θ_3			<u>0</u>

Os resultados da identificabilidade do modelo hipotético 2 demonstram que os parâmetros mais influentes do modelo não necessariamente são os parâmetros mais identificáveis, uma vez que a correlação paramétrica pode tornar o procedimento de estimação mal condicionado, podendo até mesmo inviabilizá-lo.

3.2.3 Modelo hipotético 3

O modelo hipotético 3, desenvolvido por SECCHI *et al.* (2006), reúne características que podem tornar bastante complexo o procedimento de estimação dos seus parâmetros. Este modelo apresenta uma estrutura matemática não-linear contendo parâmetros fortemente correlacionados, não sendo incomum a obtenção de regiões de confiança abertas para estes parâmetros; as Equações (E 3.6), (E 3.7) e (E 3.8) descrevem a estrutura do modelo hipotético.

$$(E 3.6) \quad y_1 = \theta_1 x_2 x_3 e^{-\theta_2/x_1} + \theta_3 x_2 x_4 e^{-\theta_4/x_1},$$

$$(E 3.7) \quad y_2 = 1 - \theta_1 x_2 x_3 e^{-\theta_2/x_1} + \theta_5 x_3 e^{-\theta_6/x_1},$$

$$(E 3.8) \quad y_3 = \theta_7 x_1 + \theta_8 (\theta_1 x_2 x_3 e^{-\theta_2/x_1} + \theta_5 x_3 e^{-\theta_6/x_1}),$$

com $y \in \mathbb{R}^3$, $x \in \mathbb{R}^4$ e $\theta \in \mathbb{R}^8$.

Apresentados pela Tabela 3.13, os conjuntos de observações experimentais são compostos por 3 pontos operacionais (OP), cada qual contendo 3 réplicas.

Na Tabela 3.13, os casos apresentados ilustram diferentes cenários experimentais frente ao problema de estimação de parâmetros. No caso 1, a variável de entrada mais importante, x_1 , é mantida constante, reduzindo a capacidade de estimativa das medições. No caso 3, os dois últimos pontos operacionais estão correlacionados. O caso 2 representa o conjunto de experimentos mais favoráveis à estimação dos parâmetros do modelo hipotético 3 dentre os três casos.

A Figura 3.8 e a Tabela 3.14 apresentam os resultados obtidos para as estratégias de identificabilidade dos parâmetros do modelo hipotético 3.

Na Tabela 3.14, a primeira coluna apresenta a ordem original dos parâmetros e as três últimas apresentam, respectivamente, os novos valores dos parâmetros selecionados pelas estratégias de SALTELLI (2002), YAO et al. (2003) e SECCHI et al. (2006). Os valores iniciais dos parâmetros do modelo hipotético 3, com os quais foram realizados os procedimentos de re-estimação de parâmetros, são apresentados na terceira coluna desta tabela. Os números que aparecem sobrescritos nesta tabela indicam a ordem de seleção, de acordo com a significância do parâmetro obtida por cada método; a presença de um traço no lugar do número indica que a ordem do parâmetro é desconhecida para a estratégia (parâmetro não selecionado). Ainda nesta tabela, os parâmetros que apresentam baixa sensibilidade aparecem sublinhados; os demais parâmetros foram estimados.

Representado pela barra, os potenciais de estimação dos parâmetros é dado na Figura 3.8. Nesta figura, os gráficos A, B e C apresentam os resultados para o caso 1, o caso 2 e o caso 3, respectivamente, que ilustram diferentes cenários experimentais. Nota-se as diferenças entre os resultados das estratégias de identificabilidade de parâmetros, como o número de parâmetros selecionados, a posição dos parâmetros selecionados e a ordem de seleção, apenas para o caso 1. Nos demais casos a seleção é a

mesma para todas as estratégias, caso 2 todos os parâmetros são selecionados e caso 3 apenas o parâmetro 6 é não selecionado.

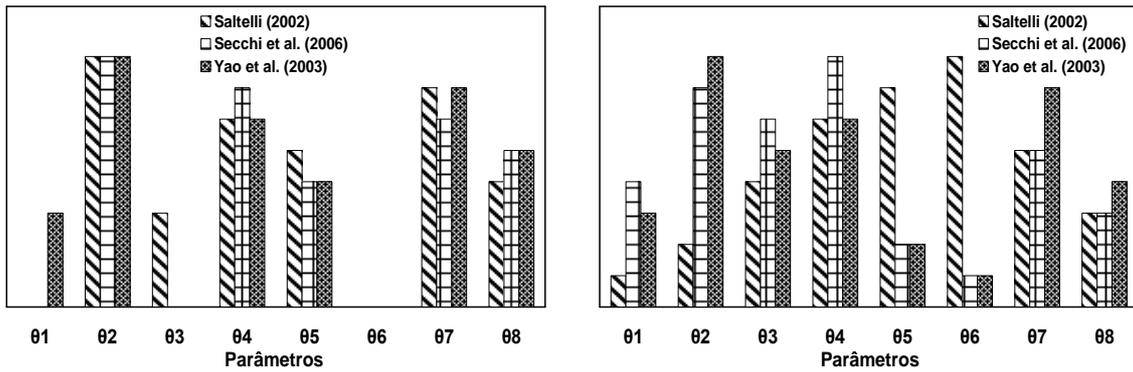
Naturalmente, nos casos 2 e 3, em que a seleção é a mesma para todas as estratégias, o procedimento de estimação levará aos mesmos valores de resíduo quadrático normalizado apresentada na Tabela 3.14. Todavia, embora a seleção das estratégias não seja concordante para o caso 1, esta análise do resíduo quadrático normalizado indica que o desempenho das estratégias foi equiparável.

O acompanhamento dos resultados das etapas de seleção e estimação a cada parâmetro selecionado apenas pode ser feito por meio da estratégia de SECCHI et al. (2006), dada a natureza do método (estimação e seleção simultâneas). Desta forma, no caso 1 os parâmetros 1 e 3 estão fortemente correlacionados com os parâmetros 2, 4, 5, 7, e 8. Uma vez que x_1 é constante neste conjunto de dados, a correlação entre os pares {1,2}, {3,4} e {5,6} é total. Por este motivo os parâmetros 1, 3 e 6, não foram selecionados pela estratégia de SECCHI et al. (2006). Neste caso 1, a estratégia de SALTELLI (2002) inclui o parâmetro 3 e a estratégia de YAO et al. (2003) inclui o parâmetro 1, que não afetam o valor da função objetivo devido à correlação entre os demais parâmetros, mas faz com que o valor destes parâmetros correlacionados sejam diferentes entre os três procedimentos. Assim, a seleção de quaisquer um dos três parâmetros (1, 3 e 6) é incorreta, o que indica que a estratégia de SECCHI et al. (2006) obteve os resultados mais adequados.

Em se tratando de modelos não lineares, a interpretação dos resultados de metodologias de identificabilidade pode ser bastante difícil. É sabido que a não-linearidade pode implicar em dificuldades adicionais, tais como a multiplicidade de soluções, a alta correlação entre os parâmetros e a presença de parâmetros não significativos. Considerando que a correlação paramétrica pode ser inferida, com base nas estimativas iniciais, os ângulos de dependência linear entre os parâmetros foram calculados e os resultados podem ser visualizados na Tabela 3.15. Nesta Tabela 3.15, os valores de ângulos que sugerem forte correlação entre os parâmetros se apresentam em **negrito e sublinhado**. Ângulos próximos a 0 e a 180 graus sugerem que os parâmetros estão fortemente correlacionados. Estes valores foram observados para todos os casos, sendo mais presente nos casos 1 e 3 onde os conjuntos de experimentos foram mal planejados.

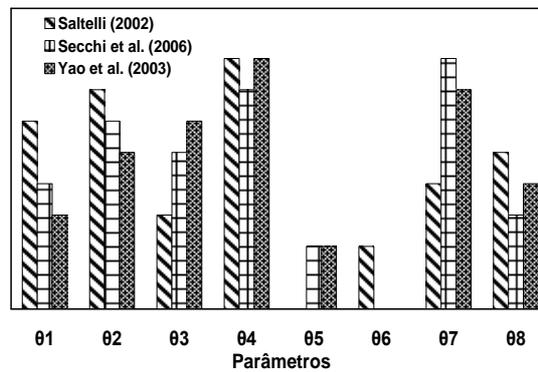
Tabela 3.13– Dados experimentais empregados pelo modelo hipotético 3.

	Variável	OP1	OP2	OP3	
Caso 1	x_1	0,98	0,98	0,98	
	x_2	0,73	0,13	0,43	
	x_3	0,23	0,45	0,72	
	x_4	0,67	0,47	0,13	
	y_1		0,676	0,178	0,758
			0,700	0,174	0,765
			0,710	1,175	0,762
	y_2		0,623	0,933	0,332
			0,621	0,919	0,338
			0,614	0,911	0,341
	y_3		7,810	6,871	9,348
			7,508	6,959	9,093
			7,383	6,903	8,889
	Caso 2	x_1	0,98	0,52	0,75
		x_2	0,73	0,13	0,43
x_3		0,23	0,45	0,72	
x_4		0,67	0,47	0,13	
y_1			0,676	0,058	0,537
			0,700	0,058	0,517
			0,710	0,059	0,536
y_2			0,623	0,954	0,527
			0,621	0,969	0,545
			0,614	0,975	0,539
y_3			7,810	3,417	6,573
			7,508	3,655	6,800
			7,383	3,532	6,781
Caso 3		x_1	0,98	0,52	0,52
		x_2	0,73	0,13	0,13
	x_3	0,23	0,45	0,45	
	x_4	0,67	0,47	0,57	
	y_1		0,676	0,058	0,059
			0,700	0,058	0,059
			0,710	0,059	0,058
	y_2		0,623	0,954	0,988
			0,621	0,969	0,984
			0,614	0,975	1,002
	y_3		7,810	3,417	3,489
			7,508	3,655	3,507
			7,383	3,532	3,494



(A) – Modelo hipotético 3: Caso 1

(B) – Modelo hipotético 3: Caso 2



(C) – Modelo hipotético: Caso 3

Figura 3.8 – Identificabilidade de parâmetros do modelo hipotético 3.

Para o caso 1, os resultados da Tabela 3.15 demonstram que os pares de parâmetros $\{1, 2\}$, $\{3, 4\}$ e $\{5, 6\}$, apresentam efeitos fortemente dependente. Desta forma, espera-se que mediante a seleção de um dos parâmetros dos pares correlacionados, o outro parâmetro não seja selecionado, mantendo o bom condicionamento do procedimento de estimação de parâmetros. Dentre as três estratégias, SECCHI *et al.* (2006) foi capaz de identificar esta correlação e por ter selecionado os parâmetros 2, 4 e 5, possivelmente devido ao grau de importância destes parâmetros para a predição, não foi possível selecionar os parâmetros 1, 3 e 6. As estratégias de SALTELLI (2002) e de YAO *et al.* (2003) selecionaram 6 parâmetros quando apenas 5 eram estimáveis, o que indica que a incerteza experimental calculada através de números aleatórios pode ter interferido na avaliação do condicionamento da matriz de covariância neste caso. Para este caso, apenas o parâmetro 6 é igualmente dito não identificável pelas três estratégias de identificabilidade.

Tabela 3.14 – Valores dos parâmetros empregados para o modelo hipotético 3.

Caso 1

Parâmetro	Original	Estimativa inicial	Saltelli (2002)	Yao et al. (2003)	Secchi et al. (2006)
θ_1	7,65	6,5	<u>6,5</u> ⁽⁻⁾	8,031 ⁽⁶⁾	<u>6,5</u> ⁽⁻⁾
θ_2	1,15	2,4	0,989 ⁽¹⁾	1,196 ⁽¹⁾	0,989 ⁽¹⁾
θ_3	3,89	2,7	3,580 ⁽⁶⁾	<u>2,7</u> ⁽⁻⁾	<u>2,7</u> ⁽⁻⁾
θ_4	1,75	1,5	1,740 ⁽³⁾	1,462 ⁽³⁾	1,462 ⁽²⁾
θ_5	0,23	0,01	0,123 ⁽⁴⁾	0,122 ⁽⁵⁾	0,112 ⁽⁵⁾
θ_6	0,79	0,15	<u>0,15</u> ⁽⁻⁾	<u>0,15</u> ⁽⁻⁾	<u>0,15</u> ⁽⁻⁾
θ_7	6,32	4,25	6,407 ⁽²⁾	6,408 ⁽²⁾	6,407 ⁽³⁾
θ_8	3,42	5,5	3,738 ⁽⁵⁾	3,740 ⁽⁴⁾	3,743 ⁽⁴⁾
Res	0,012	3,478	7×10^{-3}	7×10^{-3}	7×10^{-3}

Caso 2

Parâmetro	Original	Estimativa inicial	Saltelli (2002)	Yao et al. (2003)	Secchi et al. (2006)
θ_1	7,65	6,5	4,913 ⁽⁵⁾	4,913 ⁽⁶⁾	4,913 ⁽⁵⁾
θ_2	1,15	2,4	0,800 ⁽²⁾	0,800 ⁽¹⁾	0,800 ⁽²⁾
θ_3	3,89	2,7	9,433 ⁽³⁾	9,433 ⁽⁴⁾	9,433 ⁽³⁾
θ_4	1,75	1,5	2,585 ⁽¹⁾	2,585 ⁽³⁾	2,585 ⁽¹⁾
θ_5	0,23	0,01	0,136 ⁽⁷⁾	0,136 ⁽⁷⁾	0,136 ⁽⁷⁾
θ_6	0,79	0,15	0,587 ⁽⁸⁾	0,587 ⁽⁸⁾	0,587 ⁽⁸⁾
θ_7	6,32	4,25	6,341 ⁽⁴⁾	6,341 ⁽²⁾	6,341 ⁽⁴⁾
θ_8	3,42	5,5	3,609 ⁽⁶⁾	3,609 ⁽⁵⁾	3,609 ⁽⁶⁾
Res	0,007	2,207	$1,572 \times 10^{-4}$	$1,572 \times 10^{-4}$	$1,572 \times 10^{-4}$

Caso 3

Parâmetro	Original	Estimativa inicial	Saltelli (2002)	Yao et al. (2003)	Secchi et al. (2006)
θ_1	7,65	6,5	7,590 ⁽⁵⁾	7,590 ⁽⁶⁾	7,590 ⁽⁵⁾
θ_2	1,15	2,4	1,143 ⁽³⁾	1,143 ⁽⁴⁾	1,143 ⁽³⁾
θ_3	3,89	2,7	3,082 ⁽⁴⁾	3,082 ⁽³⁾	3,082 ⁽⁴⁾
θ_4	1,75	1,5	1,587 ⁽²⁾	1,587 ⁽¹⁾	1,587 ⁽²⁾
θ_5	0,23	0,01	0,083 ⁽⁷⁾	0,083 ⁽⁷⁾	0,083 ⁽⁷⁾
θ_6	0,79	0,15	<u>0,15</u> ⁽⁻⁾	<u>0,15</u> ⁽⁻⁾	<u>0,15</u> ⁽⁻⁾
θ_7	6,32	4,25	6,467 ⁽¹⁾	6,467 ⁽²⁾	6,467 ⁽¹⁾
θ_8	3,42	5,5	3,096 ⁽⁶⁾	3,096 ⁽⁵⁾	3,096 ⁽⁶⁾
Res	0,005	1,310	$1,611 \times 10^{-4}$	$1,611 \times 10^{-4}$	$1,611 \times 10^{-4}$

⁽¹⁾ – Posição do parâmetro quanto a sua identificabilidade.

⁽⁻⁾ – Parâmetro considerado não identificável.

No caso 2, os resultados da Tabela 3.15 sugerem forte correlação entre os pares de parâmetros {1, 2}, {3, 4} e {5, 6}. Neste caso, a estimação de todos os parâmetros do modelo hipotético 3 é possível e a consequência da correlação é visualizada na ordem

de seleção dos parâmetros selecionados; por exemplo os parâmetros 2 e 3, apresentam ordem de seleção 2 e 1 para a estratégia de SALTELLI (2002) e 3 e 4 para a estratégia de YAO *et al.* (2003).

O caso 3 apresenta resultado de identificabilidade dos parâmetros bastante similares para as três estratégias, sendo o parâmetro 6 considerado não identificável. Semelhante ao que ocorre no caso 2, a consequência da correlação é visualizada no ordenamento dos parâmetros selecionados, que apresentando os mesmos efeitos sobre a predição são deslocados em uma posição.

Tabela 3.15 – Ângulo (em graus) de dependência linear entre os parâmetros do modelo hipotético 3 calculados para os três casos experimentais.

Caso 1							
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
θ_1	<u>180</u>	84,1	95,9	<u>32,4</u>	<u>148</u>	<u>32,9</u>	<u>14,4</u>
θ_2		95,9	84,1	<u>148</u>	<u>32,4</u>	<u>147</u>	<u>166</u>
θ_3			<u>180</u>	90	90	90	90
θ_4				90	90	90	90
θ_5					<u>180</u>	<u>25,3</u>	<u>24,9</u>
θ_6						<u>155</u>	<u>155</u>
θ_7							<u>29,4</u>
Caso 2							
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
θ_1	<u>172</u>	81,9	98,2	<u>44,7</u>	128	<u>25,6</u>	<u>14,6</u>
θ_2		97,2	82,6	<u>140</u>	48	<u>153</u>	<u>164</u>
θ_3			<u>178</u>	90	90	90	90
θ_4				90	90	90	90
θ_5					<u>169</u>	<u>32,6</u>	<u>38,7</u>
θ_6						<u>143</u>	133
θ_7							<u>20,3</u>
Caso 3							
	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
θ_1	<u>177</u>	79,9	100	66,4	104	<u>36,3</u>	<u>14,7</u>
θ_2		100	79,9	116	73	<u>146</u>	<u>166</u>
θ_3			<u>177</u>	90	90	90	90
θ_4				90	90	90	90
θ_5					<u>170</u>	<u>32,2</u>	62
θ_6						<u>138</u>	108
θ_7							<u>30,9</u>

3.2.4 Modelo cinético de crescimento celular

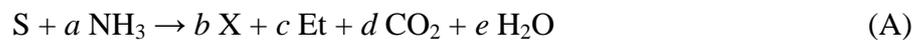
Desenvolvido por LONGHI *et al.* (2004), o modelo cinético de crescimento celular considera as seguintes hipóteses:

- A fase biótica é única (o modelo é não estruturado).

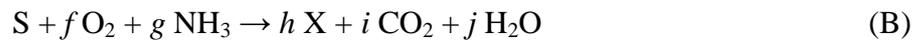
- b) A composição elementar da biomassa não muda significativamente.
- c) A lactose é metabolizada tanto aerobicamente como anaerobicamente, com diferentes taxas de eficiências.
- d) O etanol é metabolizado somente aerobicamente.
- e) Não há reserva de carboidrato intracelular.
- f) A fonte de nitrogênio está em excesso e não há limitação de substrato.
- g) Os principais produtos são: a biomassa, o CO₂, o H₂O e o etanol; sendo este último metabolicamente significativo apenas para o subproduto.

Três rotas metabólicas principais da cultura são consideradas:

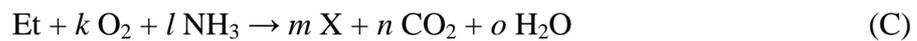
a) Fermentação:



b) Respiração I – Oxidação da lactose:



c) Respiração II – Oxidação do etanol:



em que X representa células disponíveis, S representa lactose – C₁₂H₂₂O₁₁, O₂ representa o oxigênio, Et representa o etanol – C₂H₆O. Os coeficientes estequiométricos desconhecidos são representados por *a, b, c, d, e, f, g, h, i, j, k, l, m, n* e *o* que são usados para definir os principais coeficientes de rendimento (sobre uma base mássica) para o sistema:

$$Y_{X/S}^{Ferm} = b \frac{\text{Massa molar da Biomassa}}{\text{Massa molar da Lactose}} \quad (\text{D})$$

$$Y_{X/E}^{Ferm} = \frac{b}{c} \frac{\text{Massa molar da Biomassa}}{\text{Massa molar do Etanol}} = \frac{Y_{X/S}^{Ferm}}{\phi_{X/S}^{Ferm}} \quad (\text{E})$$

$$Y_{X/S}^{Oxid} = h \frac{\text{Massa molar da Biomassa}}{\text{Massa molar da Lactose}} \quad (\text{F})$$

$$Y_{X/O}^S = \frac{h \text{ Massa molar da Biomassa}}{f \text{ Massa molar do Oxigênio}} = \frac{Y_{X/S}^{Oxid}}{\phi_{X/S}^{Oxid}} \quad (\text{G})$$

$$Y_{X/E}^{Oxid} = m \frac{\text{Massa molar da Biomassa}}{\text{Massa molar do Etanol}} \quad (\text{H})$$

$$Y_{X/O}^{Et} = \frac{m \text{ Massa molar da Biomassa}}{k \text{ Massa molar do Oxigênio}} = \frac{Y_{X/E}^{Oxid}}{\phi_{X/E}^{Oxid}} \quad (\text{I})$$

em que $\phi_{X/S}^{Ferm}$, $\phi_{X/S}^{Oxid}$ e $\phi_{X/E}^{Oxid}$ são razões estequiométricas relativas à concentração celular, Caso a composição celular seja desconhecida, todos os coeficientes do produto devem ser estimados; caso contrário, somente os três coeficientes ($Y_{X/S}^{Ferm}$, $Y_{X/S}^{Oxid}$ e $Y_{X/E}^{Oxid}$) necessitam ser estimados e os demais podem ser obtidos pela análise estequiométrica.

As taxas de limitação do crescimento de substrato (expressas em $\text{g.l}^{-1}.\text{h}^{-1}$), para as rotas metabólicas são dadas por:

a) Fermentação:

$$(E 3.9) \quad r_{Ferm} = \frac{\mu_{1máx} S}{k_1 + S} X$$

b) Respiração I:

$$(E 3.10) \quad r_{Oxid}^S = \frac{\mu_{2máx} S C_L}{(k_3 + S)(k_{Ox1} + C_L)} X$$

c) Respiração II:

$$(E 3.11) \quad r_{Oxid}^{Et} = \frac{\mu_{3máx} Et C_L}{(k_2 + Et)(k_{Ox2} + C_L)} X$$

em que C_L é a concentração de oxigênio dissolvido na fase líquida ($\text{g.l}^{-1}.\text{h}^{-1}$), Et é a concentração de etanol ($\text{g.l}^{-1}.\text{h}^{-1}$); $\mu_{1\text{máx}}, \mu_{2\text{máx}}, \mu_{3\text{máx}}, k_1, k_2, k_3, k_{Ox1}, k_{Ox2}$ são parâmetros cinéticos do modelo.

O pH foi mantido constante e a fonte de nitrogênio foi adicionada em excesso em relação à fonte de carbono.

Com base no modelo cinético acima e considerando que o biorreator opera em modo batelada em condições isotérmicas, aerado e com mistura perfeita, os balanços de massa para os componentes podem ser escritos como a seguir:

a) Biomassa:

$$(E 3.12) \quad \frac{dX}{dt} = r_{Ferm} + r_{Oxid}^S + r_{Oxid}^{Et} - r_d$$

b) Substrato:

$$(E 3.13) \quad \frac{dS}{dt} = \frac{1}{Y_{X/S}^{Ferm}} r_{Ferm} - \frac{1}{Y_{X/S}^{Oxid}} r_{Oxid}^S ;$$

c) Oxigênio na fase líquida:

$$(E 3.14) \quad \frac{dC_L}{dt} = k_l a (C_{Le} - C_L) - \left[\frac{\phi_{X/S}^{Oxid}}{Y_{X/S}^{Oxid}} r_{Oxid}^S + \frac{\phi_{X/E}^{Oxid}}{Y_{X/E}^{Oxid}} r_{Oxid}^{Et} \right] ;$$

d) Oxigênio na fase gasosa:

$$(E 3.15) \quad \frac{dC_G}{dt} = \frac{F_{ar}}{V_G} (C_{Go} - C_G) - k_l a (C_{Le} - C_L) \frac{V_L}{V_G} ;$$

e) Etanol:

$$(E 3.16) \quad \frac{dEt}{dt} = \frac{\phi_{X/S}^{Ferm}}{Y_{X/S}^{Ferm}} r_{Ferm} + \frac{1}{Y_{X/E}^{Oxid}} r_{Oxid}^{Et} ;$$

em que C_{Le} é a concentração de oxigênio na fase líquida em equilíbrio com o gás ($\text{g.l}^{-1}.\text{h}^{-1}$); C_G é a concentração de oxigênio na fase gasosa ($\text{g.l}^{-1}.\text{h}^{-1}$); k_1a é o coeficiente de transferência de massa gás-líquido de oxigênio (h^{-1}); r_d é o coeficiente da taxa específica da morte celular ($\text{g.l}^{-1}.\text{h}^{-1}$); F_{ar} é a taxa de fluxo de ar injetado no biorreator (l.h^{-1}); V_G é o volume de gás interno ao biorreator; V_L é o volume de líquido no biorreator (l); C_{Go} é a concentração no ar do fluxo alimentado (g.l^{-1}).

A concentração de oxigênio na fase líquida em equilíbrio com o gás é dada por:

$$(E 3.17) \quad C_{Le} = KC_G$$

e K é o coeficiente de equilíbrio (adimensional) , definido como:

$$(E 3.18) \quad K = \frac{C_m}{C_{Go}},$$

em que C_m é a concentração de oxigênio saturado na fase líquida (g.l^{-1}) que é função da temperatura e da pressão. A equação de C_m foi obtida pela regressão de dados experimentais para água a pressão atmosférica, resultando em:

$$(E 3.19) \quad C_m = 3,62e^{-6}T^2 - 3,312e^{-4}T + 0,01466217.$$

A concentração de gás alimentado é obtida por meio da seguinte equação, considerando gás ideal:

$$(E 3.20) \quad C_{Go} = \frac{P\bar{M}x_{O_2}}{R(T + 273,15)}$$

em que T é a temperatura do biorreator ($^{\circ}\text{C}$); x_{O_2} é a fração molar de oxigênio alimentado pelo gás (adimensional); P é a pressão do biorreator (bar); \bar{M} é o massa molar do oxigênio (g.mol^{-1}); R é a constante universal dos gases ($\text{bar.l.mol}^{-1}.\text{K}^{-1}$).

O conjunto de observações experimentais utilizado por LONGHI *et al.* (2004) é apresentado pela Tabela 3.16.

Tabela 3.16 – Observações experimentais utilizadas pelo modelo de cinética de crescimento.

N	Tempo (h)	X (g.l ⁻¹)	S (g.l ⁻¹)	Et (g.l ⁻¹)	PO ₂ (%)
1	0	0,16	48,90	0,000	102,5
2	2	0,19	51,14	0,263	95,5
3	4	0,30	47,35	0,149	81,9
4	6	1,68	46,43	0,215	25,5
5	8	6,59	33,00	2,126	0,8
6	10	13,09	9,88	11,057	0,2
7	15	20,42	0,30	6,750	0,1
8	24	22,74	0,00*	0,000*	91,2
9	27	23,11	0,00*	0,000*	100,6
10	30	22,65	0,00*	0,000*	102,9

* – Erros devido à imprecisão do dispositivo de medição, valores não empregados na estimação.

Na Figura 3.9 apresentam-se os dados experimentais e o resultado do modelo usando uma estimativa inicial dos parâmetros, nas mesmas condições usadas por SECCHI et al. (2006) e obtidas experimentalmente por LONGHI et al. (2004). O pacote computacional DASSL (PETZOLD, 1982) foi empregado para a integração numérica requerida à resolução deste modelo matemático.

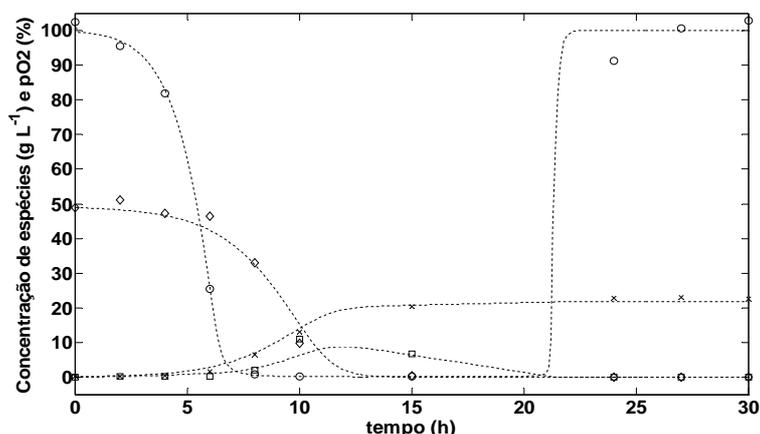


Figura 3.9 - Observações experimentais. Concentrações de: biomassa (x), substrato (◇), etanol (□), oxigênio dissolvido (○); e calculados com os parâmetros originais (– –).

Os resultados obtidos para a identificabilidade de parâmetros do modelo de LONGHI et al. (2004) são apresentados pela Tabela 3.17 e pela Figura 3.10. Nesta tabela, as três últimas colunas apresentam, respectivamente, os novos valores dos

parâmetros selecionados pelas estratégias de SALTELLI (2002), YAO et al. (2003) e SECCHI et al. (2006).

Dada a natureza do modelo, não se conhecem os valores e a ordem exatos dos parâmetros. Os valores iniciais dos parâmetros do modelo, com os quais foram realizados os procedimentos de re-estimação de parâmetros, são apresentados na segunda coluna da Tabela 3.17. Os números que aparecem sobrescritos nesta tabela indicam a ordem de seleção, de acordo com a significância do parâmetro obtida por cada método; a presença de um traço no lugar do número indica que a ordem do parâmetro é desconhecida para a estratégia (parâmetro não selecionado). Ainda nestas tabelas, os parâmetros que apresentam baixa sensibilidade aparecem sublinhados; os demais parâmetros foram re-estimados.

Tabela 3.17 – Comparação das estratégias de identificabilidade de parâmetros para o modelo de LONGHI et al. (2004).

Parâmetro	Estimativa Inicial	Saltelli (2002)	Yao et al. (2003)	Secchi et al. (2006)
$\Theta_1 = \mu_{1máx}$	0,60	0,614 ⁽¹⁾	0,617 ⁽³⁾	0,615 ⁽²⁾
$\Theta_2 = \mu_{2máx}$	0,06	0,059 ⁽⁷⁾	<u>0,06</u> ⁽⁻⁾	<u>0,06</u> ⁽⁻⁾
$\Theta_3 = \mu_{3máx}$	0,16	<u>0,16</u> ⁽⁻⁾	<u>0,16</u> ⁽⁻⁾	0,167 ⁽¹⁰⁾
$\theta_4 = k_1$	20,00	19,766 ⁽⁴⁾	19,665 ⁽⁴⁾	19,973 ⁽⁴⁾
$\theta_5 = k_{ox}$	$1,0 \times 10^{-4}$	$1,0 \times 10^{-4(10)}$	$1,0 \times 10^{-4(1)}$	$1,0 \times 10^{-4(1)}$
$\theta_6 = k_2$	4,26	4,254 ⁽⁹⁾	4,208 ⁽²⁾	4,214 ⁽³⁾
$\theta_7 = \phi_{X/S}^{Oxid} / Y_{X/S}^{Oxid}$	0,63	<u>0,63</u> ⁽⁻⁾	<u>0,63</u> ⁽⁻⁾	<u>0,63</u> ⁽⁻⁾
$\theta_8 = \phi_{X/E}^{Oxid} / Y_{X/E}^{Oxid}$	6,20	6,22 ⁽⁶⁾	<u>6,20</u> ⁽⁻⁾	6,313 ⁽⁷⁾
$\theta_9 = 1/Y_{X/S}^{Ferm}$	2,44	2,460 ⁽²⁾	<u>2,44</u> ⁽⁻⁾	2,483 ⁽⁵⁾
$\theta_{10} = 1/Y_{X/S}^{Oxid}$	2,63	2,635 ⁽⁸⁾	<u>2,63</u> ⁽⁻⁾	2,650 ⁽⁹⁾
$\theta_{11} = \phi_{X/S}^{Ferm} / Y_{X/S}^{Ferm}$	0,85	0,907 ⁽³⁾	<u>0,85</u> ⁽⁻⁾	0,873 ⁽⁶⁾
$\theta_{12} = 1/Y_{X/E}^{Oxid}$	6,67	6,660 ⁽⁵⁾	<u>6,67</u> ⁽⁻⁾	6,527 ⁽⁸⁾
Res	8,826	4,140	4,496	4,032

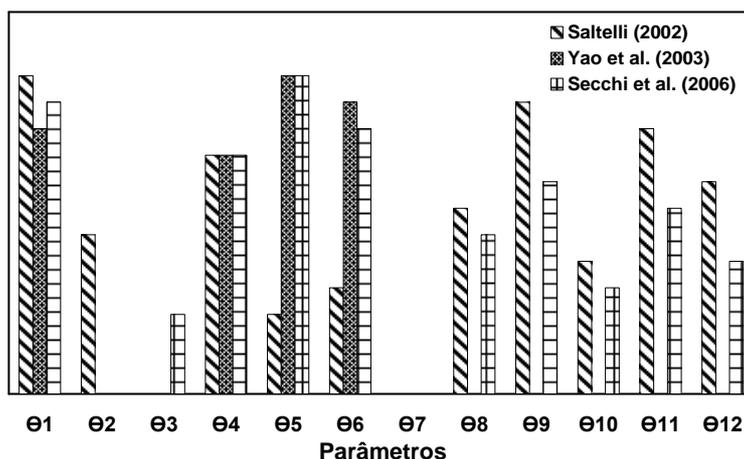


Figura 3.10 - Identificabilidade de parâmetros do modelo de LONGHI *et al.* (2004).

Representado pela barra, os potenciais de estimação dos parâmetros é dado na Figura 3.10. A comparação dos resultados das estratégias de identificabilidade demonstra discordância quanto ao potencial de identificabilidade dos parâmetros. Como indicado na Tabela 3.17, as estratégias de SALTELLI (2002) e de SECCHI *et al.* (2006) consideram apenas não identificável 2 parâmetros, enquanto a estratégia de YAO *et al.* (2003) considera não identificável 8 parâmetros. Todavia, a análise do resíduo quadrático normalizado apresenta valores da mesma ordem, o que indica que embora a seleção destas estratégias apresente resultados diferentes o desempenho foi equiparável.

A Figura 3.11 ilustra os resultados obtidos para a estimação de parâmetros a partir dos parâmetros selecionados pela estratégia de SALTELLI (2002). É possível observar melhoria na predição do modelo entre os tempos 20 e 25, onde é possível obter predições que se ajustam melhor à curva experimental da concentração de oxigênio dissolvido. O mesmo comportamento foi observado para estratégia de SECCHI *et al.* (2003), visualizado na Figura 3.13. Como indicado na Figura 3.12, a estimação dos parâmetros selecionados pela estratégia de YAO *et al.* (2003) não ocasionou mudanças no comportamento do modelo.

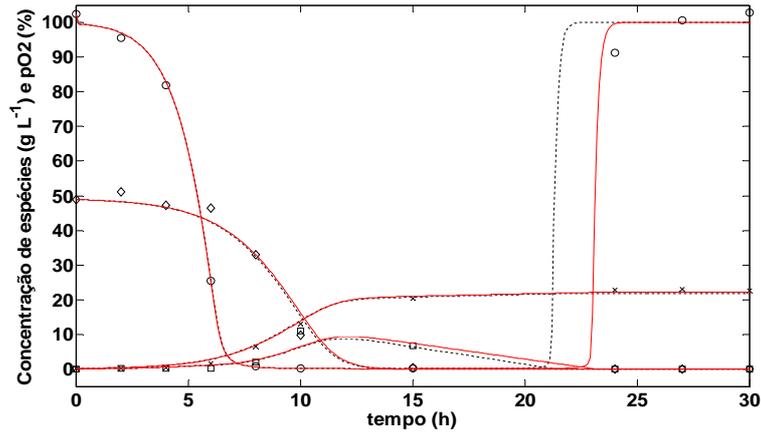


Figura 3.11 – Dados experimentais. Concentrações de: biomassa (×), substrato (◇), etanol (□), oxigênio dissolvido (○); e calculados – parâmetros originais (---) e após a estimação dos parâmetros selecionados pela estratégia de SALTELLI *et al.* (2002) (—).

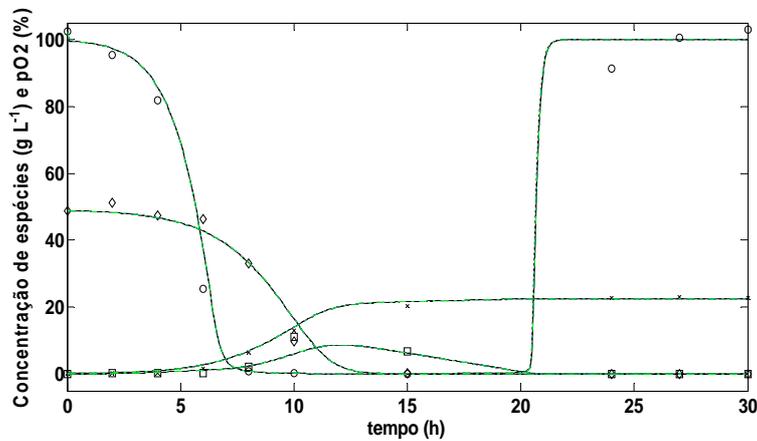


Figura 3.12 - Dados experimentais. Concentrações de: biomassa (×), substrato (◇), etanol (□), oxigênio dissolvido (○); e calculados – parâmetros originais (---) e após a estimação dos parâmetros selecionados pela estratégia de YAO *et al.* (2003) (.....).

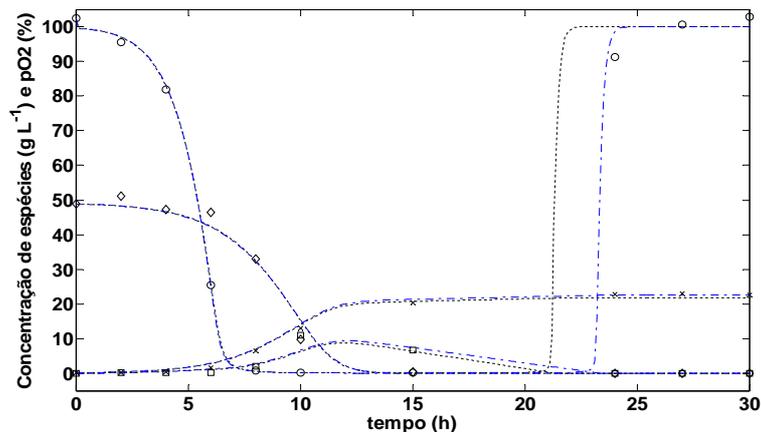


Figura 3.13 - Dados experimentais. Concentrações de: biomassa (×), substrato (◇), etanol (□), oxigênio dissolvido (○); e calculados – parâmetros originais (---) e após a estimação dos parâmetros selecionados pela estratégia de SECCHI *et al.* (2006) (-.-.).

Para este estudo de caso, os ângulos de dependência linear entre os parâmetros foram calculados e os resultados podem ser visualizados na Tabela 3.18.

Tabela 3.18 – Ângulos (em graus) de dependência linear entre os parâmetros do modelo de LONGHI et al. (2004).

	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}
θ_1	<u>10,9</u>	59,3	<u>174</u>	119	122	129	125	<u>44,7</u>	<u>30,9</u>	118	61,5
θ_2		48,6	<u>164</u>	129	132	<u>139</u>	<u>136</u>	<u>34,2</u>	<u>20,1</u>	129	50,9
θ_3			115	<u>173</u>	<u>179</u>	<u>170</u>	<u>173</u>	<u>16</u>	<u>30,1</u>	<u>175</u>	<u>5,16</u>
θ_4				67	64,1	57	60,6	130	<u>144</u>	67,3	113
θ_5					<u>6,8</u>	<u>10,1</u>	<u>6,41</u>	<u>163</u>	<u>147</u>	<u>1,31</u>	<u>178</u>
θ_6						<u>9,16</u>	<u>6,7</u>	<u>165</u>	<u>151</u>	<u>5,83</u>	<u>174</u>
θ_7							<u>3,67</u>	<u>173</u>	<u>157</u>	<u>10,4</u>	<u>170</u>
θ_8								<u>170</u>	<u>154</u>	<u>6,74</u>	<u>173</u>
θ_9									<u>17,8</u>	<u>163</u>	<u>17</u>
θ_{10}										<u>147</u>	<u>32,7</u>
θ_{11}											<u>180</u>

Como apresentado na Tabela 3.18, há indícios que os parâmetros do modelo de LONGHI *et al.* (2004) estejam fortemente correlacionados. Ainda assim as estratégias de identificabilidade selecionaram os parâmetros. Por exemplo, as estratégias de SALTELLI (2002) e de SECCHI *et al.* (2006) indicam que é possível estimar simultaneamente os parâmetros 11 e 12, enquanto o ângulo entre os vetores de sensibilidades destes parâmetros é 180°. Isto sugere que a matriz de incertezas experimentais esteja influenciando o resultado obtido, permitindo a inversão da matriz de informação de Fisher mesmo quando há dependência linear dos parâmetros.

3.3 CONCLUSÕES PARCIAIS

De um modo geral, conclui-se que a dificuldade de estimação quando muitos parâmetros estão presentes, pode ser superada empregando procedimentos de identificabilidade de parâmetros que utilizam critérios de seleção baseados em análise de sensibilidade.

Para os estudos de caso apresentados neste capítulo, os procedimentos empregados, SALTELLI (2002), YAO et al. (2003) e SECCHI et al. (2006) apresentaram ordenamento diferentes, mesmo para igual resultado de seleção. O diferente ordenamento dos parâmetros por estes procedimentos pode ter duas causas principais: a característica particular do método de abordar o problema e/ou os parâmetros podem apresentar influências muito semelhantes.

Um ponto importante desta investigação preliminar foi a verificação que o uso da análise de sensibilidade como ferramenta de identificabilidade de parâmetros pode resultar em um problema de estimação de parâmetros mal condicionado. Especialmente no caso da análise de sensibilidade global SALTELLI (2002), em que esta ferramenta baseia-se apenas análise da influência dos parâmetros sobre a predição, a depender das características do modelo e do plano experimental empregado, os parâmetros estatisticamente mais significativos podem não ser os parâmetros mais estimáveis do modelo. A estratégia de SALTELLI (2002) consiste em um procedimento para análise de sensibilidade e não leva em conta informações sobre o condicionamento do procedimento de estimação dos parâmetros selecionados, podendo levar à estimação de parâmetros ruidosos sem significado estatístico e físico ou a um problema de estimação mal condicionado. Considerando este tipo de problema, nota-se a que os desempenhos das estratégias de YAO *et al.* (2003) e de SECCHI *et al.* (2006) demonstram ser superiores para os casos em que há correlação paramétrica; YAO *et al.* (2003) e SECCHI *et al.* (2006) avaliam o condicionamento do problema de estimação a cada parâmetro selecionado, de forma que quando o problema torna-se mal condicionado estes procedimentos são finalizados, teoricamente selecionando apenas os parâmetros para os quais a estimação é numericamente possível.

Outro ponto importante, é que os procedimentos de identificabilidade de parâmetros cujos critérios de seleção são baseados em análises de sensibilidade, mostram-se extremamente sensíveis à qualidade de informação sobre os parâmetros. Dentre os três procedimentos, YAO *et al.* (2003) e SECCHI *et al.* (2006) se mostram fortemente dependentes das estimativas iniciais dos parâmetros, visto que tais procedimentos são baseados na matriz de sensibilidade local B , que emprega diretamente os valores fornecidos às estimativas iniciais dos parâmetros. Apesar do seu caráter global, este problema também pode ocorrer para análise de sensibilidade de SALTELLI (2002). Frequentemente, boas estimativas iniciais para os valores de parâmetros não estão disponíveis, sendo tomadas de sistemas similares reportados na

literatura ou mais comumente adotadas de forma arbitrária (KOU et al., 2005a; KOU et al., 2005b). Considerando a natureza não-linear e dinâmica dos sistemas das engenharias química e ambiental, esta é uma grande lacuna a ser tratada neste trabalho.

Fica claro que dada à especificidade dos sistemas, que podem apresentar características muito peculiares, o sucesso do desempenho das estratégias de identificabilidade está vinculado à adequabilidade ao problema tratado do critério de seleção de parâmetros empregado. Desta forma, a escolha adequada do procedimento de identificabilidade deve considerar as características do problema, tais como a qualidade da estimativa inicial dos parâmetros, a quantidade e qualidade dos dados experimentais e a estrutura do modelo.

Assim, este estudo preliminar do problema de identificabilidade de parâmetros demonstrou um amplo panorama para o estudo de procedimentos de identificabilidade de parâmetros, que possuem pronta aplicação aos problemas das engenharias química e ambiental. Um importante resultado obtido é a constatação que o estudo desenvolvido aponta um amplo panorama para o desenvolvimento de novos procedimentos, um dos grandes objetivos da tese.

CAPÍTULO 4. ASPECTOS CONCEITUAIS DO PROBLEMA DE IDENTIFICABILIDADE DE PARÂMETROS

A seleção de parâmetros é o foco principal da identificabilidade e não pode ser realizada adequadamente sem uma definição rigorosa a respeito da classificação do parâmetro. Mesmo porque, de acordo com os procedimentos de identificabilidade de parâmetros propostos na literatura, os parâmetros que não podem ser estimados são fixados na estimativa inicial. Naturalmente, estes valores afetam o resultado da estimação dos parâmetros identificáveis, bem como suas incertezas e a qualidade da predição. Se tais valores estão próximos dos valores "verdadeiros" (e desconhecidos) dos parâmetros, as incertezas associadas à estimação e a predição serão pequenas. Todavia, frequentemente as informações a respeito dos valores de parâmetros e suas características são desconhecidas, de modo que a caracterização das incertezas dos parâmetros e da predição após o uso dos procedimentos de identificabilidade pode ser enganosa. Esta questão vem sendo pouco debatida na literatura técnico-científica e será abordada neste Capítulo 4 da tese.

4.1 ASPECTOS CONCEITUAIS DA CLASSIFICAÇÃO DOS PARÂMETROS

Como mencionado no Capítulo 2 (Seção 2.2), a literatura técnico-científica reporta uma variedade de tipos de classificações de parâmetros quanto à identificabilidade. Todavia, os conceitos utilizados sugerem ambiguidades, de modo que as definições dadas por cada classificação podem gerar conflitos.

A Figura 4.1 apresenta a classificação clássica dos parâmetros em termos da solução do problema otimização.

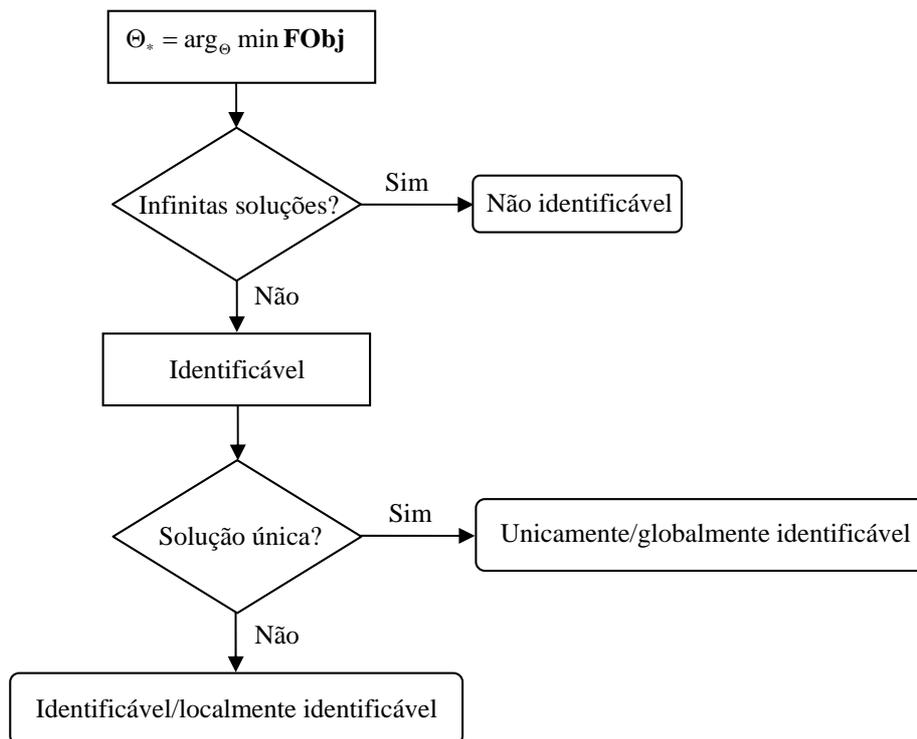


Figura 4.1 - Classificação clássica dos parâmetros em termos da solução do problema de otimização.

Neste sentido, a classificação do parâmetro quanto a sua identificabilidade pode ser dada com relação a sua influência sobre a função objetivo. É importante observar que a classificação de parâmetros como unicamente identificável está associada à presença de um único ponto de mínimo, o que resulta em um valor único do parâmetro e da função objetivo, como apresentado pela Figura 4.2A. Porém, é sabido que devido às incertezas experimentais os valores dos parâmetros não podem ser determinados com exatidão, sendo estes valores definidos em termos de um intervalo ou de uma região de confiança que podem compreender um número considerável (infinito, no caso de variáveis contínuas) de valores aceitáveis para o parâmetro.

Outra questão relevante é a classificação do parâmetro como identificável, que considera a presença de múltiplos, porém limitados, mínimos e impõe que estes mínimos resultem em funções objetivo de mesmo valor, como apresentado pela Figura 4.2C. Na prática esta condição imposta para fins de classificação raramente ocorre, sendo mais comum que múltiplos mínimos resultem em diferentes valores da função objetivo, como apresentado na Figura 4.2B.

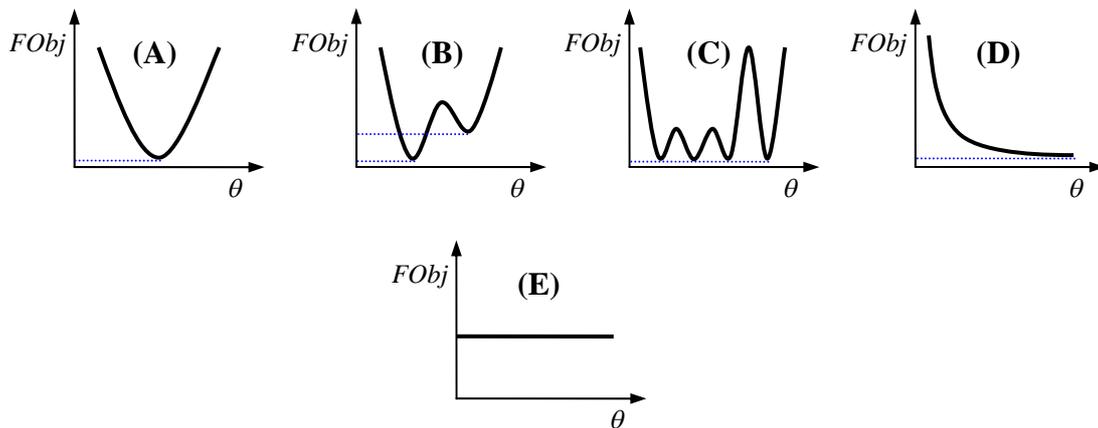


Figura 4.2 - Algumas possíveis formas de função objetivo para um parâmetro.

A Figura 4.2 apresenta algumas possíveis formas de função objetivo para um parâmetro. O caso A ilustra o cenário para o qual a função objetivo é convexa contendo um único mínimo, o que leva a obtenção de um único valor do parâmetro (e a sua região de confiança) garantindo a unicidade do valor do parâmetro. No caso B, têm-se o cenário no qual a função objetivo apresenta mais de um ponto de mínimo, local e global, sendo possível obter diferentes valores para o parâmetro (e as suas respectivas regiões de confiança); mas o valor do parâmetro (e a sua região de confiança) obtido a partir do menor valor da função objetivo, o mínimo global, seria o valor escolhido para representar o sistema, apesar dos outros valores estarem de acordo com os limites estatísticos estabelecidos para a região de confiança. Semelhante ao caso B, o caso C ilustra o cenário no qual se identificam múltiplos mínimos globais, nos quais a função objetivo apresenta o mesmo valor sendo possível obter diferentes valores para o parâmetro (e as suas respectivas regiões de confiança), sendo o parâmetro considerado não identificável. Para o caso D, infinitos valores para o parâmetro são possíveis (e as respectivas regiões de confiança), sendo o parâmetro não identificável. Por último, o caso E, ilustra o cenário em que a função objetivo não é influenciada pelo parâmetro.

É importante enfatizar que quando se trata de um parâmetro, o intervalo e a região de confiança são coincidentes; contudo, na presença de mais de um parâmetro a representação correta da incerteza paramétrica é dada pela região de confiança. Neste contexto, parece bastante conveniente que as definições utilizadas para as classificações quanto à identificabilidade sejam empregadas em termos da região de confiança dos parâmetros. Esta proposta vem da consideração que não necessariamente os múltiplos mínimos vão resultar em um mesmo valor da função objetivo, ao contrário do que é dito pela definição de parâmetros identificáveis reportada pela literatura.

A Figura 4.3 ilustra funções objetivo com mais de um ponto de mínimo para um parâmetro.

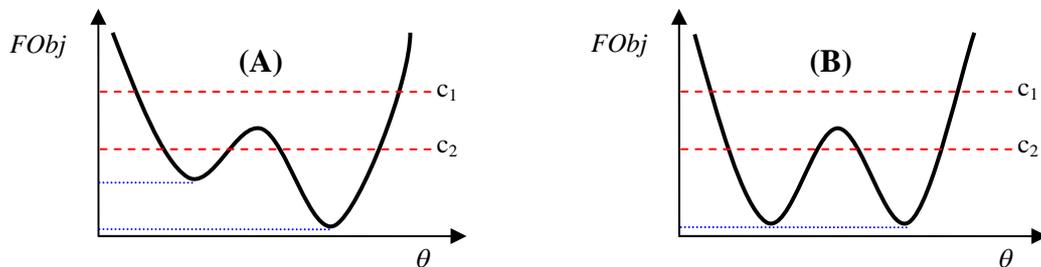


Figura 4.3 - Formas de função objetivo para um parâmetro apresentando múltiplos mínimos.

No caso A da Figura 4.3, a função objetivo apresenta valores diferentes para os mínimos, enquanto no caso B estes valores são iguais; a função objetivo está representada pela linha cheia, o mínimo está representado pela linha pontilhada e os limites de corte definidos por c_1 e c_2 estão definidos em linha tracejada. Em ambos os casos, os valores obtidos para o parâmetro e as suas respectivas regiões de confiança são diferentes. A região de confiança é determinada com base no limite de corte de valores aceitáveis da função objetivo, de modo que a depender do limite estabelecido (c_1 ou c_2) obtêm-se diferentes valores das propriedades desta região (e.g., o volume) e pode se chegar a diferentes interpretações. Nesta Figura 4.3, considerando o valor de limite de corte c_2 , a região de confiança em ambos os casos apresentará forma desconexa.

Outra consideração importante sobre a proposta de definições para a classificação dos parâmetros quanto à identificabilidade em termos da região de confiança é baseada nos valores obtidos para a função objetivo com base na tolerância admitida. Não raramente, a observação de um platô pode significar que nesta região a função objetivo varia em valores inferiores aos valores admitidos para a tolerância da

função objetivo. Tal fato pode levar a questionamentos sobre a classificação dos parâmetros como não identificáveis.

A Figura 4.4 ilustra casos em que a função objetivo é mantida a um valor fixo devido à tolerância admitida.

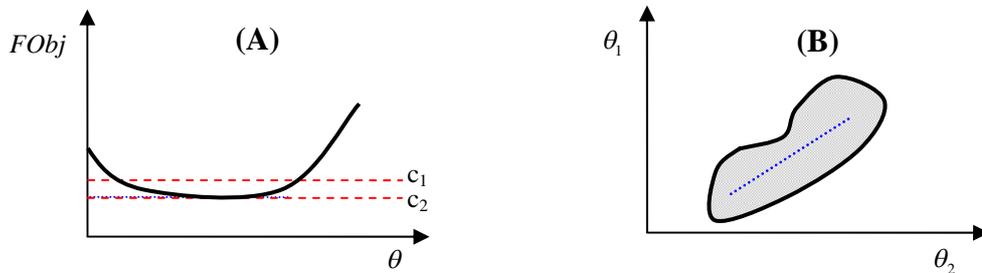


Figura 4.4 – Ilustração de infinitos mínimos obtidos por imprecisão numérica ou característica do problema estudado.

Na Figura 4.4, para o caso A, a função objetivo está representada pela linha cheia, o mínimo está representado pela linha pontilhada e os limites de corte definidos por c_1 e c_2 estão definidos em linha tracejada; para o caso B a região de confiança representada em cinza com limites definidos por linha cheia apresenta mínimos representados em linha pontilhada. Nesta figura, é possível observar para ambos os casos que a função objetivo mantém um valor fixo durante o procedimento de estimação de parâmetros. No caso A, a função objetivo apresenta variação inferior à tolerância admitida formando um platô, de modo que vários valores de parâmetros são possíveis para o mesmo valor da função objetivo. Ainda neste caso, nota-se que ao utilizar o limite de corte c_2 a região de confiança apresentará a forma aberta (ou seja, sem limites definidos) para o parâmetro. O caso B poderia caracterizar uma representação do caso A, admitindo como limite de corte c_1 . Neste caso B, embora a região de confiança seja delimitada, os valores da função objetivo (demonstrados pela linha pontilhada) mantêm-se fixos enquanto os valores dos parâmetros variam. Desta forma, devido à própria natureza do procedimento de estimação de parâmetros, podem ocorrer muitos valores de parâmetros (e suas respectivas regiões de confiança) possíveis levando o parâmetro à não identificabilidade.

A classificação do parâmetro como não identificável será inevitável no caso de parâmetros linearmente dependentes, seja esta dependência ocasionada pela estrutura do modelo e/ou pelo planejamento dos experimentos. Parâmetros linearmente dependentes

apresentam a forma da região de confiança aberta, de modo que estes parâmetros (e suas respectivas regiões de confiança) podem assumir um número infinito de valores.

Nem sempre é possível identificar e/ou tratar o problema da dependência linear entre os parâmetros, sobretudo nos casos em que esta dependência é ocasionada pela estrutura do modelo, a chamada identificabilidade estrutural. Em se tratando de modelos de estruturas simples (e.g., os modelos de Monod/Michaelis-Menten), a dependência pode ser identificada e os modelos podem ser reescritos em uma estrutura que os parâmetros apresentem menor dependência, utilizando técnicas de re-parametrização. Todavia, em se tratando de modelos complexos, dificilmente estas dependências podem ser identificadas (e.g., modelos de polimerização (SCHWAAB, 2005; TOJAL et al., 2013)) e tratadas (e.g., o modelo termodinâmico UNIQUAC (COSTA, 2011)).

Exemplo ilustrativo 4.1

Parâmetros linearmente dependentes devido à estrutura do modelo

Considere o seguinte modelo fictício:

$$y = \frac{\theta_1}{\theta_2 x + \theta_3}$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^1$ e $\theta \in \mathfrak{R}^3$ representam, respectivamente, as variáveis de saída e de entrada e os parâmetros.

Fazendo a seguinte manipulação matemática

$$y = \frac{\theta_1/\theta_1}{\theta_2/\theta_1 x + \theta_3/\theta_1},$$

o modelo fictício pode ser reescrito na forma estrutural

$$y = \frac{1}{\theta_2'' x + \theta_3''}.$$

Uma segunda forma estrutural pode ser obtida fazendo

$$y = \frac{\theta_1/\theta_3}{\theta_2/\theta_3 x + \theta_3/\theta_3},$$

$$y = \frac{\theta_1''}{\theta_2'' x + 1}.$$

Por último, fazendo

$$y = \frac{\theta_1/\theta_2}{\theta_2/\theta_2 x + \theta_3/\theta_2},$$

obtem-se uma terceira forma estrutural para o modelo original

$$y = \frac{\theta_1}{x + \theta_3}.$$

Para este exemplo, é possível observar que a forma estrutural do modelo original não está adequada sob o ponto de vista de estimação de parâmetros, uma vez que o modelo pode ser reescrito utilizando apenas dois parâmetros; ou seja, há uma relação funcional entre os parâmetros θ_1 , θ_2 e θ_3 .

Exemplo ilustrativo 4.2

Parâmetros linearmente dependentes devido aos efeitos numéricos

Considere o seguinte modelo fictício:

$$y = \frac{\theta_1}{\theta_2 x + 1}$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^1$ e $\theta \in \mathfrak{R}^2$ representam, respectivamente, as variáveis de saída e de entrada e os parâmetros.

Seja o valor do parâmetro θ_2 muito alto, o modelo fictício pode ser reescrito como

$$y = \frac{\theta_1}{\theta_2 x}.$$

Ou então, seja o valor do parâmetro θ_2 muito baixo, o modelo fictício pode ser simplificado para

$$y = \theta_1.$$

Neste exemplo, nota-se que a forma estrutural adequada do modelo para fins de estimação de parâmetro está condicionada à magnitude relativa dos parâmetros e dos dados experimentais disponíveis.

Exemplo ilustrativo 4.3

Parâmetros linearmente dependentes devido ao planejamento de experimentos

Considere o seguinte modelo fictício:

$$y = \theta_1 x_1 + \theta_2 x_2$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^2$ e $\theta \in \mathfrak{R}^3$ representam, respectivamente, a variável dependente, as variáveis independentes e os parâmetros.

Sejam as observações experimentais fixadas $x_1 = x_2$ empregando o planejamento. Ao aplicar o modelo fictício ao conjunto de experimentos, observa-se que

$$y = (\theta_1 + \theta_2) x_1,$$

$$y = \theta_1^* x_1.$$

Este exemplo demonstra que ainda que a estrutura do modelo esteja descrita adequadamente pode não ser possível estimar θ_1 e θ_2 independentemente, uma vez que para os experimentos realizados $x_1 = x_2$ o modelo pode ser reescrito em termos de um único parâmetro.

O diagnóstico adequado quanto aos problemas de estimação de parâmetros que apresentam forte correlação paramétrica devido à estrutura do modelo ou ao planejamento de experimentos, dificilmente pode ser aplicado a modelos complexos. Com isto, as definições quanto à identificabilidade de parâmetros são pouco generalizáveis, sobretudo porque há uma linha tênue entre o que é parâmetro linearmente dependente e o que é baixa estimabilidade do parâmetro.

Com base nestas informações, e admitindo-se que na prática os valores dos parâmetros são definidos com base em limites aceitáveis para a função objetivo, propõe-se representar os conceitos de identificabilidade baseando-se na região de confiança dos parâmetros. A classificação dos parâmetros quanto à identificabilidade em termos da região de confiança, está baseada nas formas destas regiões que de modo geral podem ser abertas (regiões não delimitadas), desconexas e conexas. Desta forma, os parâmetros podem ser classificados como a seguir:

- (i). Globalmente identificáveis. Apresenta região de confiança conexa, o que caracteriza um mínimo global.
- (ii). Localmente identificáveis. Apresenta região de confiança desconexa, indicando a presença de mínimos locais.
- (iii). Não-identificáveis. Apresenta região de confiança aberta, indicando um número infinito de mínimos possíveis.

A Figura 4.5 ilustra algumas formas possíveis de região de confiança dos parâmetros, conforme a classificação proposta.

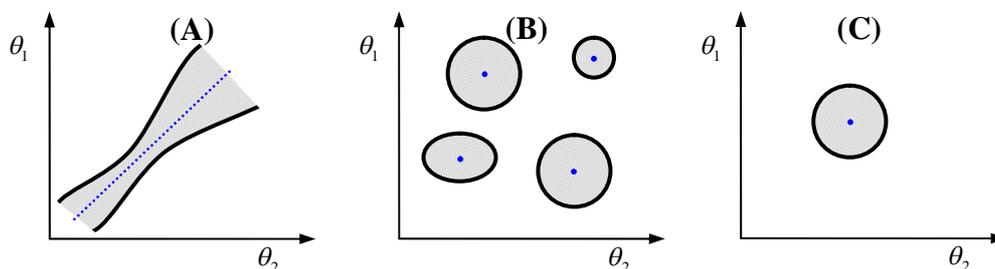


Figura 4.5 - Algumas possíveis formas de região de confiança dos parâmetros: (A) aberta contendo múltiplos mínimos, (B) desconexa contendo mínimos locais e (C) conexa contendo mínimo global.

Na Figura 4.5 estão representadas regiões de confiança abertas (gráfico A), desconexas (gráfico B) e conexas (gráfico C), respectivamente; a forma é dada pela

região cinza com limites definidos por linhas cheias e os mínimos são representados por pontos ou linhas pontilhadas. No caso A, a região de confiança representada não possui limites estabelecidos, logo um número infinito de valores pode ser assumido pelos parâmetros. Este caso pode indicar que os parâmetros são linearmente dependentes, sendo não-identificáveis. Para o caso B a região de confiança apresenta múltiplos mínimos, o que demonstra a possibilidade de se obter diferentes valores para o mesmo parâmetro, sendo que o valor do parâmetro e a sua região de confiança, obtidos a partir do menor valor da função objetivo, dados pelo mínimo global, são os mais adequados para representar o sistema. Este caso enquadra usualmente os modelos não-lineares, muito comuns aos processos das engenharias química e ambiental. No caso C a região de confiança única e delimitada indica a presença de um único mínimo da função objetivo, garantindo a unicidade do parâmetro. Neste caso, configura-se um exemplo particular que são os modelos lineares associados a planejamentos adequados.

Há ainda outros casos possíveis. Citada anteriormente (Figura 4.4B), a região de confiança única delimitada que apresenta os valores da função objetivo fixos enquanto os valores dos parâmetros variam, ocorre raramente. Esta forma de região de confiança pode ocorrer artificialmente devido à tolerância definida para a função objetivo, bem como quando os parâmetros são delimitados por restrições (caracterizando o caso da região de confiança aberta, mas delimitada pelas restrições). Tal fato é característico da não identificabilidade, já que são inúmeros valores que podem ser admitido pelos parâmetros (e suas respectivas regiões de confianças). Também incomuns, podem ocorrer regiões de confianças conexas delimitadas que apresentam múltiplos, porém finitos mínimos. Comumente associada a modelos não-lineares, esta forma de região pode ser ocasionada pelo valor definido para o limite de corte da função objetivo. Uma vez que este número de mínimos é finito, considera-se que os parâmetros são identificáveis.

Na prática, a classificação de parâmetros quanto à identificabilidade é um procedimento bastante complexo. Desta forma, é natural a multiplicidade de termos e jargões que compõem o vocabulário da identificabilidade, bem como suas contradições, seus equívocos e suas ambiguidades. Nota-se que a identificabilidade de parâmetros apresenta aspectos específicos envolvendo características do modelo e do planejamento de experimentos, de modo que qualquer tentativa formal de descrição do problema pode levar à perda de generalização. Porém, o diagnóstico e o tratamento do problema de

identificabilidade requerem conceitos generalizáveis para o estabelecimento de metodologias.

A introdução de conceitos de região de confiança é uma tentativa de tornar factível a classificação de parâmetros quanto à identificabilidade e, como as demais abordagens, também apresentam limitações. Estes conceitos podem ser muito úteis pela facilidade de obtenção e clareza da informação, mas são reféns do valor admitido para a região de busca dos parâmetros e o limite de corte da função objetivo. Com base nas variáveis de decisão e nas restrições, a região de busca pode ser construída por um método de otimização, que se não for adequado ao problema pode fornecer valores inconsistentes. O limite de corte é definido geralmente por um teste estatístico (e.g., o teste Chi-quadrado quando se admite que os parâmetros apresentam distribuição de probabilidade normal) que, de um modo geral, o torna uma informação confiável.

Uma vez que as definições utilizadas para a classificação de parâmetros quanto à identificabilidade neste estudo estão baseadas nas formas da região de confiança, a generalização destes conceitos é apresentada pela Figura 4.6.

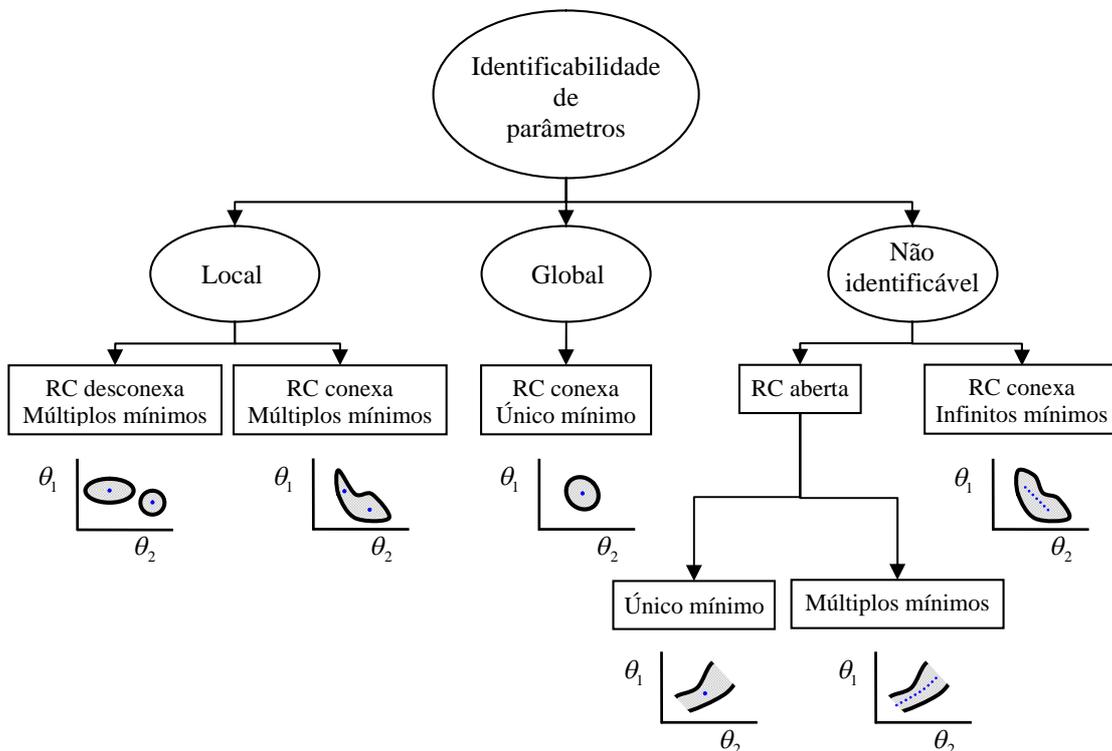


Figura 4.6 - Generalização do procedimento de seleção de parâmetros.

De acordo com a Figura 4.6 e classificados utilizando a nomenclatura convencional da literatura, os parâmetros são definidos em três categorias principais

quanto à forma da região de confiança (RC): (i) localmente identificáveis, (ii) globalmente identificáveis e (iii) não identificáveis. Para fins de visualização, são utilizados gráficos onde as formas das regiões de confiança são representadas em cinza, cujos limites estão demarcados por linha cheia, sendo a representação de mínimo indicada pelos pontos ou pela linha pontilhada.

O parâmetro será dito identificável ou localmente identificável, quando for verificada a presença de múltiplos, mas finitos mínimos que pode ocorrer tanto em regiões desconexas quanto em regiões únicas; considerando uma função objetivo não-linear, estas regiões de confiança podem ser encontradas a depender do limite de corte determinado para a função objetivo. Naturalmente, na presença de um único mínimo e uma única região de confiança o parâmetro será dito unicamente ou globalmente identificável. Por fim, quando um número infinito de mínimos, seja em região de confiança aberta ou fechada, os parâmetros serão ditos não identificáveis; as formas destas regiões podem depender da tolerância admitida para a função objetivo e/ou do limite de corte.

É importante ressaltar que, na prática, a distinção entre parâmetro globalmente identificável ou parâmetro localmente identificável é muito pouco utilizada, sendo mais recomendado o uso apenas do termo identificável. Isto porque, após o procedimento de identificabilidade ambas estas classes recebem o mesmo tratamento, a estimação. Assim, como a forma corrente adotada na literatura atual, a distinção necessária e suficiente realizada pelo procedimento de identificabilidade sobre os parâmetros é dada por duas classes apenas: (i) identificável e (ii) não-identificável.

A abordagem acerca de classificação dos parâmetros quanto à identificabilidade, utilizando o conceito da região de confiança dos parâmetros, também é apresentada similarmente por RAUE et al. (2009), em estudos independentes a esta tese. De acordo com RAUE et al. (2009), o conceito de região de confiança dos parâmetros pode ser empregado para definir o tipo de não identificabilidade que afeta a estimação do parâmetro θ_i , embora os parâmetros que apresentam a não identificabilidade estrutural também possam apresentar a não identificabilidade prática. Desta forma, o parâmetro θ_i pode apresentar (RAUE et al., 2009): (i) identificabilidade, se o seu intervalo de confiança é finito $[\sigma_{\theta_i}^-, \sigma_{\theta_i}^+]$, (ii) não identificabilidade estrutural, se o seu intervalo de confiança é infinito $[-\infty, +\infty]$, ou (iii) não identificabilidade prática, se um dos limites do seu intervalo de confiança é infinito $[\sigma_{\theta_i}^-, \sigma_{\theta_i}^+]$, $\sigma_{\theta_i}^- = -\infty$ ou $\sigma_{\theta_i}^+ = +\infty$; como mostra

a Figura 4.7. Assim como a classificação de parâmetros quanto ao tipo global ou local, o uso dos termos identificabilidade estrutural e identificabilidade prática é pouco usual na literatura.

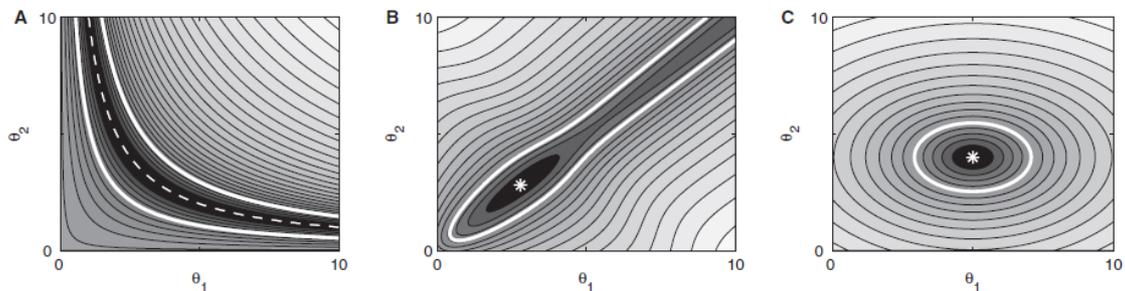


Figura 4.7 - Classificação proposta por RAUE et al. (2009), quanto ao tipo de identificabilidade de parâmetros utilizando a região de confiança dos parâmetros: A - não identificabilidade estrutural, B - não identificabilidade prática, C - identificável.

4.2 ASPECTOS CONCEITUAIS DA CARACTERIZAÇÃO ESTATÍSTICA DE INCERTEZAS

Usualmente, a estimação de parâmetros é realizada após o procedimento de identificabilidade, como uma etapa final ou exclusiva deste procedimento (Figura 2.6). Assim, a qualidade dos valores das estimativas iniciais admitidas para os parâmetros é fundamental para a seleção adequada. Infelizmente, boas estimativas iniciais dos valores de parâmetros são raramente conhecidas; em alguns casos as estimativas podem ser obtidas a partir de sistemas similares reportados na literatura ou, mais comumente assumidas arbitrariamente (KOU et al., 2005a; KOU et al., 2005b).

A qualidade das estimativas iniciais dos parâmetros é mandatória durante e após o procedimento de identificabilidade, visto que o subconjunto de parâmetros selecionado deverá ser estimado. A estimação dos parâmetros selecionados pode não ser garantida, especialmente durante a etapa de otimização, em que os parâmetros podem exceder os limites pré-estabelecidos, e também durante a avaliação do modelo, em que algumas combinações de valores de parâmetros podem levar a soluções inadequadas. Embora tais problemas sejam comuns em procedimentos de identificabilidade de parâmetros, têm sido negligenciados na literatura. Portanto, quando baseado em valores incertos, os procedimentos de identificabilidade de parâmetros podem levar à seleção inadequada dos parâmetros e, por conseguinte, problemas de estimação de parâmetros mal condicionados.

Naturalmente, a incerteza das estimativas iniciais propaga-se para os valores dos parâmetros estimados e suas incertezas e, conseqüentemente, para a incerteza de predição do modelo. Uma das dificuldades ocasionadas pelos procedimentos de identificabilidade é que ao fixar os valores dos parâmetros não identificáveis, geralmente correlacionados com os parâmetros identificáveis, restringem-se os valores que podem ser obtidos pela estimação. Como agravante, a estimação pode obter valores espúrios para os parâmetros identificáveis se os parâmetros não identificáveis são fixados em valores inapropriados; neste caso, é comum obter valores de parâmetros que violam os limites de validade e/ou que levam a erros numéricos durante a resolução do modelo. Outra dificuldade que também deriva da estratégia de fixar os valores dos parâmetros não identificáveis, é a inviabilidade de contabilizar a incerteza atribuída aos valores destes parâmetros, uma vez que apenas os parâmetros estimados podem ter suas incertezas calculadas pelos procedimentos convencionais de estimação. Como consequência, a incerteza de predição contabiliza apenas a incerteza dos parâmetros estimados e a correlação existente entre estes parâmetros. Um dos agravantes de se usar estimativas iniciais inadequadas está no fato de que os parâmetros não identificáveis apresentarão uma incerteza elevada que não pode ser conhecida e, portanto, não pode ser contabilizada pela incerteza de predição, mas que estará presente podendo levar à resultados desastrosos quando do uso do modelo matemático em pontos experimentais diferentes dos utilizados na estimação dos parâmetros identificáveis.

A existência de correlação entre os parâmetros não identificáveis e os parâmetros identificáveis, condição pela qual esses são assim classificados na maioria dos casos, pode invalidar todo o procedimento de estimação e de caracterização de incertezas, especialmente no caso de baixa qualidade das estimativas iniciais. Quando o parâmetro está correlacionado com outro parâmetro, seu valor não pode ser estimado independentemente, de forma que os parâmetros identificáveis terão seus valores estimados a fim de compensar o erro da estimativa nos parâmetros não identificáveis que foram fixados. Neste contexto, a matriz de covariância dos parâmetros pode perder a sua validade, visto que a incerteza do parâmetro obtida está associada à incerteza de outros parâmetros cuja incerteza não pode ser conhecida. Da mesma forma, não é possível conhecer a interação entre estes dois conjuntos de parâmetros, identificáveis e não identificáveis, após a estimação.

Ao fixar o valor de um parâmetro, assume-se uma hipótese forte no problema de estimação: o valor no qual o parâmetro foi fixado é o seu valor "verdadeiro" - não é

necessário a estimação e informações acerca da sua incerteza são desprezadas. Portanto, mesmo considerando-se um cenário em que as estimativas iniciais admitidas para os valores dos parâmetros são apropriadas, a incerteza dos parâmetros e a incerteza da predição podem ser apenas parcialmente conhecidas, invalidando os resultados das análises estatísticas obtidos em problemas de estimação de parâmetros que fazem uso dos procedimentos de identificabilidade.

Para contornar o problema de inconsistência das estimativas iniciais dos valores de parâmetros, uma etapa anterior ao procedimento de identificabilidade pode ser adicionada, em que métodos globais de otimização sejam empregados para obter estimativas mais adequadas aos valores exatos dos parâmetros. Todavia, a maioria dos casos de interesse trata-se de modelos dinâmicos complexos com grandes restrições aos valores de parâmetros, de forma que valores de limites dos parâmetros mal estabelecidos frequentemente acarretam problemas numéricos. Portanto, mesmo métodos de otimização que não utilizam derivada como o Enxame de Partículas (KENNEDY e EBERHART, 1995) ou o Algoritmo Genético (GOLDBERG, 1989), podem falhar em obter tais valores.

Assim, existem questões importantes sobre os procedimentos usuais de identificabilidade, quando o subconjunto de parâmetros não identificáveis tem seus valores fixados na estimativa inicial:

- (i). Qual a confiabilidade dos valores dos parâmetros e das suas incertezas obtidas após a estimação?
- (ii). Qual a confiabilidade da predição do modelo após a estimação, especialmente nas regiões que não contêm observações experimentais?
- (iii). Como julgar a qualidade do modelo, uma vez que as incertezas não podem ser completamente conhecidas?

Neste contexto, parece razoável questionar a validade dos procedimentos de identificabilidade reportados na literatura técnico-científica. Para determinar se a incerteza calculada usando procedimentos de identificabilidade é uma informação consistente, que permite julgar a qualidade dos resultados obtidos após a estimação dos parâmetros, é importante saber o quanto da incerteza calculada inclui a incerteza real. Note que a incerteza real representa a incerteza calculada quando todos os parâmetros são estimados. Em se tratando de problemas de estimação mal condicionados, que são o

principal foco dos procedimentos de identificabilidade, torna-se inviável conhecer a incerteza real.

Na literatura técnico-científica a qualidade das informações previamente admitidas aos parâmetros, bem como os prejuízos acarretados por sua incerteza sobre os procedimentos de identificabilidade tem sido pouco discutida. OMLIN e REICHERT (1999) e CHU e HAHN (2007) comentam este problema. A influência de informações prévias a respeito da qualidade dos parâmetros é observada em OMLIN e REICHERT (1999), comparando métodos clássicos e a inferência Bayesiana para estimação de valores de parâmetros, quando se tem parâmetros identificáveis e não identificáveis. CHU e HAHN (2007) demonstram que uma variação de 10% sobre o valor nominal dos parâmetros não identificáveis pode levar a mudanças significativas nos valores dos parâmetros identificáveis; em especial no caso de parâmetros correlacionados, ao fixar o valor de um parâmetro não identificável que esteja correlacionado com um ou mais parâmetros identificáveis, afeta fortemente o procedimento de identificabilidade.

Parece claro que a abordagem rigorosa do problema de identificabilidade deve contabilizar a incerteza de todos os parâmetros bem como a sua consequência sobre a incerteza da predição, mesmo frente ao mau condicionamento do procedimento de estimação, de forma que se conheça estatisticamente a confiabilidade dos resultados obtidos e assim o risco de adotar o modelo em estratégias importantes do processo.

4.2.1 Propagação da incerteza dos parâmetros não identificáveis para a predição

O objetivo do procedimento de identificabilidade é a seleção dos parâmetros mais importantes e menos correlacionados do modelo para a estimação. A incerteza dos parâmetros estimados não pode ser adequadamente conhecida, quando outros parâmetros incertos do modelo não fazem parte da estimação. Naturalmente, o quanto da incerteza paramétrica se propaga para predição também não pode ser adequadamente contabilizado. Todavia, o sucesso do procedimento de identificabilidade está vinculado à boa capacidade preditiva do modelo para as regiões de interesse.

Neste sentido, propõe-se que o objetivo do procedimento de identificabilidade deve consistir em identificar quais os parâmetros podem ser fixados em seus valores nominais sem prejuízo da predição. A ideia básica desta proposta está representada na Figura 4.8, que apresenta 4 casos em que se tem diferentes cenários para a incerteza dos parâmetros e a incerteza da predição, de um modelo linear contendo dois parâmetros θ_1

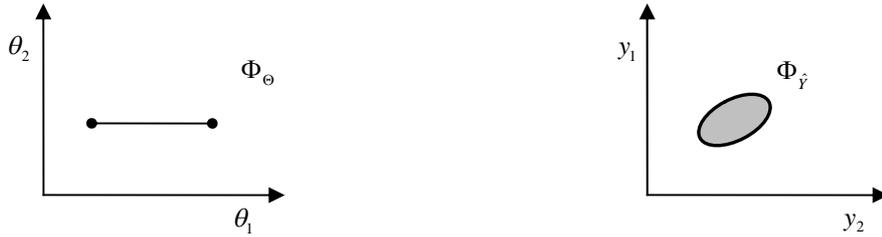
e θ_2 , em que Φ_θ e $\Phi_{\hat{y}}$ representam os volumes da região de incerteza dos parâmetros e da região de incerteza da predição, respectivamente. Nesta figura os casos 1 e 2 ilustram os cenários em que ambos os parâmetros podem ser estimados simultaneamente, i.e. θ_1 e θ_2 são identificáveis. O problema de identificabilidade aparece quando o volume da região de confiança dos parâmetros é muito grande ou infinito, como pode ser visto nos casos 3 e 4, cenários em que os parâmetros não podem ser simultaneamente identificáveis. Contudo, frente ao problema de identificabilidade, o volume da região de confiança da predição deve ser finito (caso 3) para assegurar a confiabilidade do modelo. Desta forma, o caso 3, representa os cenários em que os parâmetros não identificáveis podem ser fixados em valores arbitrários sem que haja prejuízo à predição, visto que tais parâmetros não são significativos para o comportamento do modelo. A princípio, isto deveria ser verificado para as condições experimentais de interesse, para que se tenha segurança ao fixar os valores dos parâmetros não identificáveis sem comprometer a confiabilidade da predição.

O cenário a princípio incorreto para a identificabilidade de parâmetros é aquele no qual a incerteza paramétrica se reflete em uma incerteza de predição muito elevada (caso 4). Desta forma, o caso 4, representa o cenário em que não há segurança ao fixar valores dos parâmetros não identificáveis. Os valores de y obtidos fora das regiões de experimentação investigadas, mas ainda em regiões de interesse, torna-se se tão arbitrários quanto a arbitrariedade do valor do parâmetro fixado. Neste cenário o procedimento de identificabilidade pode fornecer resultados enganosos.

Assim, a questão importante a ser respondida para a adequada aplicação do procedimento de identificabilidade de parâmetros é: será o volume da região de confiança da predição limitado mesmo quando o volume da região de incerteza paramétrica é extremamente elevado?

Caso 1

Cenário em que é conhecido o valor exato do parâmetro θ_2 e o valor do parâmetro θ_1 é conhecido em faixa associada a valores da predição.



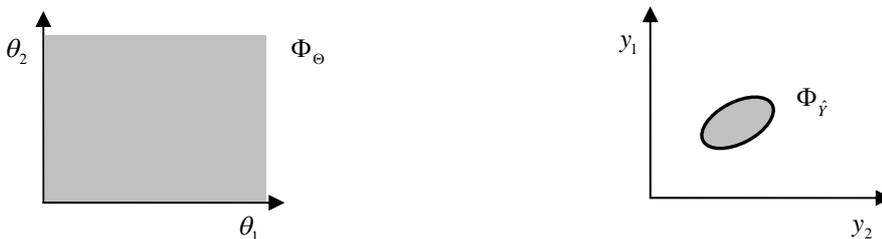
Caso 2

Cenário em que diferentes combinações de valores dos parâmetros θ_1 e θ_2 são consideradas possíveis e obtém-se uma larga região de incerteza da predição.



Caso 3

Cenário em que diferentes combinações de valores dos parâmetros θ_1 e θ_2 sejam possíveis e resultam em uma região de confiança não delimitada ou muito extensa e a região de incerteza da predição é delimitada.



Caso 4

Cenário em que as diferentes combinações de valores dos parâmetros θ_1 e θ_2 sejam possíveis e resultam em uma região de confiança não delimitada ou muito extensa e, semelhante à incerteza dos parâmetros, a região de incerteza da predição é não delimitada.



Figura 4.8 - Alguns cenários para a incerteza paramétrica e a incerteza preditiva de um modelo linear contendo dois parâmetros θ_1 e θ_2 ; Φ_θ e $\Phi_{\hat{y}}$ representam os volumes da região de incerteza dos parâmetros e da região de incerteza da predição, respectivamente.

4.2.2 Quando o uso do procedimento de identificabilidade leva a predição confiável

Pode-se argumentar que o uso do procedimento de identificabilidade é sempre apropriado; a questão é o que fazer com o resultado obtido. Na presença de parâmetros não identificáveis é necessário avaliar as incertezas decorrentes dos mesmos. Tais parâmetros são um indicativo de modelo mal estruturado ou experimentos insuficientes, resultado importante do procedimento.

Não obstante, indubitavelmente, um dos principais objetivos da identificabilidade de parâmetros é obter o modelo como instrumento de predição confiável. Logo, independente da qualidade dos parâmetros, o uso do procedimento de identificabilidade deve ser cuidadoso quando o valor desprezado de incerteza da predição torna-se muito elevado. Assim, será assumido que:

O uso da identificabilidade torna o modelo próprio para a predição quando não há perda de informação ao considerar apenas um subconjunto de parâmetros como estimável.

Este caso é caracterizado para a incerteza de predição é finita, mesmo sendo a incerteza dos parâmetros infinita. Para demonstrar em que situações a identificabilidade é própria para ajuste final do modelo, é conveniente destacar que a matriz *FIM* pode ser decomposta em valores e vetores característicos, na forma:

$$(E 4.1) \quad FIM = B^T (V_y)^{-1} B = D\Lambda D^T$$

em que *D* e Λ representam respectivamente os vetores característicos e os valores característicos da matriz *FIM*.

A matriz Λ é uma matriz diagonal. Para um procedimento bem posto, a matriz *FIM* é inversível e todos os valores característico λ_j são estritamente positivos. Para um procedimento mal posto, um ou mais elementos λ_j são próximos ou iguais a zero.

Para uma dada condição em que se deseja obter a predição com o modelo, a matriz de covariâncias das incertezas de predição \hat{V}_y pode ser escrita como:

$$(E 4.2) \quad \hat{V}_Y = B(FIM)^{-1} B^T.$$

Substituindo a Equação (E 4.2) na Equação (E 4.1), tem-se que:

$$(E 4.3) \quad \hat{V}_Y = B(D\Lambda D^T)^{-1} B^T.$$

Sendo a matriz *FIM* inversível, a matriz de covariâncias de predição pode ser escrita como:

$$(E 4.4) \quad \hat{V}_Y = B D \Lambda^{-1} D^T B^T,$$

ou seja, as direções características da matriz *FIM* ou à matriz V_{Θ} permanecem as mesmas. Uma vez que a incerteza dos parâmetros se propaga para a incerteza da predição do modelo, tais direções características da matriz *FIM* podem ser úteis em identificar os cenários em que a identificabilidade pode ser apropriadamente aplicada aos problemas de estimação mal condicionados.

A matriz *B* é constituída por vetores associados às sensibilidades de cada parâmetro, ao observar as colunas, conforme Figura 4.9, ou seja, $B = [b_{\theta_1} \ b_{\theta_2} \ b_{\theta_3} \ \dots \ b_{\theta_{NP}}]$. Contudo, para a demonstração a seguir, a matriz será representada observando-se os vetores formados pelas linhas e não pelas colunas, ou seja, $B^T = [b_{y_1} \ b_{y_2} \ b_{y_3} \ \dots \ b_{y_{NY}}]$.

A matriz de vetores característicos *D*, com dimensões $NP \times NP$, contém os vetores característicos, ou seja, $D = [d_1 \ d_2 \ d_3 \ \dots \ d_{NP}]$. O resultado da multiplicação de *BD* pode ser escrito como:

$$(E 4.5) \quad BD = \begin{bmatrix} b_{y_1}^T d_1 & b_{y_1}^T d_2 & \dots & b_{y_1}^T d_{NP} \\ b_{y_2}^T d_1 & b_{y_2}^T d_2 & \dots & b_{y_2}^T d_{NP} \\ \dots & \dots & \dots & \dots \\ b_{y_{NY}}^T d_1 & b_{y_{NY}}^T d_2 & \dots & b_{y_{NY}}^T d_{NP} \end{bmatrix}.$$

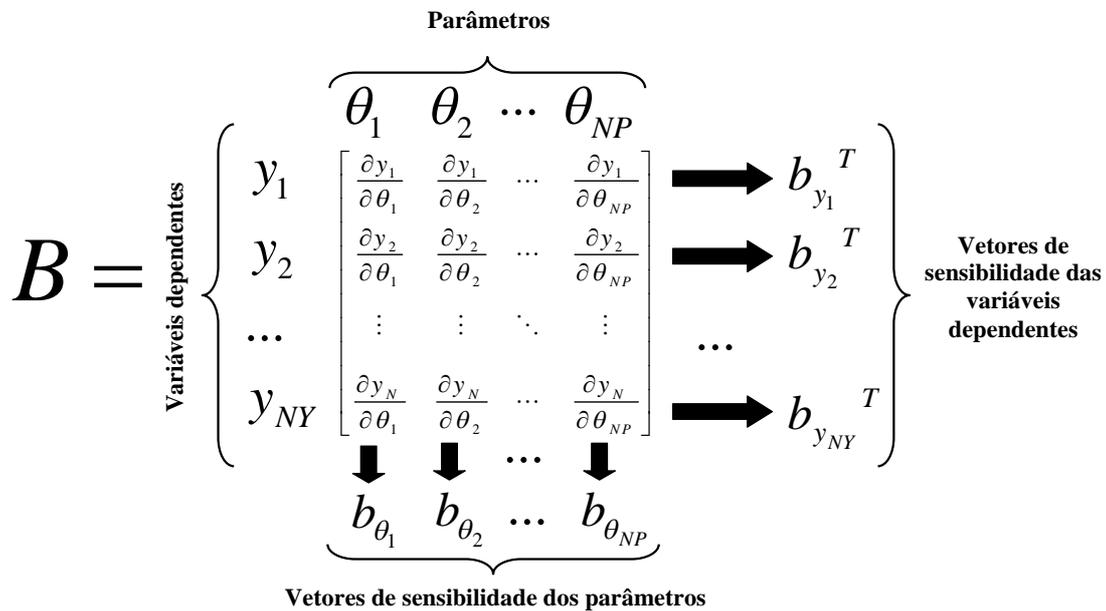


Figura 4.9 - Descrição da matriz de sensibilidade local.

Para ilustrar, pode-se concentrar no efeito da multiplicação da matriz B pela primeira direção característica d_1 : cada elemento resultante corresponde ao produto escalar entre as linhas da matriz B e d_1 . O primeiro elemento da multiplicação $B d_1$ corresponde ao produto escalar das sensibilidades da variável y_1 e do vetor d_1 ; o segundo elemento de $B d_1$ corresponde ao produto escalar das sensibilidades da variável y_2 e do vetor d_1 ; e assim por diante. Como o vetor d_1 possui tipicamente norma unitária, $\|d_1\|=1$, cada elemento da matriz $B d_1$, associado à cada y_i , corresponde essencialmente à projeção do vetor b_{y_i} sobre a direção d_1 , ou seja, $b_{y_i}^T d_1 = \|b_{y_i}\| \|d_1\| \cos(\beta_{b_{y_i}, d_1}) = \|b_{y_i}\| \cos(\beta_{b_{y_i}, d_1})$, em que $\beta_{b_{y_i}, d_1}$ corresponde ao ângulo entre o vetor b_{y_i} e o vetor da direção característica d_1 :

$$(E 4.6) \quad B d_1 = \begin{bmatrix} b_{y_1}^T d_1 \\ b_{y_2}^T d_1 \\ \dots \\ b_{y_{NY}}^T d_1 \end{bmatrix} = \|d_1\| \begin{bmatrix} \|b_{y_1}\| \cos(\beta_{b_{y_1}, d_1}) \\ \|b_{y_2}\| \cos(\beta_{b_{y_2}, d_1}) \\ \dots \\ \|b_{y_{NY}}\| \cos(\beta_{b_{y_{NY}}, d_1}) \end{bmatrix}.$$

Se, para dado y_i , a norma da sensibilidade dos parâmetros $\|b_{y_i}\|$ é não nula, então a projeção dos vetores de sensibilidade sobre a direção d_1 somente poderá ser nula se o vetor b_{y_i} e a direção d_1 forem perpendiculares entre si, ou seja, $\beta_{b_{y_i}, d_1} = \frac{\pi}{2}$. Logo, caso $\beta_{b_{y_i}, d_1} = \frac{\pi}{2}$, se os parâmetros forem modificados apenas na direção d_1 , não haverá sensibilidade sobre a predição da variável y_i . Se isto ocorrer para todas as variáveis y_i , os parâmetros podem ser quaisquer na direção d_1 , sem que a predição seja afetada, i.e:

$$(E 4.7) \quad \text{se } \beta_{b_{y_i}, d_1} = \frac{\pi}{2} \quad \forall (i = 1, \dots, NY) \quad \text{então } B d_1 = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

Isto implicará no fato de que a incerteza na direção d_1 pode ser infinita; contudo, isto não representa qualquer efeito sobre a incerteza de predição. Isto é conhecido como o espaço nulo da matriz, então, se d_1 é um espaço nulo da matriz B a incerteza nesta direção d_1 pode ser infinita. Associado a uma direção característica d_j , há um valor característico λ_j . No problema de otimização, este valor característico indica a informação contida na direção d_j . Então, quando a informação contida na direção é muito baixa $\lambda_j \rightarrow 0$, o inverso deste valor, $\frac{1}{\lambda_j} \rightarrow \infty$, resultará em uma indeterminação no cálculo da inversa da matriz FIM e da matriz \hat{V}_Y . Contudo, se a sensibilidade dos vetores da predição na direção

d_j for nula, $B d_j = 0$, os parâmetros podem variar de forma infinita na direção d_j sem que isto afete a predição. Isto quer dizer que \hat{V}_y pode ser definida, mesmo que a matriz *FIM* seja não inversível.

Exemplo ilustrativo 4.4

Não identificabilidade estrutural dos parâmetros

Considere o seguinte modelo fictício:

$$y = \theta_1 x_1$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^1$ e $\theta \in \mathfrak{R}^1$ representam, respectivamente, as variáveis de saída e de entrada e os parâmetros.

Caso sejam realizados três experimentos, $x_1 = [1 \ 2 \ 3]$, o parâmetro θ_1 é perfeitamente estimável e a região de confiança da predição é delimitada para qualquer condição experimental.

Supondo que, inadvertidamente, o modelo seja modificado, incluindo-se dois parâmetros cuja soma resulta no parâmetro θ_1 :

$$y = \theta_1'' x_1 + \theta_2'' x_1$$

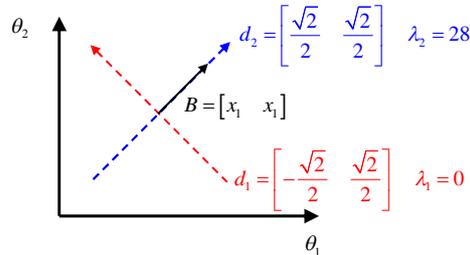
Esta nova estrutura do modelo apresenta uma dependência linear entre os parâmetros e, portanto, não se deve esperar que a estimação dos parâmetros θ_1'' e θ_2'' seja um problema bem posto. Isto quer dizer que, em uma direção característica da matriz de incerteza paramétrica V_θ , um valor característico será zero, levando a infinitos valores de θ_1'' e θ_2'' . No entanto, a informação da predição deve permanecer preservada, pois a informação contida nos dados experimentais acerca da predição do modelo não é perdida.

De fato, a matriz *FIM* e a sua decomposição em valores e vetores característicos levam à:

$$FIM = \begin{bmatrix} 14 & 14 \\ 14 & 14 \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}}_D \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 28 \end{bmatrix}}_\Lambda \underbrace{\begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}}_{D^T},$$

em que um dentre os valores característicos é nulo. Isto acarreta indeterminação na inversão da matriz *FIM* e, conseqüentemente, leva a uma região de confiança dos parâmetros infinita. Contudo, para qualquer condição de predição, como só há uma variável de saída, a matriz de sensibilidade é dada por $B = [x_1 \ x_1]$, que pode ser escrita como $B = x_1 [1 \ 1]$. Este vetor é perpendicular ao primeiro valor característico da matriz *FIM*,

$d_1 = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$, cuja direção possui valor característico nulo, i.e., o produto escalar da matriz de sensibilidade pelo primeiro valor característico resulta em $Bd_1 = [0]$:



Neste exemplo nota-se que o procedimento de fixar um dos parâmetros não deve afetar a incerteza da predição do modelo quando a não identificabilidade do parâmetro é oriunda da estrutura do modelo.

Exemplo ilustrativo 4.5

Não identificabilidade prática dos parâmetros

Considere o seguinte modelo fictício:

$$y = \theta_1 x_1 + \theta_2 x_2$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^2$ e $\theta \in \mathfrak{R}^2$ representam, respectivamente, as variáveis de saída e de entrada e os parâmetros.

Caso sejam realizados três experimentos, $x_1 = [1 \ 2 \ 3]$ e $x_2 = [1 \ 2 \ 3]$, em o planejamento resulte em condições iguais para as variáveis x_1 e x_2 , os parâmetros θ_1 e θ_2 não podem ser estimados independentemente; como mostra o exemplo ilustrativo 4.3.

De fato, a matriz *FIM* e a sua decomposição em valores e vetores característicos levam à:

$$FIM = \begin{bmatrix} 14 & 14 \\ 14 & 14 \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}}_D \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 28 \end{bmatrix}}_\Lambda \underbrace{\begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}}_{D^T},$$

Idêntico ao exemplo ilustrativo 4.1, um dos valores característicos é nulo. Isto acarreta indeterminação na inversão da matriz *FIM* e, conseqüentemente, uma região de confiança dos parâmetros infinita.

Entretanto, deve ser observado que para uma dada condição experimental, em que o planejamento seja adequado, $x_1 \neq x_2$, a matriz *B* não é necessariamente perpendicular ao vetor característico d_1 , visto que $B = [x_1 \ x_2]$; i.e., $Bd_1 \neq [0]$, em que as variações dos parâmetros na direção d_1 afetam a predição. Isto deve levar à indeterminação no cálculo da matriz \hat{V}_Y . Desta forma, a única condição na qual o cálculo de \hat{V}_Y

poderia ser efetuado é o caso em que $x_1 = x_2$, recaindo em um cenário similar ao descrito no exemplo ilustrativo 4.4.

Neste exemplo nota-se que o procedimento de fixar um dos parâmetros deve afetar a incerteza da predição do modelo quando a não identificabilidade do parâmetro é oriunda do plano experimental empregado.

Quando a direção característica d_j que contém infinitos valores possíveis de parâmetros não se constitui um espaço nulo da matriz de predição, a incerteza na predição pode se tornar infinita. Isto significa que, tomando um tamanho do passo κ tendendo a infinito, os limites das variáveis de predição podem ser também infinitos:

$$(E.4.8) \quad \lim_{|\kappa| \rightarrow \infty} |Y(\Theta' + \kappa \cdot d_j)| = \infty.$$

Exemplo ilustrativo 4.6

Não identificabilidade prática dos parâmetros

Considere o modelo do exemplo ilustrativo 4.5. Suponha que, para as condições experimentais dadas, os seguintes valores da variável de resposta tenham sido obtidos experimentalmente:

x_1	x_2	y^E
1	1	3,01
2	2	5,98
3	3	8,95

Após a estimação de parâmetros, foi calculada a incerteza de predição para a condição $x_1 = x_2 = 4$ e para a condição $x_1 = 4$ e $x_2 = 4,01$, considerando valores de passos κ elevados na direção característica associada a um valor característico nulo.

A partir dos dados do exemplo ilustrativo 4.5, ao se fixar um dos parâmetros arbitrariamente e estimar o outro, têm-se:

Valor arbitrado de θ_1	Valor estimado de θ_2	$FObj$
0,5	2,487	$6,860 \times 10^{-4}$

em que se sabe que a direção característica com valor característico nulo é $d_1 = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$. Partindo dos

valores estimados, os parâmetros podem ser modificados na direção d_1 como:

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 0,5 \\ 2,487 \end{bmatrix} + \kappa \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}.$$

Tomando valores de κ elevados, pode-se calcular o valor da predição na medida em que se caminha na direção d_1 para cada uma das condições informadas no enunciado, conforme segue:

	Condição 1	Condição 2
x_1	4	4
x_2	4	4,01

κ	θ_1	θ_2	y^C - Condição 1	y^C - Condição 2
0	0,5	2,487	11,95	11,95
10000	-9999,5	10002,487	11,95	21,95
1000000	-999999,5	1000002,487	11,95	1011,95
10000000	-9999999,5	10000002,49	11,95	10011,95
100000000	-99999999,5	100000002,5	11,95	100011,95

Na condição 1, a direção característica com incerteza infinita é um espaço nulo da matriz de predição. Para tal caso, a variação dos parâmetros nesta direção não implica em valores infinitos na incerteza de predição. Por outro lado, na condição 2, há uma pequena variação do x_2 , suficiente para levar a uma incerteza ilimitada da predição.

4.2.3 Informação desprezada ao fixar o valor dos parâmetros

Uma forma de tratar rigorosamente o problema de identificabilidade de parâmetros consiste em determinar a região de confiança dos parâmetros e a correspondente incerteza associada à predição. Isto deve ser feito mesmo para os problemas de estimação mal condicionados, em que não é possível obter a V_{θ} . Em outras palavras, deve-se determinar a região de confiança dos parâmetros, que pode ser não delimitada, e os correspondentes valores preditos do modelo para todos os valores de parâmetros presentes nesta região em cada condição experimental de interesse.

Os valores obtidos na região de incerteza de predição do modelo para uma dada condição experimental representam os valores possíveis que o experimentador pode esperar obter a partir de experimentos. Todavia para problemas de estimação mal condicionados, em que a matriz V_{θ} pode não ser obtida e alguns parâmetros são fixados na estimativa

inicial, a avaliação da incerteza pode tornar-se completamente enganosa. De fato, como mencionado anteriormente, a região de confiança dos parâmetros calculada a partir de apenas alguns parâmetros pode não ter qualquer sentido, de forma que o uso destes valores de parâmetros associados às condições experimentais para fins de cálculo da região de confiança da predição podem levar a conclusões enganosas. Em outras palavras, o modelo pode não ser adequado para a predição, embora os resultados da caracterização de incerteza demonstrem o contrário.

Uma vez que a diferença entre a incerteza da predição real e a incerteza da predição obtida ao final do procedimento de identificabilidade pode ser bastante elevada, é importante investigar a área desprezada da incerteza da predição quando parâmetros têm seus valores fixados na estimativa inicial. Para tal investigação, propõe-se um procedimento numérico composto por três etapas básicas: (i) caracterização da incerteza dos parâmetros, (ii) caracterização da incerteza da predição para dada condição experimental e (iii) caracterização da área de incerteza desprezada; como apresentado a seguir:

(i). Etapa 1 – Caracterização da incerteza dos parâmetros

A - Obter combinações dos parâmetros dentro da região de busca segundo o número de pontos pré-determinados.

B - Calcular a função objetivo para cada uma das combinações de valores parâmetros.

C - Selecionar o menor dentre os valores da função objetivo obtido.

D - Calcular o limite de corte da função objetivo empregando o teste F, a partir do menor valor obtido da função objetivo.

E - Os valores de parâmetros com função objetivo abaixo do limite de corte são considerados adequados; constituindo a região de confiança dos parâmetros.

(ii). Etapa 2 – Caracterização da incerteza da predição para dada condição experimental

A - Efetuar simulações com o modelo a partir de todas as combinações de parâmetros contidos na região de confiança dos parâmetros. O valor máximo e mínimo das variáveis de saída y_i é armazenado, $y_{i,max}$ e $y_{i,min}$.

Nota: Caso se deseja varrer condições de experimentação, basta repetir o procedimento acima para cada condição investigada.

(iii). Etapa 3 – Cálculo da área desprezada da predição quando parâmetros são fixados.

A - Calcular o limite de corte da função objetivo considerando o teste F, de acordo com o menor valor de função objetivo obtido na etapa 1B. O valor do teste F é segundo o número de parâmetros não fixados.

B - Uma vez dispondo do valor da função objetivo para cada combinação de parâmetros, verificar quais destas combinações encontram-se dentro da região de confiança segundo o teste F.

C - Para cada valor das combinações dos parâmetros, para uma dada condição de experimentação, efetuar simulações com todos os parâmetros contidos nesta região de confiança. O valor máximo e mínimo das variáveis de saída y'_i é armazenado, $y'_{i,\max}$ e $y'_{i,\min}$.

D - Para cada variável de saída, os possíveis valores de resposta desprezados, o volume desprezado da região de confiança de predição $\Phi_{\hat{y}}^{desp}$ pode ser obtido

considerando-se um hipercubo como:
$$\Phi_{\hat{y}}^{desp} = 1 - \prod_{i=1}^{NY} \frac{y'_{i,\max} - y'_{i,\min}}{y_{i,\max} - y_{i,\min}}.$$

Além dos dados experimentais, basicamente dois tipos de informações relacionadas tanto aos parâmetros quanto às variáveis independentes do modelo são requeridas para essa investigação: (i) a região de busca e, (ii) o número de pontos. A região de busca representa a faixa de valores a ser investigada e é delimitada pelos limites inferior e superior, enquanto o número de pontos representa quantos valores desta região devem avaliados pelo procedimento numérico; da relação entre a região de busca e do número de pontos obtém-se os pontos para avaliação, que são os valores de parâmetros e de variáveis independentes que devem ser avaliados pelo procedimento.

Naturalmente, para um maior número de pontos é esperado obter regiões de confiança melhor caracterizadas. Entretanto, a caracterização adequada da região de confiança requer que o número de pontos esteja distribuído em uma região de busca válida. Desta forma, a região de busca das variáveis independentes deve ser definida de forma a

englobar os dados experimentais disponíveis. Semelhantemente, a região de busca dos parâmetros deve englobar as estimativas iniciais dos parâmetros.

O procedimento numérico proposto a fim de investigar o comportamento da incerteza quando parâmetros têm seus valores fixados no problema de identificabilidade foi implementado em linguagem computacional Fortran 95.

É importante enfatizar, que por se tratar de um problema de natureza combinatorial, o procedimento numérico requer grande número de avaliações do modelo, de forma que para modelos contendo elevado número de parâmetros e/ou de variáveis independentes, esta abordagem pode tornar-se computacionalmente proibitiva. Outra questão importante, é que a análise de modelos mais complexos, tais como as estruturas lógico-matemática empregadas para descrever sistemas dinâmicos, pode introduzir dificuldades numéricas adicionais à implementação. Desta forma, tais assuntos deverão ser tratados em trabalhos futuros. Todavia, deve ficar claro que o procedimento numérico proposto não deve ser útil apenas para tratamento de modelos complexos, mas principalmente de sistemas reais, a despeito da complexidade de tais sistemas.

Dois estudos de caso são apresentados: (i) modelo linear simples e (ii) modelo não linear simples; cada exemplo é composto por um modelo de estrutura matemática simples contendo apenas dois parâmetros, o que permite maior clareza na análise e interpretação dos resultados. Nestes exemplos, os valores numéricos foram definidos a fim de garantir a identificabilidade de todos os parâmetros dos modelos, visto que a caracterização estatisticamente rigorosa da incerteza dos parâmetros e da incerteza da predição somente é possível quando todos os parâmetros são estimáveis.

4.2.3.1 Estudo de caso 1: Modelo hipotético 4

Contendo apenas dois parâmetros e uma variável, o modelo linear nos parâmetros apresentado na Equação (E 4.9) possui estrutura bem simples; por este motivo, a análise dos resultados torna-se trivial.

$$(E 4.9) \quad y = \theta_1 x_1 + \theta_2 x_1^2,$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^1$ e $\theta \in \mathfrak{R}^2$.

A dependência do desempenho do procedimento de estimação de parâmetros com a incerteza experimental assumida é esperada mesmo para problemas de estimação bem condicionados. Esta é uma informação muito importante, mas geralmente é negligenciada e arbitrariamente definida devido ao árduo trabalho experimental requerido para caracterizar os erros experimentais; na prática, frequentemente os experimentos são difíceis, demorados e caros. Apesar da grande relevância, a influência da incerteza experimental sobre o procedimento de identificabilidade de parâmetros tem sido negligenciada pela literatura.

A fim de avaliar os efeitos da incerteza experimental sobre o problema de identificabilidade, especialmente quando parâmetros são fixados em valores arbitrários, foram elaborados três estudos de casos, considerando para cada caso um valor de incerteza experimental. Tais casos ilustram diferentes cenários experimentais frente ao problema de identificabilidade de parâmetros. No caso 1, a variância experimental é definida como sendo 10% do valor da medida e neste contexto representa o valor máximo tolerável. Para o caso 2, a variância experimental é considerada como sendo 25% do valor da medida e neste contexto representa um valor elevado, realidade experimental de muitas áreas. Por fim, o caso 3, define a variância experimental como sendo 1% do valor da medida e neste caso representa um valor baixo, que representa o valor usual adotado na maioria dos experimentos em que esta entidade não pode ser rigorosamente caracterizada por meio da estatística. Assim, para cada caso tratado neste exemplo tem-se um conjunto de dados experimentais, que foram simulados a partir do modelo, sem considerar erros nas variáveis de entrada, utilizando os valores exatos dos parâmetros definidos como $\theta_1 = 1,0$ e $\theta_2 = 1,0$, adicionando às saídas ruídos com distribuição normal de média igual a 0,0 e de variância experimental conhecida definida de acordo com cada caso.

Os conjuntos de dados experimentais empregados para o estudo do modelo linear nos parâmetros são apresentados pela Tabela 4.1.

Com relação aos parâmetros do modelo simples linear nos parâmetros, as informações utilizadas são apresentadas pela Tabela 4.2. Uma vez que neste contexto os parâmetros deste modelo não apresentam qualquer restrição física, o intervalo fornecido para os limites inferior e superior dos parâmetros foram definidos em valores que levavam a uma caracterização estatística adequada da incerteza dos parâmetros e conseqüentemente da incerteza da predição do modelo; para todos os casos tratados o limite inferior dos

parâmetros foi definido em valores abaixo de zero. Embora os valores exatos dos parâmetros seja uma informação conhecida, é sabido que a faixa válida de valores de parâmetros depende da quantidade e da qualidade dos dados experimentais; dado que não pode ser conhecido sem prévia caracterização da incerteza dos parâmetros. Assim, o intervalo dos parâmetros foi definido a partir de uma investigação preliminar desta faixa válida, em que inicialmente fornece-se um amplo intervalo em torno do valor exato do parâmetro, que permite identificar se este intervalo fornecido deve ser reduzido ou aumentado, de forma a delimitar adequadamente o intervalo para construção adequada da região de confiança dos parâmetros.

Tabela 4.1 – Conjuntos de dados experimentais empregados para o estudo do modelo simples linear nos parâmetros.

Caso 1: $\sigma_y^2 = 0,1$

N	x	y^E
1	0,100	0,043
2	1,000	2,141
3	1,100	2,339

Caso 2: $\sigma_y^2 = 0,25$

N	x	y^E
1	0,100	-0,057
2	1,000	2,352
3	1,100	2,382

Caso 3: $\sigma_y^2 = 0,01$

N	x	y^E
1	0,100	0,103
2	1,000	2,014
3	1,100	2,312

Para o tratamento rigoroso do problema de identificabilidade, as informações relacionadas à variável independente são fornecidas pela Tabela 4.3. Assim como para os parâmetros, neste contexto a variável independente no modelo simples linear não apresenta qualquer restrição física, de forma que para todos os casos tratados os limites inferior (-2,0) e superior (2,0) da variável independente foram definidos de acordo com a faixa

experimental. Uma vez que foi utilizada a mesma faixa experimental, empregou-se o mesmo intervalo para todos os casos tratados neste exemplo.

Tabela 4.2 – Informações relacionadas aos parâmetros do modelo linear nos parâmetros simples.

Caso 1: $\sigma_y^2 = 0,1$

	θ_1	θ_2
Valor exato – θ^*	1,0	1,0
Limite inferior – θ_L	-30,0	-30,0
Limite superior – θ_U	30,0	30,0
Número de pontos – θ_{Grid}	600	600

Caso 2: $\sigma_y^2 = 0,25$

	θ_1	θ_2
Valor exato – θ^*	1,0	1,0
Limite inferior – θ_L	-60,0	-60,0
Limite superior – θ_U	60,0	60,0
Número de pontos – θ_{Grid}	600	600

Caso 3: $\sigma_y^2 = 0,01$

	θ_1	θ_2
Valor exato – θ^*	1,0	1,0
Limite inferior – θ_L	-2,0	-2,0
Limite superior – θ_U	4,0	4,0
Número de pontos – θ_{Grid}	600	600

Tabela 4.3 – Informações relacionadas à variável independente do modelo linear nos parâmetros simples, utilizada nos casos 1, 2 e 3.

Número de pontos – x_{Grid}	Limite inferior – x_L	Limite superior – x_U
400	-2,0	2,0

Na Figura 4.10 são apresentados os resultados para a caracterização de incertezas dos parâmetros do modelo simples linear nos parâmetros. De acordo com a Figura 4.10, para todos os casos estudados, os parâmetros θ_1 e θ_2 apresentam forte correlação. Como esperado, esta figura ilustra a propagação da incerteza experimental para a incerteza dos parâmetros, em que se observa o alargamento da faixa válida de valores de parâmetros para os casos em que o erro experimental é mais elevado (Figura 4.10a e Figura 4.10b). Tomando como referência o caso 1 ilustrado pela Figura 4.10a, em que a incerteza

experimental foi definida como 10% da medida, a comparação com os demais casos demonstra claramente o grande impacto sobre a estimação e caracterização dos valores de parâmetros quando esta entidade estatística apresenta valor elevado (caso 2, Figura 4.10b), ou como em muitos problemas em que esta informação não pode ser estatisticamente calculada e é admitida abaixo do valor real (caso 3, Figura 4.10c) sendo subestimada.

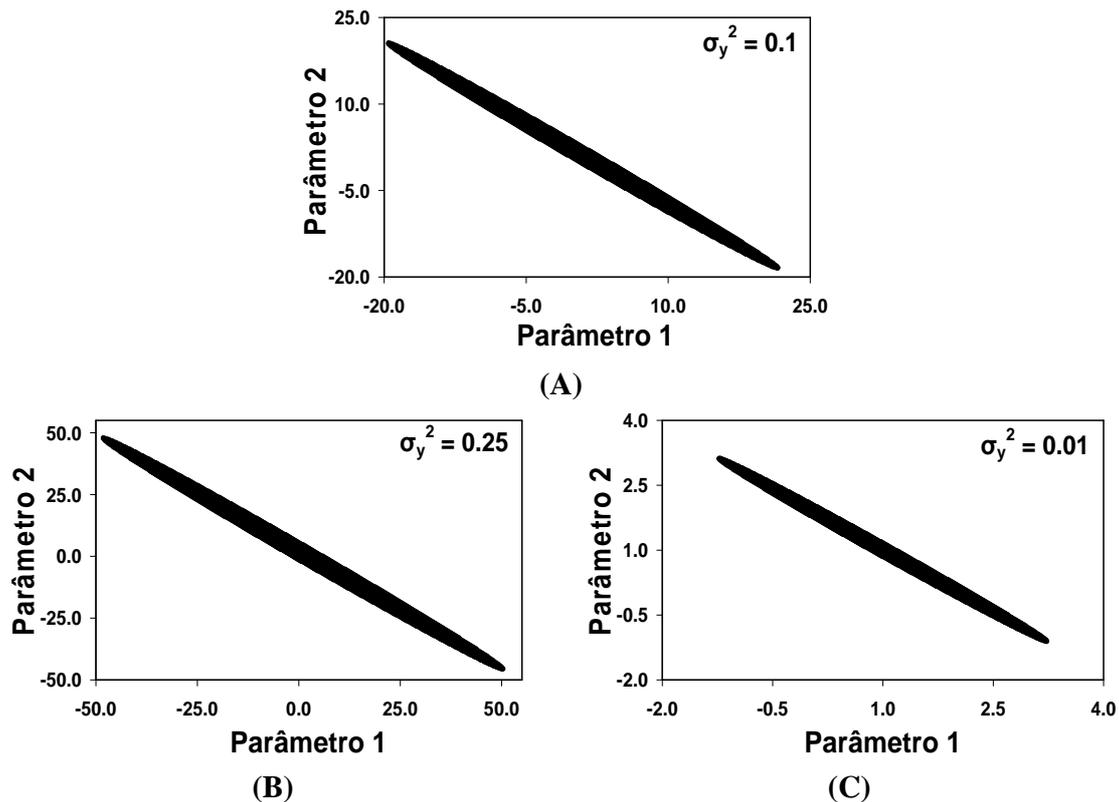


Figura 4.10 - Caracterização da incerteza dos parâmetros do modelo hipotético 4, quando todos os parâmetros são identificáveis, para diferentes valores de incerteza experimental: A - caso 1, B - caso 2 e, C - caso 3.

A Figura 4.11 apresenta as regiões de confiança da predição para os três casos tratados neste exemplo: (i) Caso 1 - $\sigma_y^2 = 0,1$, (ii) Caso 2 - $\sigma_y^2 = 0,25$ e, (iii) Caso 3 - $\sigma_y^2 = 0,01$. Como esperado, estes resultados demonstram claramente a propagação da incerteza experimental e da incerteza dos parâmetros sobre a incerteza da predição.

A Figura 4.11 tem uma forte implicação sobre o uso de procedimentos de identificabilidade, pois demonstra claramente que ao se fixar o valor de um parâmetro, o uso do modelo pode tornar-se inadequado para predições em regiões de extrapolação e até mesmo de interpolação dos dados experimentais; há uma grande incerteza associada às

regiões sem informação experimental e um grande desvio da predição que não pode ser contabilizado pode estar sendo aceito quando um ou mais parâmetros são fixados. Este resultado é muito importante, pois demonstra que quando não se está no espaço nulo da matriz B , o uso do procedimento de identificabilidade é limitado pela flexibilidade do modelo, em que um conjunto de parâmetros limitado faz com que a função objetivo seja aceitável até um determinado limite de corte.

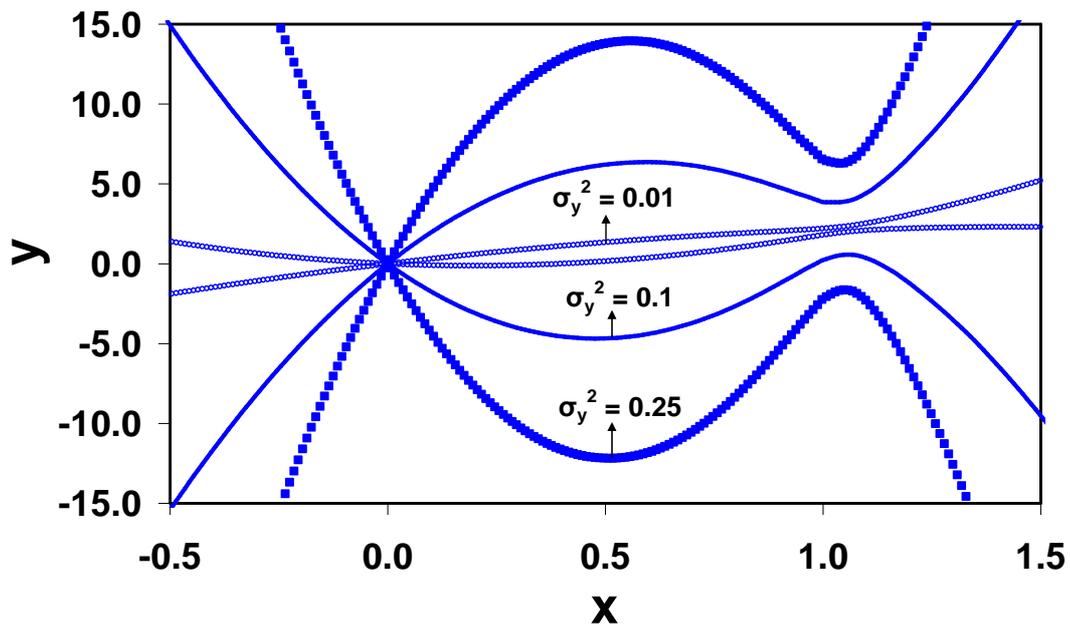


Figura 4.11 - Caracterização da incerteza da predição do modelo hipotético 4, quando todos os parâmetros são identificáveis, para diferentes valores de incerteza experimental (casos 1, 2 e 3).

O volume desprezado, dado em unidades adimensionais, refere-se ao volume da região de confiança da incerteza da predição que não pode ser contabilizada quando o parâmetro θ_i tem seu valor fixado em um dado valor; como mostra a Figura 4.12, os gráficos A, C e E em que θ_2 é estimado enquanto θ_1 é fixado, e os gráficos B, D e F em que o contrário ocorre. Na Figura 4.12 ao fixar o valor de um dos parâmetros do modelo simples linear, o menor valor obtido para o volume relativo desprezado está acima de 0,9, o que revela elevado valor de incerteza de predição que não pode ser contabilizado (maior que 90%) quando este parâmetro é definido como não identificável. Note que, na Figura 4.12, a coordenada x do gráfico representa a faixa válida de valores do parâmetro θ_i , em

que fixando o parâmetro para um dado valor obtém-se o respectivo volume desprezado de região de confiança da incerteza de predição; o menor valor do volume de região de confiança desprezado está relacionado ao valor exato dos parâmetros, $\theta_1 = \theta_2 = 1,0$, o que configura o melhor caso, em que se ultrapassa 94% de região da incerteza que não pode ser contabilizada.

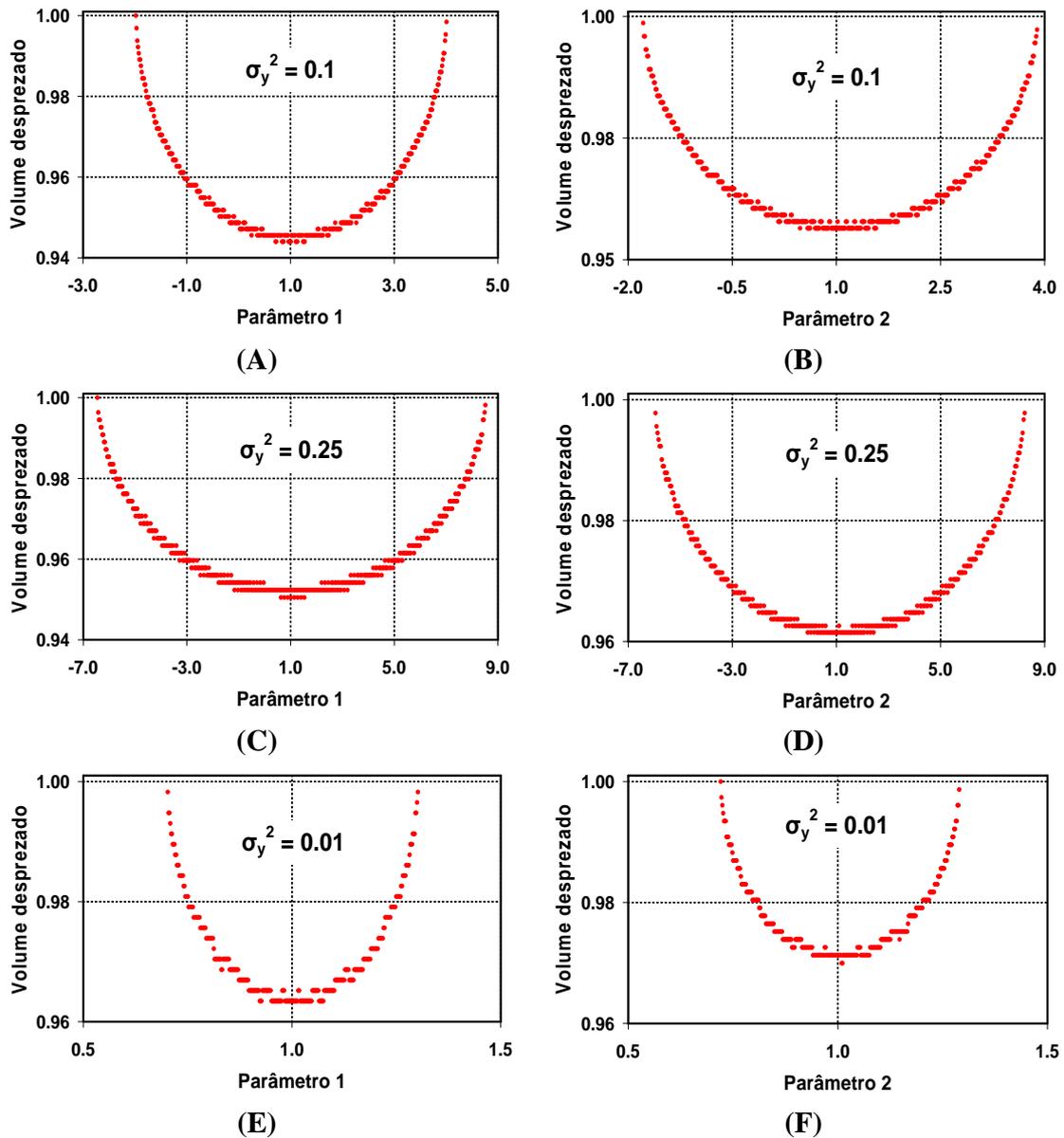


Figura 4.12 – Volume desprezado relativo da incerteza da predição do modelo hipotético 4, quando todos os parâmetros são identificáveis, para diferentes valores de incerteza experimental: A-C-E- quando se fixa o parâmetro 2, e B-D-F- quando se fixa o parâmetro 1.

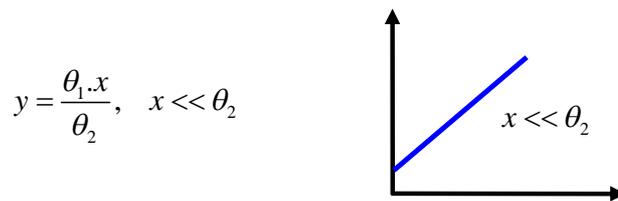
4.2.3.2 Estudo de caso 2: modelo hipotético 5

Contendo dois parâmetros e apenas uma variável independente, o modelo representado pela Equação (E 4.10) é provavelmente o mais popular na classe de modelos matemáticos empregados para descrever processos biotecnológicos e ecológicos. Empregado por MICHAELLIS-MENTEN (1913) para descrever as cinéticas enzimáticas, por MONOD (1942) para descrever crescimento bacteriano, este modelo apresenta comportamento compatível com muitas observações no estudo da cinética de sistemas biológicos. Estrutura similar foi usada por LANGMUIR (1916) para descrever fenômenos de adsorção, também muito comum a estes sistemas. Recentemente tal modelo foi empregado em estudos de identificabilidade de parâmetros devido a sua grande relevância para descrição do comportamento biológico (OMLIN e REICHERT, 1999).

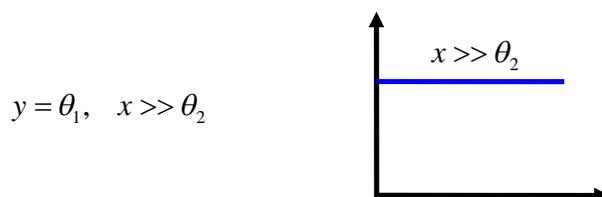
$$(E\ 4.10) \quad y = \theta_1 \frac{x}{\theta_2 + x}$$

em que as variáveis y e x representam respectivamente a taxa de conversão e a concentração de substrato, e os parâmetros θ_1 e θ_2 representam a taxa de conversão máxima e a concentração na qual a taxa de conversão é metade deste valor máximo, respectivamente.

De acordo com a Equação (E 4.10), as taxas de conversão são lineares para baixas concentrações de substrato, mas para elevadas concentrações as taxas de conversão tornam-se constantes, como pode ser verificado na Figura 4.13. Note que a forma estrutural do modelo adequada para a estimação de parâmetros está condicionada a magnitude dos parâmetros e das observações experimentais disponíveis. Este tipo de correlação é muito difícil de avaliar, especialmente em se tratando de modelos não lineares contendo grande número de parâmetros, de forma que nos modelos matemáticos típicos empregados para representar os processos relativos às engenharias química e ambiental não se é possível evitar este efeito antes do procedimento de estimação.



(A)



(B)

Figura 4.13 - Relações entre a taxa de conversão y e a concentração de substrato x do modelo hipotético 5: A - o valor da concentração de substrato é muito baixo quando comparado ao valor do segundo parâmetro, levando a uma taxa de conversão linear, B - o valor da concentração de substrato é muito elevado quando comparado ao valor do segundo parâmetro, levando à saturação.

Para este exemplo, cinco experimentos foram simulados com o modelo, como mostra a Tabela 4.4. Os experimentos foram corrompidos com ruído gerado de acordo com a distribuição normal de probabilidade, com média zero e variância $\sigma_y^2 = 0,01$. Foram admitidos como valores exatos dos parâmetros: $\theta_1 = 2,5$ e $\theta_2 = 15,0$.

Tabela 4.4 – Dados experimentais para estudo da identificabilidade do modelo hipotético 5.

N	x	y^E
1	40,0	1,877
2	45,0	1,926
3	50,0	1,965
4	55,0	2,069
5	60,0	2,023

A Tabela 4.5 e a Tabela 4.6 apresentam, respectivamente, os valores empregados para a investigação dos parâmetros e para a investigação das variáveis independentes do modelo simples não linear. De acordo com o comportamento dos sistemas biológicos, contexto no qual este modelo é amplamente empregado, foi admitido como restrição física

que os parâmetros não devem assumir valores negativos, o que leva a definir em zero o limite inferior da região de busca dos parâmetros. É importante ressaltar, que apesar da simplicidade deste exemplo, o uso de intervalos amplos torna-se computacionalmente proibitivo devido ao exorbitante número de avaliações resultantes das combinações de valores dos parâmetros e das variáveis independentes.

Tabela 4.5 – Regiões de busca dos parâmetros do modelo hipotético 5.

	θ_1	θ_2
Valor exato – θ^*	2,5	15,0
Limite inferior – θ_L	0,0	0,0
Limite superior – θ_U	10,0	100,0
Número de pontos – θ_{Grid}	800	800

Tabela 4.6 – Região de busca da variável independente do modelo hipotético 5.

Número de pontos – x_{Grid}	Limite inferior – x_L	Limite superior – x_U
300	0,0	70,0

A Figura 4.14 apresenta a região de confiança dos parâmetros - A, e a região de confiança da predição - B. De acordo com a Figura 4.14A, para todos os casos estudados, os parâmetros θ_1 e θ_2 apresentam forte correlação e, apesar das restrições físicas, a região de confiança dos parâmetros indica valores abaixo de zero para o parâmetro θ_2 . A Figura 4.14B demonstra que para baixas concentrações de substrato, a taxa apresenta um amplo intervalo de confiança, que vai da linearidade à saturação, sendo que saturação nesta região de baixa concentração não condiz com o comportamento de sistemas biológicos.

Similarmente ao observado para o exemplo hipotético 4, a Figura 4.14B apresenta o risco de se utilizar o modelo para predição, ao manter os valores dos parâmetros não identificáveis nas estimativas iniciais. O amplo intervalo de confiança da predição indica que o modelo pode mudar completamente o significado do fenômeno, em região na qual a estimação dos parâmetros do modelo não dispunha de dados experimentais. Tal resultado corrobora a observação de que quando não se está no espaço nulo da matriz B , a flexibilidade do modelo é um fator chave para o adequado uso do procedimento de identificabilidade.

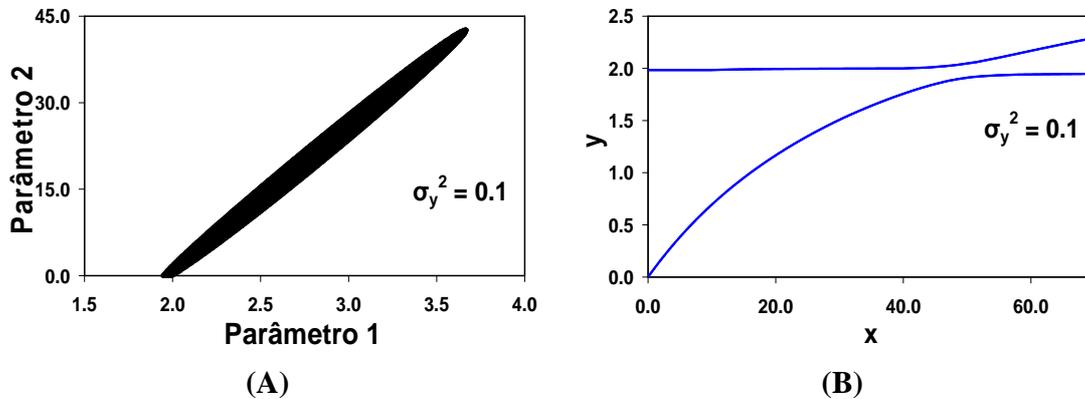


Figura 4.14 - Caracterização das incertezas dos parâmetros e da predição do modelo hipotético 5 quando todos os parâmetros são identificáveis: A - região de confiança da incerteza dos parâmetros e; B - região de confiança da incerteza da predição.

O volume da região de confiança da incerteza da predição que não pode ser contabilizada quando o parâmetro θ_i tem seu valor fixado em um dado valor é apresentado pela Figura 4.15; no gráfico A, θ_2 é estimado enquanto θ_1 é fixado, e no gráfico B, ocorre o contrário. Na Figura 4.12, ao fixar o valor de um dos parâmetros do modelo simples não linear, o menor valor obtido para o volume desprezado relativo está acima de 0,75, o que revela elevado valor de incerteza de predição desprezado, mesmo quando os parâmetros são fixados em seus valores verdadeiros, $\theta_1 = 2,5$ e $\theta_2 = 15,0$.

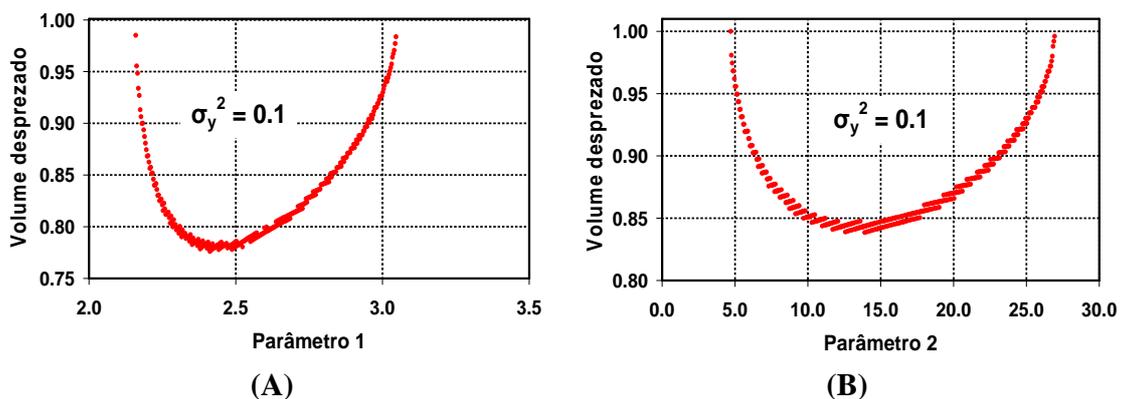


Figura 4.15 - Volume desprezado relativo da incerteza da predição do modelo hipotético 5: A - ao fixar o valor do parâmetro 2 e, B - ao fixar o valor do parâmetro 1.

Naturalmente, em se tratando de modelos complexos, o maior desafio é a complexa inter-relação entre os parâmetros e a predição do modelo, que requer que o espaço de parâmetros seja densamente amostrado para capturar todos os cenários estatisticamente possíveis da predição. Para modelos com elevado número de parâmetros isto é numericamente custoso ou mesmo inviável, o que é mais comumente conhecido na literatura como a maldição da dimensionalidade (MARIMONT e SAPHIRO, 1979).

Neste contexto, o problema de identificabilidade de parâmetros vem sendo abordado pela literatura em termos do perfil de verossimilhança χ_{PL}^2 (do termo em inglês *Profile Likelihood - PL*) (RAUE et al., 2009; HUG et al., 2013; KREUTZ et al., 2013), em trabalhos independentes desta tese.

O perfil de verossimilhança (VENZON e MOOLGAVKAR, 1988; MURPHY e van der VAART, 2000) pode ser calculado para cada parâmetro individualmente:

$$(E 4.11) \quad \chi_{PL}^2(\theta_i) = \min_{\theta_{j \neq i}} [\chi^2(\theta)],$$

que representa a re-otimização de $\chi^2(\theta)$ com respeito a todos os parâmetros $\theta_{j \neq i}$, para cada valor de θ_i mantido fixo. Então, o perfil de verossimilhança do parâmetro mantém o χ^2 tão pequeno quanto possível ao longo de θ_i .

De acordo com RAUE et al. (2009), nos procedimentos de identificabilidade prática, a curvatura da função χ^2 é computada utilizando a sua aproximação quadrática dada pela matriz Hessiana ou a matriz *FIM*, exata para modelos lineares. Desta forma, tais procedimentos são apropriados quando as relações funcionais entre os parâmetros, emergente da não identificabilidade estrutural, são lineares (RAUE et al., 2009). Portanto, em se tratando de modelos complexos, a não identificabilidade prática dos parâmetros pode não ser detectada, porque a aproximação quadrática pode não ser capaz de explicar o comportamento da função $\chi_{PL}^2(\theta_i)$.

RAUE et al. (2009) compararam o perfil de probabilidade da função $\chi_{PL}^2(\theta_i)$ com a aproximação quadrática, neste caso a matriz Hessiana, para classificação do parâmetro θ_i quanto à identificabilidade em quatro casos, como mostram a Figura 4.16 e a Tabela 4.7.

Os resultados apresentados pelas Figura 4.16 e Tabela 4.7 corroboram a observação de que em cenários de modelos complexos, a aproximação quadrática pode não ser adequada para avaliação da identificabilidade dos parâmetros.

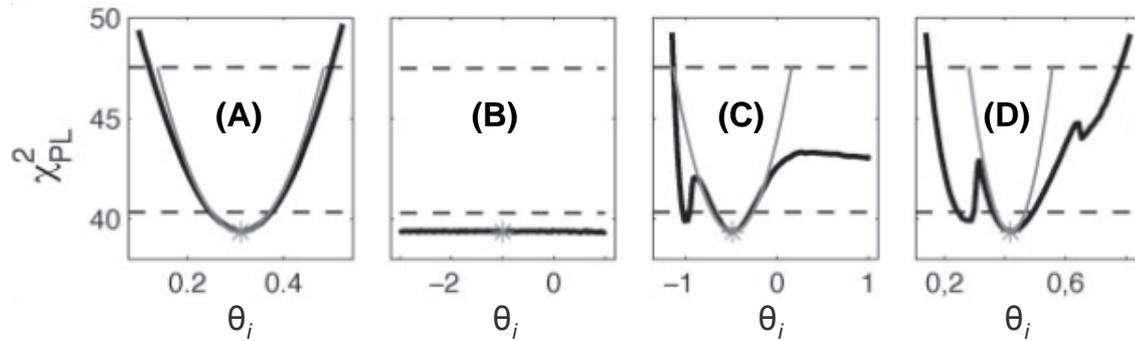


Figura 4.16 - Classificação do parâmetro θ_i quanto a identificabilidade em quatro casos distintos (Adaptado de RAUE et al. (2009)): A - identificável, B - não identificabilidade estrutural, C - não identificabilidade prática, e D - identificável. Em que representam: (—) perfil de probabilidade do parâmetro θ - χ_{PL}^2 , (- -) aproximação quadrática, (*) o valor estimado do parâmetro θ_i , (- -) o intervalo de confiança do parâmetro θ_i .

Tabela 4.7 – Classificação do parâmetro θ_i quanto a identificabilidade em 4 casos distintos (Adaptado de RAUE et al. (2009)):

	Valor estimado de θ_i	σ_{PL}^-	σ_{PL}^+	σ_{Hess}^-	σ_{Hess}^+	Identificabilidade
A	0,31	0,12	0,50	0,14	0,48	Identificável
B	1,00	$-\infty$	$+\infty$	-33,00	31,00	Não identificabilidade estrutural
C	-0,49	-1,14	$+\infty$	-1,14	0,15	Não identificabilidade prática
D	0,42	0,15	0,78	0,28	0,56	Identificável

Na Figura 4.16, os casos B e C tratam da não identificabilidade do parâmetro θ_i , de acordo com o perfil de verossimilhança, limite inferior (σ_{PL}^-) e limite superior (σ_{PL}^+) dados pela Tabela 4.7: (B) não identificabilidade estrutural $\sigma_{PL} = [-\infty \quad +\infty]$, e (C) não identificabilidade prática $\sigma_{PL} = [-1,14 \quad +\infty]$. Entretanto, este resultado não pode ser observado para a aproximação quadrática, limite inferior (σ_{Hess}^-) e limite superior (σ_{Hess}^+), em especial para o caso C, em que a aproximação quadrática não descreve o comportamento limitado da função $\chi_{PL}^2(\theta_i)$.

Análogo ao perfil de verossimilhança dos parâmetros do modelo, $\chi_{PL}^2(\theta_i)$, têm-se o perfil de verossimilhança da predição χ_{PPL}^2 (HUG et al., 2013; KREUTZ et al., 2013),

cuja ideia básica consiste em calcular o intervalo de confiança da predição sem explicitamente amostrar o espaço de parâmetros. Neste sentido, HUG et al. (2013) propuseram uma metodologia baseada no perfil de probabilidade da predição (PPL - *predict profile likelihood*) χ^2_{PPL} (VENZON e MOOLGAVKAR, 1988; MURPHY e van der VAART, 2000), a fim de calcular o conjunto de predições que são consistentes com as medidas experimentais disponíveis sem explicitamente calcular a incerteza dos parâmetros. Esta metodologia é baseada no conceito de observabilidade, que busca caracterizar quando a dinâmica de algumas variáveis que não podem ser medidas pode ser inferida por meio de um conjunto de experimentos viáveis. O esforço de se avaliar o espaço de dimensões NP dos parâmetros é reduzido para a avaliação do espaço de dimensão um da predição, sendo requerido apenas o cálculo confiável do valor de máxima probabilidade.

4.3 CONCLUSÕES PARCIAIS

Aspectos conceituais do problema de identificabilidade quanto à classificação dos parâmetros e à caracterização de incertezas remontam à questão do uso apropriado de tais procedimentos. Parece natural que a definição da classificação dos parâmetros quanto à identificabilidade seja dada em termos da região de confiança dos parâmetros, mas há uma grande dificuldade em caracterizar as incertezas em problemas de estimação mal condicionados, o que pode levar a resultados enganosos e elevado risco no uso indiscriminado dos procedimentos de identificabilidade.

Fica claro que a seleção dos parâmetros deve ser realizada levando em consideração os parâmetros que podem ser fixados na estimativa inicial, sem prejuízo da predição do modelo. Obviamente, esta tarefa não é trivial, mas é de suma importância para assegurar a validade dos procedimentos de identificabilidade.

CAPÍTULO 5. DESENVOLVIMENTO DE NOVOS PROCEDIMENTOS DE IDENTIFICABILIDADE DE PARÂMETROS

A fim de reduzir a dependência da identificabilidade com a estimativa inicial dos parâmetros, bem como com o critério de seleção adotado, propõem-se dois novos procedimentos, com diferentes focos: (i) avaliação intensiva - a classificação robusta dos parâmetros e (ii) seleção conjunto a conjunto - a redução do número de avaliações. Ambos os procedimentos consideram simultaneamente a identificabilidade e a estimação. Como estudo de caso, destaca-se o árduo problema de identificabilidade de parâmetros das rotas metabólicas do microorganismo *E. coli* K-12 W3110, que apresenta 131 parâmetros e 5 condições iniciais desconhecidas, que devem ser estimados a partir de um conjunto escasso de experimentos. Os procedimentos propostos apresentam grande robustez aos típicos problemas numéricos, assegurando o ajuste do modelo aos dados experimentais.

Na maioria dos procedimentos reportados pela literatura, a estimação é uma etapa final ou exclusiva à identificabilidade. Uma vez que a identificabilidade do parâmetro é realizada com base na estimativa inicial dos parâmetros, que podem ser ou não adequadas aos valores reais dos parâmetros, este procedimento pode resultar em problemas de estimação mal condicionados. Esta é uma questão importante e, embora ocorra com certa frequência, tem sido negligenciada pela literatura.

O parâmetro selecionado pode levar a problemas na estimação por mau condicionamento, pela inclusão do vetor de sensibilidade deste parâmetro à matriz *FIM*, ou ainda por problemas numéricos na resolução do modelo e convergência do otimizador relacionados com estimativas iniciais inadequadas e também limites mal estabelecidos, uma vez que informações sobre os parâmetros são pouco conhecidas ou desconhecidas. Em se tratando de modelos dinâmicos complexos, informações sobre os parâmetros são raramente bem conhecidas e outras informações importantes do sistema, e.g. condições iniciais, também frequentemente não o são.

Outra questão relevante é o quanto cada parâmetro incluído afeta os valores estimados dos demais parâmetros selecionados bem como a qualidade da predição do modelo. Ao sugerir que a estimação seja simultânea à identificabilidade, a ideia é verificar o quanto a inclusão do parâmetro modifica o ajuste do modelo quando o conjunto de parâmetros é selecionado. Além disso, a estimação do conjunto de parâmetros selecionados modifica os seus valores, permitindo que, a cada etapa da seleção, valores mais adequados destes parâmetros sejam empregados na seleção dos demais parâmetros.

A estratégia para amenizar este problema é realizar a estimação simultânea à seleção, o que garante a estimação dos parâmetros selecionados (e.g.: SECCHI et al., 2006; WU et al., 2007; WU et al., 2011; MCLEAN et al., 2012). A estimação simultânea à identificabilidade resulta em um procedimento que permite uma avaliação mais adequada dos parâmetros, reduzindo as dificuldades de identificabilidade devido a estimativas iniciais inadequadas, problemas de mau condicionamento e outros problemas numéricos e obtendo ao final os valores estimados de parâmetros que influenciam a predição do modelo.

A fim de garantir a estimação dos parâmetros selecionados, uma abordagem bastante comum é adotar a singularidade da matriz *FIM* como um dos critérios de paradas em algoritmos de identificabilidade (e.g.: WEIJERS e VANROLLEGHEM, 1997;

SANDINK et al., 2001; LI et al., 2004; SECCHI et al., 2006; LUND e FOSS, 2008; THOMPSON et al., 2008). Quando tal ponto é atingido pelo algoritmo, a etapa de seleção dos parâmetros é interrompida e os parâmetros restantes são classificados como parâmetros não identificáveis. Especialmente quando a identificabilidade realiza a seleção e a estimação simultâneas (e.g.: SECCHI et al., 2006; WU et al., 2007; WU et al., 2011; McLEAN et al., 2012), não é desejável fixar os parâmetros restantes como parâmetros não identificáveis sem uma avaliação do potencial de estimação destes parâmetros, porque o problema de estimação é modificado a cada parâmetro selecionado.

Um grande desafio para tratar em procedimentos de identificabilidade, mesmo em procedimentos que incluam a estimação simultânea, é que os parâmetros não selecionados são avaliados nas suas estimativas iniciais, que são provavelmente inadequadas. A literatura aborda o uso da técnica Monte Carlo (CHU e HAHN, 2007) e a seleção e estimação simultâneas como alternativas para assegurar o bom condicionamento do problema de estimação de parâmetros selecionados (e.g.: SECCHI et al., 2006; WU et al., 2007; WU et al., 2011; McLEAN et al., 2012). Mais apropriadamente, a avaliação dos parâmetros subsequentes deve ser realizada com base nos valores de parâmetros selecionados re-estimados (McLEAN et al., 2012), a fim de reduzir a influência da estimativa inicial sobre os parâmetros. Neste sentido, a literatura reporta um trabalho interessante proposto por McLEAN et al. (2012) que permite avaliar a identificabilidade de todos os parâmetros do modelo. De acordo com o procedimento de identificabilidade proposto por McLEAN et al. (2012), os parâmetros selecionados são re-estimados e estes valores são empregados para avaliação dos parâmetros subsequentes, em uma busca em árvore que apresenta um elevado esforço computacional. Outro desafio é que mesmo quando boas estimativas iniciais para os valores de parâmetros estão disponíveis, em modelos complexos a verificação por problemas de identificabilidade (e.g., parâmetros estatisticamente não significativos ou parâmetros correlacionados devido ao mau planejamento de experimentos) é conceitualmente e numericamente uma tarefa árdua.

Em tal cenário uma importante questão a ser respondida é: como reduzir a dependência do procedimento de identificabilidade com as estimativas iniciais dos valores de parâmetros e com o critério de identificabilidade adotado? Uma vez que esta pergunta não está prontamente respondida na literatura, este capítulo aborda o desenvolvimento de

procedimentos numéricos que visam minimizar os impactos da incerteza das estimativas iniciais dos parâmetros e da escolha do critério de seleção sobre o procedimento de identificabilidade.

Os procedimentos numéricos apresentados neste capítulo foram implementados em linguagem computacional FORTRAN 95 e para a obtenção de todos os resultados foi utilizado um computador pessoal Intel Core 2 Duo com 3 GB RAM.

Quando empregada em tais procedimentos numéricos, a estimação de parâmetros foi realizada por meio do pacote computacional ESTIMA&PLANEJA (SCHWAAB et al., 2011), baseado no método Gauss-Newton. A estimação de parâmetros utilizando o método Gauss-Newton é um procedimento que inclui: (i) avaliação do modelo, (ii) *linsearch* (busca em linha) para cálculo do passo com verificação de faixa válida de valores dos parâmetros; (iii) avaliação da matriz *FIM* e sua inversa. Como cada uma destas etapas pode falhar interrompendo a estimação, nesta tese, algumas modificações foram introduzidas neste código para melhoria da convergência, a fim de tratar problemas de identificabilidade que apresentam modelos complexos e dados experimentais escassos.

5.1 CLASSIFICAÇÃO ROBUSTA DOS PARÂMETROS

Um importante aspecto dos procedimentos de identificabilidade de parâmetros é manter fixo os valores dos parâmetros não identificáveis enquanto os parâmetros selecionados são estimados. Uma vez que a estimação de todos os parâmetros do modelo não é possível, a estimação do subconjunto de parâmetros mais estimáveis pode regularizar o problema de estimação mal condicionado e simplificar o problema de otimização associado (KRAVARIS et al., 2012). Todavia, nos procedimentos clássicos de identificabilidade, ao atingir o critério de parada, os parâmetros restantes são classificados como parâmetros não identificáveis e fixados na estimativa inicial, sem avaliar a possibilidade de inclusão destes parâmetros no conjunto de parâmetros identificáveis. Este assunto é abordado na literatura por McLEAN et al. (2012), que propuseram um procedimento de identificabilidade em que o critério de identificabilidade é empregado como ferramenta chave apenas para a seleção do primeiro parâmetro, sendo as demais seleções realizadas de acordo com um critério baseado em valores da função objetivo

resultantes da combinação entre o conjunto de parâmetros selecionados e cada um dos parâmetros restantes. O procedimento de identificabilidade de McLEAN et al. (2012) resulta em um problema combinatorial que requer $\left[\frac{NP(NP+1)}{2} - 1 \right]$ avaliações, o que representa elevado esforço computacional em problemas de identificabilidade contendo elevado número de parâmetros.

Neste sentido, esta tese propõe um novo procedimento numérico de identificabilidade, visando a avaliação de todos os parâmetros do modelo, assegurando que os parâmetros restantes são não identificáveis, quanto ao critério de identificabilidade estabelecido. Este novo procedimento também considera a estimação simultânea à seleção, de forma que os parâmetros que foram avaliados e foram considerados não identificáveis, possam ser novamente avaliados quando um parâmetro é incluído com sucesso no conjunto de parâmetros selecionados. Isto porque, a inclusão bem sucedida de um parâmetro no conjunto de parâmetros selecionados, leva a estimação dos parâmetros deste conjunto modificando o problema de identificabilidade, visto que se têm estes parâmetros em novos valores cuja avaliação da correlação com os demais parâmetros pode ser completamente diferente da avaliação realizada com o valor anterior. Assim, mediante tais mudanças de valores dos parâmetros selecionados, parâmetros avaliados como não identificáveis podem tornar-se estimáveis. Isto é realizado até que todos os parâmetros do modelo tenham sido avaliados quanto a sua identificabilidade. Uma vez que um parâmetro seja incluído no conjunto de parâmetros selecionados, que indica que sua estimação foi bem sucedida, este não será mais removido do conjunto, visto que não será mais avaliado, apenas re-estimado juntamente com demais parâmetros deste conjunto, a cada inclusão bem sucedida de um novo parâmetro. O procedimento numérico proposto para avaliação intensiva tem um menor esforço computacional quando comparado ao procedimento de McLEAN et al. (2012), sendo similar a este apenas no pior caso, que é o cenário em que todos os parâmetros do modelo são avaliados antes de uma inclusão bem sucedida de um parâmetro.

Assim, o novo procedimento numérico faz uma avaliação intensiva dos parâmetros do modelo quanto a sua identificabilidade, permitindo que parâmetros já avaliados sem sucesso possam ser re-avaliados, juntamente com os parâmetros não avaliados, quando valores mais adequados dos parâmetros selecionados são obtidos; isto é particularmente importante em procedimentos que utilizam critérios de identificabilidade baseados em

valores dos parâmetros. Além disso, este novo procedimento permite contornar o problema de seleção de um parâmetro que de acordo com o critério de identificabilidade é estimável, mas que na prática a sua estimação não é viável; ou seja, reduz a dependência com o critério de identificabilidade estabelecido, que frente a modelos complexos pode levar a uma seleção que resulta em um problema de estimação mal condicionado.

5.1.1 Procedimento numérico I: Avaliação intensiva dos parâmetros

O procedimento numérico proposto, apresentado pela Figura 5.1, realiza intensiva avaliação dos parâmetros quanto à identificabilidade, considerando a etapa de estimação simultânea à seleção, de forma que se avaliando os parâmetros um a um, este procedimento apenas pode ser interrompido quando todos os parâmetros tenham sido avaliados. Essencialmente, a avaliação intensiva é composta por três etapas: (i) avaliação dos parâmetros de acordo com o critério de identificabilidade adotado, (ii) seleção do parâmetro melhor avaliado quanto a sua identificabilidade na etapa (i), e (iii) re-estimação dos parâmetros selecionados. Estas etapas são realizadas até que todos os parâmetros sejam avaliados.

De acordo com o procedimento numérico proposto têm-se três conjuntos: (i) parâmetros selecionados $\Theta^{(S)}$, (ii) parâmetros não selecionados $\Theta^{(NS)}$, e (iii) parâmetros avaliados $\Theta^{(E)}$. Inicialmente, todos os parâmetros estão incluídos no conjunto de parâmetros não selecionados $\Theta^{(NS)}$. Então, ordenam-se os parâmetros quanto à identificabilidade, a partir do critério adotado. O parâmetro considerado mais estimável é incluído no conjunto de parâmetros selecionados e removido do conjunto de parâmetros não selecionados. O parâmetro selecionado é estimado juntamente com todos os parâmetros do conjunto $\Theta^{(S)}$. Se a estimação do parâmetro selecionado é bem sucedida, os parâmetros não selecionados são ordenados novamente de acordo com o critério de identificabilidade adotado e o procedimento é repetido. Caso contrário, se a estimação falha ou ocorre algum problema numérico, o último parâmetro incluído deve ser removido do conjunto de parâmetros selecionados e transferido para o conjunto de parâmetros avaliados $\Theta^{(E)}$. O conjunto de parâmetros avaliados $\Theta^{(E)}$ contém os parâmetros que foram incluídos no conjunto de parâmetros selecionados $\Theta^{(S)}$, mas sem sucesso. O procedimento é interrompido quando o conjunto $\Theta^{(NS)}$ estiver vazio, i.e., não há mais parâmetros a serem incluídos no conjunto

$\Theta^{(S)}$. Isto ocorre quando todos os parâmetros já tenham sido avaliados quanto à identificabilidade, com sucesso $\Theta^{(S)}$ ou sem sucesso $\Theta^{(E)}$. Assim, obtém-se uma minuciosa avaliação da identificabilidade dos parâmetros.

Como nos demais procedimentos reportados pela literatura (e.g.: WEIJERS e VANROLLEGHEM, 1997; BRUN et al., 2001; SANDINK et al., 2001; YAO et al., 2003; LI et al., 2004; SECCHI et al., 2006; CHU e HAHN, 2007; SUN e HAHN, 2006; LUND e FOSS, 2008; CHU et al., 2011), o objetivo do procedimento proposto consiste em incluir o próximo parâmetro mais estimável do conjunto $\Theta^{(NS)}$. Todavia, nesta proposta se o parâmetro é incluído com sucesso, então a re-estimação possivelmente modificará os valores dos parâmetros do conjunto $\Theta^{(S)}$. Deste modo, os parâmetros do $\Theta^{(E)}$ podem tornar a ser estimáveis; então, no caso de estimação bem sucedida do conjunto $\Theta^{(S)}$ após a inclusão de um parâmetro, todos os parâmetros do conjunto $\Theta^{(E)}$ são transferidos para o conjunto $\Theta^{(NS)}$ a fim de que estes parâmetros sejam avaliados nas próximas iterações.

Adicionalmente, métodos como a matriz de ganhos relativos (do termo em inglês *Relative Gain Array – RGA*) (BRISTOL, 1966; CAO e ROSSITER, 1997) e a matriz não quadrática de ganhos relativos (do termo em inglês *Non-Square Relative Gain Array – NSRGA*) (CAO e ROSSITER, 1997) podem ser aplicados como critérios para se obter uma avaliação prévia da correlação entre os parâmetros (SANDINK et al., 2001; BOTELHO, 2012). Especialmente em procedimentos de identificabilidade de parâmetros, dada as dimensões da matriz de sensibilidade B , o uso da matriz não quadrática de ganhos relativos *NSRGA* (Equação (E 5.1)) é mais adequado que o uso da matriz de ganhos relativos *RGA* (Equação (E 5.2)); ressalta-se que para matrizes quadradas com bom condicionamento a pseudo-inversa da matriz equivale a sua inversa.

$$(E 5.1) \quad RGA = B \otimes (B^{-1})^T,$$

$$(E 5.2) \quad NSRGA = B \otimes (B^+)^T,$$

em que B^+ denota a pseudo-inversa da matriz de sensibilidade e \otimes é o produto de Hadamard (ou produto elemento a elemento).

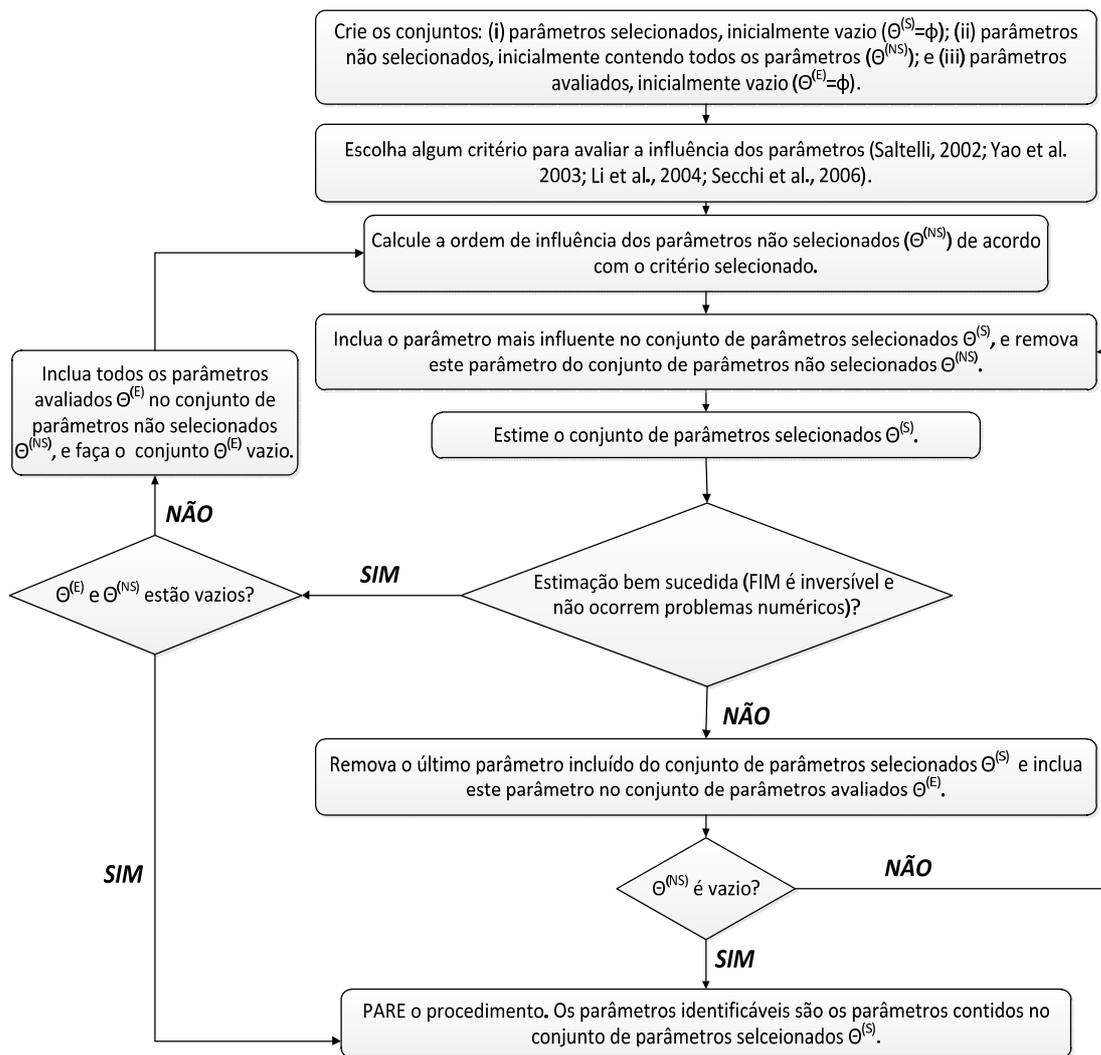


Figura 5.1 - Procedimento numérico proposto para avaliação intensiva da identificabilidade de parâmetros.

A *NSRGA* não deve ser considerada uma análise rigorosa, visto que a correlação entre os parâmetros somente pode ser verificada par a par; além disto, uma vez que esta análise é baseada na matriz B , os resultados são fortemente dependentes da qualidade da estimativa inicial dos parâmetros. Assim, a avaliação *a priori* da correlação entre os parâmetros poder ser uma tarefa árdua.

Para fins didáticos, a Figura 5.2 ilustra a solução de um problema de identificabilidade hipotético contendo 3 parâmetros via o procedimento numérico proposto descrito pela Figura 5.1.

De acordo com a Figura 5.2, os parâmetros são ordenados decrescentemente de acordo com o critério adotado (e.g.: WEIJERS e VANROLLEGHEM, 1997; SANDINK et al., 2001; BRUN et al., 2001; YAO et al., 2003; LI et al., 2004; SECCHI et al., 2006; CHU e HAHN, 2007; SUN e HAHN, 2006; LUND e FOSS, 2008; CHU et al., 2011). O parâmetro mais relevante de acordo com o ordenamento é incluído no conjunto de parâmetros selecionados, representado pelo quadrado preenchido na Figura 5.2. Com a adição deste parâmetro ao conjunto de parâmetros selecionados há duas possibilidades: (i) o conjunto é estimável, (ii) o conjunto é não estimável. Se conjunto de parâmetros selecionados pode ser estimado - o último parâmetro selecionado foi incluído com sucesso, então, os parâmetros avaliados, $\Theta^{(E)}$, são movidos para o conjunto $\Theta^{(NS)}$ e os parâmetros não selecionados são re-ordenados e o mais relevante deve ser incluído no conjunto de parâmetros selecionados. Em contrapartida, se o conjunto de parâmetros selecionados não pode ser estimado - o último parâmetro selecionado não foi incluído com sucesso, então, este parâmetro deve ser removido do conjunto de parâmetros selecionados e adicionado ao conjunto de parâmetros avaliados, neste caso representado pelo quadrado xadrez na Figura 5.2, e o próximo parâmetro de acordo com o ordenamento deve ser selecionado. Uma vez que o número de parâmetros é igual 3, no pior caso o número de avaliações requeridas é 6.

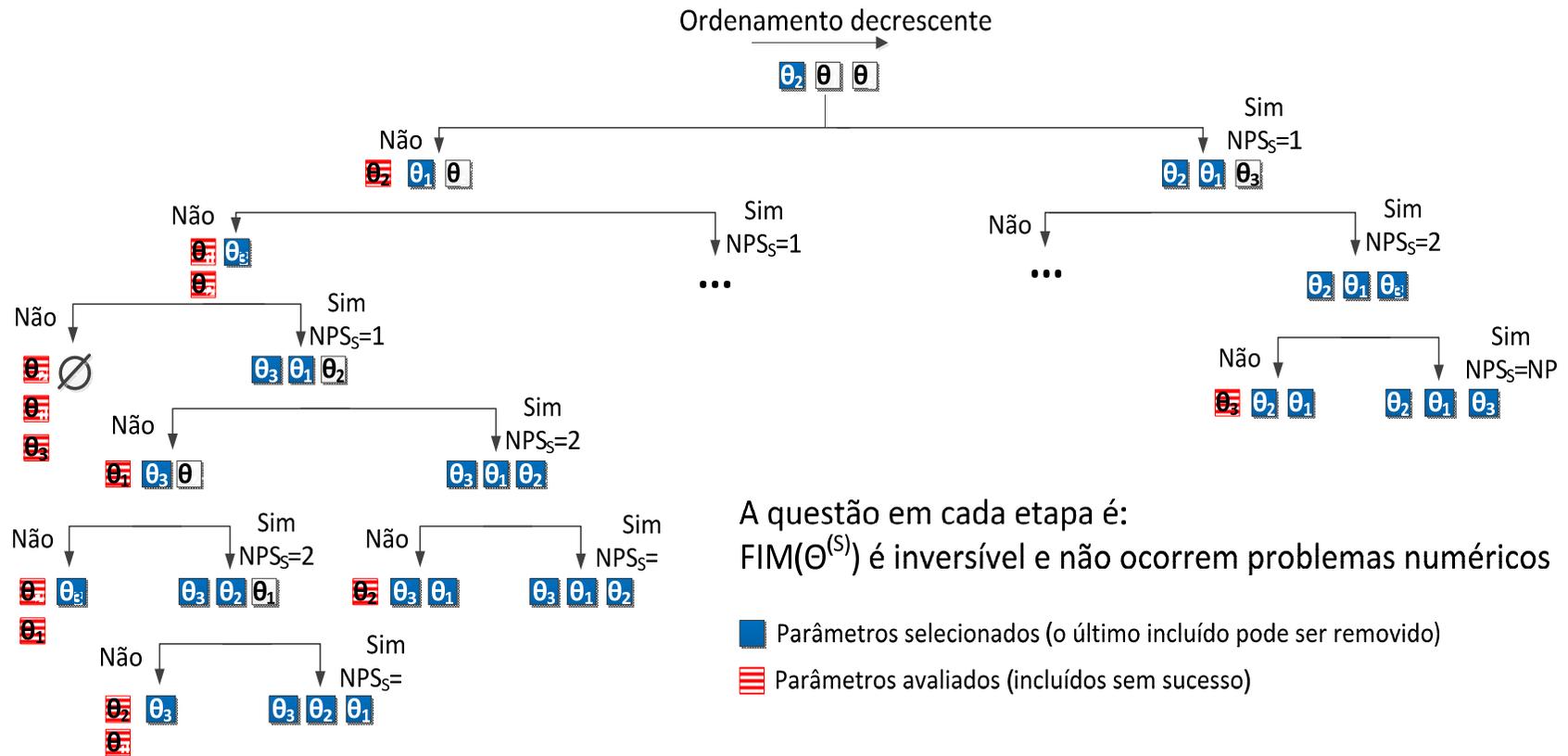


Figura 5.2 - Possíveis combinações empregando a avaliação intensiva (Figura 5.1) para solução de um problema de identificabilidade hipotético contendo 3 parâmetros; NP e NPS_S , respectivamente, representam o número de parâmetros do modelo e o número de parâmetros selecionados com sucesso, sendo \emptyset a representação para o conjunto de parâmetros selecionados vazio.

5.1.2 Estudo de caso: modelo das rotas metabólica da *E. coli* K-12 W3110

A *Escherichia coli* (*E. coli*) é um dos microorganismos mais estudados, sendo amplamente empregado como objeto base para o desenvolvimento de novos métodos de análises de rotas metabólicas, tido como modelo geral para o sistema biológico de bactérias. Embora se tenha um teor considerável de informação a respeito deste microorganismo, usualmente os dados experimentais disponíveis são obtidos dentro de diferentes condições por meio de técnicas heterogêneas, cuja escolha deve estar associada a um fenômeno específico de interesse do metabolismo (CHASSAGNOLLE et al., 2002; HOQUE et al., 2005).

Uma vez que rotas metabólicas são apenas parcialmente observáveis, somente uma fração dos metabólitos intracelulares cruciais para descrever as rotas pode ser diretamente medida. Frequentemente, esta dificuldade experimental compromete a confiabilidade dos mecanismos cinéticos propostos para modelos matemáticos de rotas metabólicas e torna inviável a estimação dos parâmetros. Além disso, esta dificuldade afeta fortemente a modelagem dinâmica, pois em muitos casos se torna necessário a estimação das condições iniciais para muitos metabólitos intracelulares não medidos. Tais problemas requerem o uso de procedimentos de identificabilidade de parâmetros.

Rotas metabólicas em larga escala da glicose, da via das pentoses-fosfato e do sistema fosfotransferase para a *E. coli* K-12 W3110 são apresentadas pela Figura 5.3. A glicose é o substrato mais utilizado experimentalmente, por ser simultaneamente fonte de carbono e energia, ser facilmente metabolizada e ter preço relativamente baixo.

O modelo empregado para representar tais vias do metabolismo da *E. coli* K-12 W3110 é dado por uma complexa estrutura matemática formada por (CHASSAGNOLE et al., 2002; Di MAGGIO et al., 2010): (i) 7 correlações temporais que buscam representar o comportamento de co-metabólitos, como mostra a Tabela 5.1, (ii) 18 equações diferenciais ordinárias que descrevem o balanço de massa para o substrato empregado, a glicose, e os metabólitos intracelulares, como mostra a Tabela 5.2 e, (iii) 30 expressões algébricas que denotam as taxas cinéticas das enzimas, como mostra Tabela 5.3. Nesta estrutura matemática complexa estão presentes 131 parâmetros, descritos pela Tabela 5.4.

O pacote computacional DASSL (PETZOLD, 1982) foi empregado para a integração numérica requerida à resolução deste modelo matemático.

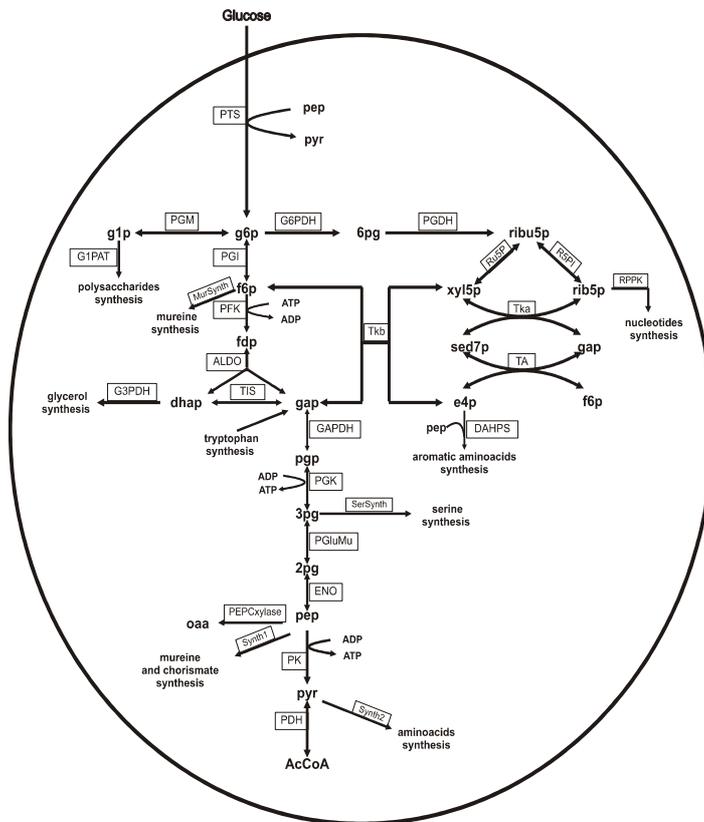


Figura 5.3 – Metabolismo do carbono central da *Escherichia coli* (Di Maggio et al., 2010).

Tabela 5.1 – Correlações temporais para co-metabólitos do microorganismo *E. coli*.

Co-metabólitos	Correlações temporais	Eq.
Adenosina trifosfato: <i>atp</i>	$C_{atp} = 4,27 - 4,163 \frac{t}{0,657 + 1,43t + 0,364t^2}$	(E 5.3)
Adenosina difosfato: <i>adp</i>	$C_{adp} = 0,582 + 1,73 (2,731^{-0,15t}) (0,12t + 2,14 \times 10^{-4} t^3)$	(E 5.4)
Adenosina monofosfato: <i>amp</i>	$C_{amp} = 0,123 + 7,25 \frac{t}{7,25 + 1,47t + 0,17t^2} + 1,073 \frac{t}{1,29 + 8,05t}$	(E 5.5)
Difosfopiridina dinucleotídio-fosfato, reduzido: <i>nadph</i>	$C_{nadph} = 0,062 + 0,332 (2,718^{-0,46t}) \begin{pmatrix} 1,66 \times 10^{-2} t^{1,58} \\ + 1,66 \times 10^{-4} t^{4,73} \\ + 1,16 \times 10^{-10} t^{7,89} \\ + 1,36 \times 10^{-13} t^{11} \\ + 1,23 \times 10^{-16} t^{14,2} \end{pmatrix}$	(E 5.6)
Difosfopiridina dinucleotídio-fosfato, oxidado: <i>nadp</i>	$C_{nadp} = 0,159 + 5,54 \times 10^{-3} \frac{t}{2,8 + 0,271t + 0,01t^2} + 0,182 \frac{t}{4,81 + 0,526t}$	(E 5.7)
Difosfopiridina dinucleotídio, reduzido: <i>nadh</i>	$C_{nadh} = 9,34 \times 10^{-2} + 1,1 \times 10^{-3} (2,371^{-0,123t}) (0,844t + 0,104t^3)$	(E 5.8)
Difosfopiridina dinucleotídio, oxidado: <i>nad</i>	$C_{nad} = 1,314 (1 + 2,73^{(-0,0435t - 0,342)}) - \frac{(t + 7,871) (2,73^{(-0,0218t - 0,171)})}{8,481 + t}$	(E 5.9)

Tabela 5.2 – Balanço de massa para a glicose e os metabólitos do microorganismo *E. coli*.

Substrato	Eq.
Glicose: <i>glc</i>	$\frac{dC_{glc}^{extracellular}}{dt} = D(C_{glc}^{feed} - C_{glc}^{extracellular}) + f_{pulse} - \frac{C_x r_{PTS}}{\rho_x}$ (E 5.10)
Metabólitos	Eq.
Glicose-6-fosfato: <i>g6p</i>	$\frac{dC_{g6p}}{dt} = r_{PTS} - r_{PGI} - r_{G6PDH} - r_{PGM} - \mu C_{g6p}$ (E 5.11)
Frutose-6-fosfato: <i>f6p</i>	$\frac{dC_{f6p}}{dt} = r_{PGI} - r_{PFK} + r_{TKb} + r_{TA} - 2r_{MurSynth} - \mu C_{f6p}$ (E 5.12)
Frutose-1,6-difosfato: <i>fdp</i>	$\frac{dC_{fdp}}{dt} = r_{PFK} - r_{ALDO} - \mu C_{fdp}$ (E 5.13)
Gliceraldeído-3-fosfato: <i>gap</i>	$\frac{dC_{gap}}{dt} = r_{ALDO} + r_{TIS} - r_{GAPDH} + r_{TKa} + r_{TKb} - r_{TA} + r_{TrpSynth} - \mu C_{gap}$ (E 5.14)
Fosfato di-hidroxiacetona: <i>dhap</i>	$\frac{dC_{dhap}}{dt} = r_{ALDO} - r_{TIS} - r_{G3PDH} - \mu C_{dhap}$ (E 5.15)
1,3-difosfoglicerato: <i>pgp</i>	$\frac{dC_{pgp}}{dt} = r_{GAPDH} - r_{PGK} - \mu C_{pgp}$ (E 5.16)
3-fosfoglicerato: <i>3pg</i>	$\frac{dC_{3pg}}{dt} = r_{PGK} - r_{PGGluMu} - r_{SerSynth} - \mu C_{3pg}$ (E 5.17)
2-fosfoglicerato: <i>2pg</i>	$\frac{dC_{2pg}}{dt} = r_{PGluMu} - r_{ENO} - \mu C_{2pg}$ (E 5.18)
Fosfoenolpiruvato: <i>pep</i>	$\frac{dC_{pep}}{dt} = r_{ENO} - r_{PK} - r_{PTS} - r_{PEPCxylase} - r_{DAHPS} - r_{Synth1} - \mu C_{pep}$ (E 5.19)
Piruvato: <i>pyr</i>	$\frac{dC_{pyr}}{dt} = r_{PK} + r_{PTS} - r_{PDH} - r_{Synth2} + r_{MetSynth} + r_{TrpSynth} - \mu C_{pyr}$ (E 5.20)
6-fosfogluconato: <i>6pg</i>	$\frac{dC_{6pg}}{dt} = r_{G6PDH} - r_{PGDH} - \mu C_{6pg}$ (E 5.21)
Ribulose-5-fosfato: <i>ribu5p</i>	$\frac{dC_{ribu5p}}{dt} = r_{PGDH} - r_{Ru5P} - r_{R5PI} - \mu C_{ribu5p}$ (E 5.22)
Xilulose-5-fosfato: <i>xyl5p</i>	$\frac{dC_{xyl5p}}{dt} = r_{Ru5P} - r_{TKa} - r_{TKb} - \mu C_{xyl5p}$ (E 5.23)
Sedoeptulose-7-fosfato: <i>sed7p</i>	$\frac{dC_{sed7p}}{dt} = r_{TKa} - r_{TA} - \mu C_{sed7p}$ (E 5.24)
Ribose-5-fosfato: <i>rib5p</i>	$\frac{dC_{rib5p}}{dt} = r_{R5PI} - r_{TKa} - r_{RPPK} - \mu C_{rib5p}$ (E 5.25)
Eritrose-4-fosfato: <i>e4p</i>	$\frac{dC_{e4p}}{dt} = r_{TA} - r_{TKb} - r_{DAHPS} - \mu C_{e4p}$ (E 5.26)
Glicose-1-fosfato: <i>glp</i>	$\frac{dC_{glp}}{dt} = r_{PGM} - r_{GIPAT} - \mu C_{glp}$ (E 5.27)

Tabela 5.3 - Taxas cinéticas de enzimas (CHASSAGNOLE et al., 2002).

Taxas cinéticas	Eq.
$r_{PGI} = \frac{r_{PGI}^{\max} \left(C_{g6p} - \frac{C_{f6p}}{K_{PGI,eq}} \right)}{K_{PGI,g6p} \left[I + \frac{C_{f6p}}{K_{PGI,f6p} \left(I + \frac{C_{6pg}}{K_{PGI,f6p,6pg,inh}} \right)} + \frac{C_{6pg}}{K_{PGI,g6p,6pg,inh}} \right] + C_{g6p}}$	(E 5.28)
$r_{PFK}^{\max} \left[\frac{eC_{f6p}}{K_{Rf6p}} \left(I + \frac{eC_{f6p}}{K_{Rf6p}} \right)^{n_{PFK}-1} \left(I + \frac{C_{adp}}{K_{Radp}} \right)^{n_{PFK}} + L\theta ce' \frac{C_{f6p}}{K_{Rf6p}} \left(I + ce' \frac{C_{f6p}}{K_{Rf6p}} \right)^{n_{PFK}-1} \right]$	(E 5.29)
$r_{PFK} = \frac{\left(I + \frac{eC_{f6p}}{K_{Rf6p}} \right)^{n_{PFK}} \left(I + \frac{C_{adp}}{K_{Radp}} \right)^{n_{PFK}} + L \left(I + ce' \frac{C_{f6p}}{K_{Rf6p}} \right)^{n_{PFK}-1} \frac{C_{pep}}{K_{Tpep}}}{\left(I + \frac{eC_{f6p}}{K_{Rf6p}} \right)^{n_{PFK}} \left(I + \frac{C_{adp}}{K_{Radp}} \right)^{n_{PFK}} + L \left(I + ce' \frac{C_{f6p}}{K_{Rf6p}} \right)^{n_{PFK}-1} \frac{C_{pep}}{K_{Tpep}}}$	(E 5.30)
$r_{ALDO} = \frac{r_{ALDO}^{\max} \left(C_{fdp} - \frac{C_{gap} C_{dhap}}{K_{ALDO,eq}} \right)}{\left(K_{ALDO,fdp} + C_{fdp} + \frac{K_{ALDO,gap} C_{dhap}}{K_{ALDO,eq} V_{ALDO,bif}} + \frac{K_{ALDO,dhap} C_{gap}}{K_{ALDO,eq} V_{ALDO,bif}} \right) + \frac{C_{fdp} C_{gap}}{K_{ALDO,gap,inh}} + \frac{C_{dhap} C_{gap}}{K_{ALDO,eq} V_{ALDO,bif}}}$	(E 5.31)
$r_{TIS} = \frac{r_{TIS}^{\max} \left(C_{dhap} - \frac{C_{gap}}{K_{TIS,eq}} \right)}{K_{TIS,dhap} \left(1 + \frac{C_{gap}}{K_{TIS,gap}} \right) + C_{dhap}}$	(E 5.32)
$r_{GAPDH} = \frac{r_{GAPDH}^{\max} \left(C_{gap} C_{nad} - \frac{C_{pgp} C_{nadh}}{K_{GAPDH,eq}} \right)}{\left(K_{GAPDH,gap} \left(1 + \frac{C_{pgp}}{K_{GAPDH,pgp}} \right) + C_{gap} \right) \left(K_{GAPDH,nad} \left(1 + \frac{C_{nadh}}{K_{GAPDH,nadh}} \right) + C_{nad} \right)}$	(E 5.33)
$r_{PGK} = \frac{r_{PGK}^{\max} \left(C_{adp} C_{pgp} - \frac{C_{atp} C_{3pg}}{K_{PGK,eq}} \right)}{\left(K_{PGK,adp} \left(1 + \frac{C_{atp}}{K_{PGK,atp}} \right) + C_{adp} \right) \left(K_{PGK,pgp} \left(1 + \frac{C_{3pg}}{K_{PGK,3pg}} \right) + C_{pgp} \right)}$	(E 5.34)
$r_{PGluMu} = \frac{r_{PGluMu}^{\max} \left(C_{3pg} - \frac{C_{2pg}}{K_{PGluMu,eq}} \right)}{K_{PGluMu,3pg} \left(1 + \frac{C_{2pg}}{K_{PGluMu,2pg}} \right) + C_{3pg}}$	(E 5.35)

$$r_{ENO} = \frac{r_{ENO}^{\max} \left(C_{2pg} - \frac{C_{pep}}{K_{ENO,eq}} \right)}{K_{ENO,2pg} \left(1 + \frac{C_{pep}}{K_{ENO,pep}} \right) + C_{2pg}} \quad (E 5.36)$$

$$r_{PK} = \frac{r_{PK}^{\max} C_{pep} C_{adp} \left(\frac{C_{pep}}{K_{PK,pep}} + 1 \right)^{(n_{PK}-1)}}{K_{PK,pep} \left[L_{PK} \left(\frac{1 + \frac{C_{atp}}{K_{PK,atp}}}{\frac{C_{fdp}}{K_{PK,fdp}} + \frac{C_{amp}}{K_{PK,amp}} + 1} \right)^{n_{PK}} + \left(\frac{C_{pep}}{K_{PK,pep}} + 1 \right)^{n_{PK}} \right] (C_{adp} + K_{PK,adp})} \quad (E 5.37)$$

$$r_{G6PDH} = \frac{r_{G6PDH}^{\max} C_{g6p} \left(\frac{C_{nadp}}{K_{G6PDH,nadp}} \right)^{1+k_{dn} \frac{C_{nadh}^{hd}}{k_{di}^{hd} + C_{nadh}^{hd}}}}{\left(K_{G6PDH,g6p} \left(I + k_{ign} \frac{C_{nadp}^{hg}}{k_{ig}^{hg} + C_{nadp}^{hg}} \right) + C_{g6p} \right) \left(I + \frac{C_{nadph}}{K_{G6PDH,nadphinh}} + \frac{C_{nadh}}{K_{G6PDH,nadphinh}} \right) D} \quad (E 5.38)$$

$$D = \left(I + \left(\frac{C_{nadp}}{K_{G6PDH,nadp}} \right)^{1+k_{dn} \frac{C_{nadh}^{hd}}{k_{di}^{hd} + C_{nadh}^{hd}}} + \left(\frac{C_{nadph}}{K_{G6PDH,nadphinh2}} \right)^{h_{nadph}} \right) \left(I + \left(\frac{C_{nadh}}{K_{G6PDH,nadh}} \right)^{h_{nadh}} \right) \quad (E 5.39)$$

$$r_{PDH} = \frac{r_{PDH}^{\max} C_{pyr}^{n_{PDH}}}{K_{PDH,pyr} + C_{pyr}^{n_{PDH}}} \quad (E 5.40)$$

$$r_{PGDH} = \frac{r_{PGDH}^{\max} C_{6pg} C_{nadp}}{(C_{6pg} + K_{PGDH,6pg}) \left(C_{nadp} + K_{PGDH,nadp} \left(1 + \frac{C_{nadph}}{K_{PGDH,nadph,inh}} \right) \left(1 + \frac{C_{atp}}{K_{PGDH,atp,inh}} \right) \right)} \quad (E 5.41)$$

$$r_{R5PI} = r_{R5PI}^{\max} \left(C_{rib5p} - \frac{C_{rib5p}}{K_{R5PI,eq}} \right) \quad (E 5.42)$$

$$r_{Ru5p} = r_{Ru5p}^{\max} \left(C_{rib5p} - \frac{C_{xyl5p}}{K_{Ru5p,eq}} \right) \quad (E 5.43)$$

$$r_{TKa} = r_{TKa}^{\max} \left(C_{rib5p} C_{xyl5p} - \frac{C_{sed7p} C_{gap}}{K_{TKa,eq}} \right) \quad (E 5.44)$$

$$r_{TKb} = r_{TKb}^{\max} \left(C_{xyl5p} C_{e4p} - \frac{C_{f6p} C_{gap}}{K_{TKb,eq}} \right) \quad (E 5.45)$$

$$r_{TA} = r_{TA}^{\max} \left(C_{gap} C_{sed7p} - \frac{C_{e4p} C_{f6p}}{K_{TA,eq}} \right) \quad (E 5.46)$$

$$r_{PEPCxylase} = \frac{r_{PEPCxylase}^{\max} C_{pep} \left(1 + \left(\frac{C_{fdp}}{K_{PEPCxylase,fdp}} \right)^{n_{PEPCxylase,fdp}} \right)}{K_{PEPCxylase,pep} + C_{pep}} \quad (E 5.47)$$

$$r_{Synthesis1} = \frac{r_{Synthesis1}^{\max} C_{pep}}{K_{Synthesis1,pep} + C_{pep}} \quad (E 5.48)$$

$$r_{Synthesis2} = \frac{r_{Synthesis2}^{max} C_{pyr}}{K_{Synthesis2,pyr} + C_{pyr}} \quad (E 5.49)$$

$$r_{SerSynthesis} = \frac{r_{SerSynthesis}^{max} C_{3pg}}{K_{SerSynthesis,3pg} + C_{3pg}} \quad (E 5.50)$$

$$r_{RPPK} = \frac{r_{RPPK}^{max} C_{rib5p}}{K_{RPPK,rib5p} + C_{rib5p}} \quad (E 5.51)$$

$$r_{G3PDH} = \frac{r_{G3PDH}^{max} C_{dhap}}{K_{G3PDH,dhap} + C_{dhap}} \quad (E 5.52)$$

$$r_{DAHPS} = \frac{r_{DAHPS}^{max} C_{e4p}^{n_{DAHPS,e4p}} C_{pep}^{n_{DAHPS,pep}}}{\left(K_{DAHPS,e4p} + C_{e4p}^{n_{DAHPS,e4p}} \right) \left(K_{DAHPS,pep} + C_{pep}^{n_{DAHPS,pep}} \right)} \quad (E 5.53)$$

$$r_{PGM} = \frac{r_{PGM}^{max} \left(C_{g6p} - \frac{C_{g1p}}{K_{PGM,eq}} \right)}{K_{PGM,g6p} \left(1 + \frac{C_{g1p}}{K_{PGM,g1p}} \right) + C_{g6p}} \quad (E 5.54)$$

$$r_{G1PAT} = \frac{r_{G1PAT}^{max} C_{g1p} C_{atp} \left(1 + \left(\frac{C_{fdp}}{K_{G1PAT,fdp}} \right)^{n_{G1PAT,fdp}} \right)}{\left(K_{G1PAT,g1p} + C_{g1p} \right) \left(K_{G1PAT,atp} + C_{atp} \right)} \quad (E 5.55)$$

$$r_{MurSynthesis} = r_{MurSynthesis}^{max} \quad (E 5.56)$$

$$r_{TrpSynthesis} = r_{TrpSynthesis}^{max} \quad (E 5.57)$$

$$r_{MetSynthesis} = r_{MetSynthesis}^{max} \quad (E 5.58)$$

Tabela 5.4 - Descrição dos parâmetros presentes nas taxas cinéticas de enzimas (CHASSAGNOLE et al., 2002; Di MAGGIO et al., 2010).

Enzimas	Parâmetros	Descrição
Sistema fosfotransferase: PTS	K_{PTS1}	Constante de meia saturação (mM)
	K_{PTS2}	Constante (mM)
	K_{PTS3}	Constante
	$K_{PTS,g6p}$	Constante de inibição (mM)
	$n_{PTS,g6p}$	Constante
	r_{PTS}^{max}	Taxa máxima de reação (mM s ⁻¹)
Fosfoglucoisomerase: PGI	$K_{PGI,g6p}$	Constante de meia saturação (mM)
	$K_{PGI,f6p}$	Constante de inibição (mM)
	$K_{PGI,eq}$	Constante de equilíbrio
	$K_{PGI,g6p,6pg,inh}$	Constante de inibição (mM)
	$K_{PGI,f6p,6pg,inh}$	Constante de inibição (mM)
	r_{PGI}^{max}	Taxa máxima de reação (mM s ⁻¹)
Fosfofrutoquinase: PFK	$K_{PFK,f6p,s}$	Constante de meia saturação (mM)
	$K_{PFK,atp,s}$	Constante de meia saturação (mM)
	$K_{PFK,adp,a}$	Constante de ativação (mM)
	$K_{PFK,adp,b}$	Constante de ativação (mM)
	$K_{PFK,adp,c}$	Constante de ativação (mM)

	$K_{PFK,amp,a}$	Constante de ativação (mM)
	$K_{PFK,amp,b}$	Constante de ativação (mM)
	$K_{PFK,pep}$	Constante de inibição (mM)
	L_{PFK}	Constante alostérica
	n_{PFK}	Número de sítios ligantes
	r_{PFK}^{max}	Taxa máxima de reação (mM s ⁻¹)
Aldolase: ALDO	$K_{ALDO,fdp}$	Constante de meia saturação (mM)
	$K_{ALDO,dhap}$	Constante de meia saturação (mM)
	$K_{ALDO,gap}$	Constante de meia saturação (mM)
	$K_{ALDO,gap,inh}$	Constante de inibição (mM)
	$V_{ALDO,blf}$	Constante de reação de retorno
	$K_{ALDO,eq}$	Constante de equilíbrio (mM)
	r_{ALDO}^{max}	Taxa máxima de reação (mM s ⁻¹)
Triosephosphate isomerase: TIS	$K_{TIS,dhap}$	Constante de meia saturação (mM)
	$K_{TIS,gap}$	Constante de meia saturação (mM)
	$K_{TIS,eq}$	Constante de equilíbrio
	r_{TIS}^{max}	Taxa máxima de reação (mM s ⁻¹)
Gliceraldeído-3-fosfato desidrogenase: GAPDH	$K_{GAPDH,gap}$	Constante de meia saturação (mM)
	$K_{GAPDH,pgp}$	Constante de inibição (mM)
	$K_{GAPDH,nad}$	Constante de meia saturação (mM)
	$K_{GAPDH,nadh}$	Constante de inibição (mM)
	$K_{GAPDH,eq}$	Constante de equilíbrio
	r_{GAPDH}^{max}	Taxa máxima de reação (mM s ⁻¹)
Fosfoglicerato quinase: PGK	$K_{PGK,pgp}$	Constante de meia saturação (mM)
	$K_{PGK,3pg}$	Constante de inibição (mM)
	$K_{PGK,adp}$	Constante de meia saturação (mM)
	$K_{PGK,atp}$	Constante de inibição (mM)
	$K_{PGK,eq}$	Constante de equilíbrio
	r_{PGK}^{max}	Taxa máxima de reação (mM s ⁻¹)
Fosfoglicerato mutase: PGluMu	$K_{PGluMu,3pg}$	Constante de meia saturação (mM)
	$K_{PGluMu,2pg}$	Constante de inibição (mM)
	$K_{PGluMu,eq}$	Constante de equilíbrio
	r_{PGluMu}^{max}	Taxa máxima de reação (mM s ⁻¹)
Enolase: ENO	$K_{ENO,2pg}$	Constante de meia saturação (mM)
	$K_{ENO,pep}$	Constante de inibição (mM)
	$K_{ENO,eq}$	Constante de equilíbrio
	r_{ENO}^{max}	Taxa máxima de reação (mM s ⁻¹)
Piruvato quinase: PK	$K_{PK,pep}$	Constante de meia saturação (mM)
	$K_{PK,adp}$	Constante de meia saturação (mM)
	$K_{PK,atp}$	Constante de inibição (mM)
	$K_{PK,fdp}$	Constante de ativação (mM)
	$K_{PK,amp}$	Constante de ativação (mM)
	L_{PK}	Constante alostérica
	n_{PK}	Número de sítios ligantes
	r_{PK}^{max}	Taxa máxima de reação (mM s ⁻¹)

Piruvato desidrogenase: <i>PDH</i>	$K_{PDH,pyr}$	Constante de meia saturação (mM)
	n_{PDH}	Número de sítios ligantes
	r_{PDH}^{max}	Taxa máxima de reação (mM s ⁻¹)
Fosfoenolpiruvato carboxilase: <i>PEPCxylase</i>	$K_{PEPCxylase,pep}$	Constante de meia saturação (mM)
	$K_{PEPCxylase,fdp}$	Constante de ativação (mM)
	$n_{PEPCxylase,fdp}$	Número de sítios ligantes
	$r_{PEPCxylase}^{max}$	Máxima taxa de reação (mM s ⁻¹)
Fosfoglicomutase: <i>PGM</i>	$K_{PGM,g6p}$	Constante de meia saturação (mM)
	$K_{PGM,g1p}$	Constante de inibição (mM)
	$K_{PGM,eq}$	Constante de equilíbrio
	r_{PGM}^{max}	Taxa máxima de reação (mM s ⁻¹)
Glicose-1-fosfato adeniltransferase: <i>G1PAT</i>	$K_{G1PAT,g1p}$	Constante de meia saturação (mM)
	$K_{G1PAT,atp}$	Constante de meia saturação (mM)
	$K_{G1PAT,fdp}$	Constante de ativação (mM)
	$n_{G1PAT,fdp}$	Número de sítios ligantes
	r_{G1PAT}^{max}	Taxa máxima de reação (mM s ⁻¹)
Ribose fosfato pirofosfoquinase: <i>RPPK</i>	$K_{RPPK,rib5p}$	M–M half-saturation constant (mM)
	r_{RPPK}^{max}	Taxa máxima de reação (mM s ⁻¹)
Glicerol-3-fosfato desidrogenase: <i>G3PDH</i>	$K_{G3PDH,dhap}$	Constante de meia saturação (mM)
	r_{G3PDH}^{max}	Taxa máxima de reação (mM s ⁻¹)
Síntese serina: <i>SerSynth</i>	$K_{SerSynth,3pg}$	Constante de meia saturação (mM)
	$r_{SerSynth}^{max}$	Taxa máxima de reação (mM s ⁻¹)
Síntese 3-deoxy-D-arabino-heptulosonato-7-fosfato: <i>DAHPS</i>	$K_{DAHPS,e4p}$	Constante de meia saturação (mM)
	$K_{DAHPS,pep}$	Constante de meia saturação (mM)
	$n_{DAHPS,e4p}$	Número de sítios ligantes
	$n_{DAHPS,pep}$	Número de sítios ligantes
	r_{DAHPS}^{max}	Taxa máxima de reação (mM s ⁻¹)
Glicose-6-fosfato desidrogenase: <i>G6PDH</i>	$K_{G6PDH,g6p}$	Constante de meia saturação (mM)
	$K_{G6PDH,nadp}$	Constante de meia saturação (mM)
	$K_{G6PDH,nadph,nadph,inh}$	Constante de inibição (mM)
	$K_{G6PDH,nadph,g6ph,inh}$	Constante de inibição (mM)
	r_{G6PDH}^{max}	Taxa máxima de reação (mM s ⁻¹)
6-Fosfogluconato desidrogenase: <i>PGDH</i>	$K_{PGDH,6pg}$	Constante de meia saturação (mM)
	$K_{PGDH,nadp}$	Constante de meia saturação (mM)
	$K_{PGDH,nadph,inh}$	Constante de inibição (mM)
	$K_{PGDH,atp,inh}$	Constante de inibição (mM)
	r_{PGDH}^{max}	Taxa máxima de reação (mM s ⁻¹)
Ribulose fosfato epimerase: <i>RU5P</i>	$K_{RU5P,EQ}$	Constante de equilíbrio (mM)
	r_{RU5P}^{max}	Taxa máxima de reação (mM s ⁻¹)
Ribose fosfato isomerase: <i>R5PI</i>	$K_{R5PI,eq}$	Constante de equilíbrio (mM)
	r_{R5PI}^{max}	Taxa máxima de reação (mM s ⁻¹)
Transquetolase a: <i>TKa</i>	$K_{TKa,eq}$	Constante de equilíbrio (mM)

	r_{TKa}^{\max}	Taxa máxima de reação (mM s^{-1})
Transquetolase b: <i>TKb</i>	$K_{TKb,eq}$	Constante de equilíbrio (mM)
	r_{TKb}^{\max}	Taxa máxima de reação (mM s^{-1})
Transaldolase: <i>TA</i>	$K_{TA,eq}$	Constante de equilíbrio (mM)
	r_{TA}^{\max}	Taxa máxima de reação (mM s^{-1})
Síntese 1: <i>Synth1</i>	$K_{Synth1,pep}$	Constante de meia saturação (mM)
	r_{Synth1}^{\max}	Taxa máxima de reação (mM s^{-1})
Síntese 2: <i>Synth2</i>	$K_{Synth2,pyr}$	Constante de meia saturação (mM)
	r_{Synth2}^{\max}	Taxa máxima de reação (mM s^{-1})
Síntese mureína: <i>MurSynth</i>	$r_{MurSynth}^{\max}$	Taxa máxima de reação (mM s^{-1})
Síntese triptofano: <i>TrpSynth</i>	$r_{TrpSynth}^{\max}$	Taxa máxima de reação (mM s^{-1})
Síntese metionina: <i>MetSynth</i>	$r_{MetSynth}^{\max}$	Taxa máxima de reação (mM s^{-1})

As estimativas iniciais para os valores dos parâmetros do modelo matemático do metabolismo da *E. coli* K-12 W3110 foram obtidos a partir da literatura (Di MAGGIO et al., 2010). Assim como a informação experimental utilizada, obtida em HOQUE et al. (2005), empregando-se um horizonte de tempo de 300 segundos. Uma vez que as rotas metabólicas são experimentalmente somente parcialmente observáveis, os dados experimentais não estão disponíveis para todos os metabólitos, de modo que apenas os seguintes perfis temporais para os seguintes metabólitos foram empregados: glicose (*glc*), piruvato (*pyr*), fosfato de di-hidroxiacetona (*dhap*), eritrose-4-fosfato (*e4p*), frutose-1,6-bifosfato (*fdp*), ribose-5-fosfato (*rib5p*), ribulose-5-fosfato (*ribu5p*), 2-fosfoglicerato (*2pg*), fosfoenolpiruvato (*pep*) e 6-fosfogluconato (*6pg*). Os demais metabólitos intracelulares não foram medidos e então suas condições iniciais não são conhecidas; uma vez que as condições iniciais são informações imprescindíveis para os modelos matemáticos de rotas metabólicas, seus valores devem ser estimados junto com os parâmetros.

A incerteza experimental foi admitida proporcional aos valores medidos da variável dependente (y_i) na forma:

$$(E 5.59) \quad \sigma_{y_i} = a \cdot y_i + b$$

em que a e b podem ser arbitrariamente escolhidos ou obtidos experimentalmente ou mesmo estimados. Embora esta informação seja de suma importância para o procedimento de estimação e identificabilidade de parâmetros, como demonstrado no Capítulo 4, a caracterização da incerteza experimental requer replicações dos experimentos, o que é um problema significativo para a maioria dos sistemas investigados pelas engenharias química e ambiental. Nesta tese, foram admitidos $a = 10^{-2}$ e $b = 10^{-6}$ (particularmente o valor de b foi definido de acordo com a magnitude dos valores das medidas experimentais).

A fim de obter uma avaliação preliminar a respeito da interação entre os efeitos dos parâmetros sobre a predição do modelo matemático da *E. coli* K-12 W3110, aplicou-se a *NSRGA* e também foram calculados os ângulos de correlação entre os vetores de sensibilidade de todos os parâmetros deste modelo. A Figura 5.4 apresenta os resultados para os ângulos de correlação entre os vetores de sensibilidade dos parâmetros, demonstrando que para a informação experimental disponível a grande maioria dos parâmetros apresenta fraca correlação ou não estão correlacionados. Tais resultados concordam com os resultados obtidos via *NSRGA*.

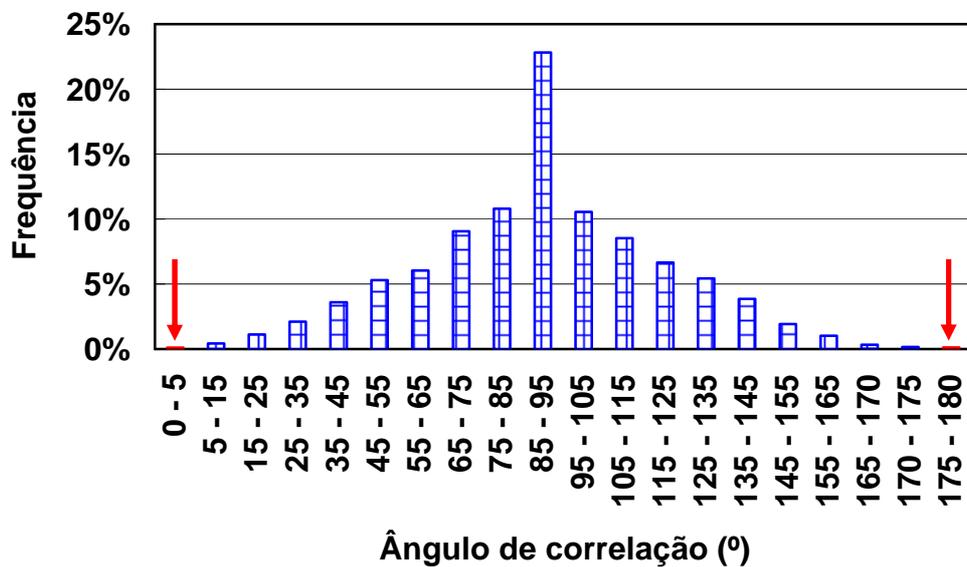


Figura 5.4 - Frequência de ocorrência dos parâmetros do modelo matemático das rotas metabólicas da *E. coli* K-12 W3110 em faixas de valores do ângulo de correlação.

Na Figura 5.4 as setas assinalam a baixa frequência dos parâmetros na faixa de ângulo que apresentam forte correlação, 0° e 5° e entre 175° e 180° , sendo estes parâmetros

apresentados pela Tabela 5.5. Embora o ângulo de correlação entre os parâmetros seja uma informação útil que permite identificar previamente a correlação entre os parâmetros, os resultados desta análise são fortemente dependentes da estimativa inicial dos parâmetros. Desta forma, com a etapa de re-estimação realizada a cada seleção, os valores dos parâmetros podem modificar-se e conseqüentemente modificam-se também os valores do ângulo de correlação entre os parâmetros. Além disso, a análise permite a avaliação dos parâmetros apenas par a par. O mesmo ocorre para a análise *NSRGA*. Assim, os resultados destas análises devem ser utilizados com cautela, sendo nesta etapa reservados à compreensão do cenário inicial a que foi empregado o procedimento de identificabilidade.

Tabela 5.5 – Parâmetros com forte correlação no modelo matemático das rotas metabólicas da *E. coli* K-12 W3110: ângulos de correlação com valores entre 0° e 5° e entre 175° e 180°.

Pares de parâmetros	Ângulo de correlação (°)
$K_{PTS,1} - K_{PTS,3}$	0,5
$K_{ALDO,eq} - r_{TIS}^{max}$	175,3
$K_{GAPDH,nad} - K_{PGK,eq}$	176,5
$K_{GAPDH,eq} - L_{PK}$	179,5
$K_{PK,amp} - r_{PGM}^{max}$	176,5
$K_{GIPAT,glp} - K_{GIPAT,atp}$	179,3
$K_{PK,amp} - r_{Ser,Synth}^{max}$	2,5
$K_{PK,fdp} - K_{Synth1,pep}$	0,0
$n_{DAHPS,pep} - K_{ALDO,fdp}$	179,3
$K_{GAPDH,eq} - K_{G6PDH,tgn}$	2,5
$L_{PK} - K_{G6PDH,tgn}$	177,4
$K_{PGDH,atp} - K_{G1PAT,glp}$	2,4
$K_{PGDH,atp} - K_{G1PAT,atp}$	1,6
$K_{PGDH,atp} - K_{PGDH,nadp}$	177,5
$K_{TA,eq} - K_{G1PAT,glp}$	2,3
$K_{TA,eq} - K_{G1PAT,atp}$	1,5
$K_{TA,eq} - K_{PGDH,atp}$	0,7
$r_{PTS}^{max} - K_{G6PDH,dt}$	175,2
$r_{PGDH}^{max} - K_{GAPDH,nad}$	177,6

Os resultados numéricos são apresentados pelas Figura 5.5 e Tabela 5.6, comparando a simulação com as estimativas iniciais da literatura (DI MAGGIO et al., 2010) e os valores re-estimados dos parâmetros de acordo com o procedimento numérico proposto para a avaliação intensiva (Figura 5.1). Foram identificados 58 parâmetros, juntamente com 5 condições iniciais para a concentração de metabólitos intracelulares não medidos ($pgp=1,215\times 10^{-5}$, $3pg=1,911$, $xyl5p=1,564$, $sed7p=4,085\times 10^{-2}$ e $glp=6,500\times 10^{-2}$).

Em modelos matemáticos de rotas metabólicas, a condição inicial para a concentração de metabólitos intracelulares é uma informação essencial. A ausência desta informação foi um dos grandes desafios de se empregar a avaliação intensiva no modelo matemático da *E. coli* K-12 W3110, impossibilitando o uso de outras fontes de dados experimentais presentes na literatura, para as quais havia uma grande quantidade de metabólitos não medidos presentes. A partir do observado ao longo do estudo da identificabilidade do modelo matemático da *E. coli* K-12 W3110, pode-se dizer que é altamente não recomendável manter o valor das condições iniciais em zero, visto que pode levar a problemas numéricos na integração ou os valores preditos podem ser totalmente inconsistentes com a realidade experimental. Desta forma, como mencionado anteriormente, recomenda-se a estimação da concentração inicial dos metabólitos não medidos, junto com os parâmetros.

Nesta tese adotou-se a seguinte estratégia para se obter estimativas iniciais para as concentrações iniciais de metabólitos cruciais:

- (i). Inicialmente fixe em zero o valor da concentração inicial dos metabólitos não medidos e simule o modelo até o próximo ponto de integração; note que o passo de integração deve ser pequeno para este caso.
- (ii). Tome o valor calculado para a concentração de metabólitos não medidos e adicione como condições iniciais para estes metabólitos; por rigor, o tempo inicial para simulação do modelo deve ser o tempo no qual tais condições iniciais foram obtidas.
- (iii). Este dado é fornecido como estimativa inicial da concentração inicial dos metabólitos não medidos, sendo que tais concentrações podem ser melhoradas ao longo do procedimento de identificabilidade.

Embora não seja rigorosa, a estratégia para se obter a concentração inicial dos metabólitos intracelulares pode fornecer uma medida da ordem de magnitude da variação inicial dos metabólitos intracelulares.

A Figura 5.5 demonstra a melhoria da predição do modelo pelo uso do procedimento proposto de avaliação intensiva. Uma vez que o modelo ajustou-se razoavelmente aos dados experimentais, pode-se concluir que a seleção de parâmetros empregando o procedimento para avaliação intensiva dos parâmetros (Figura 5.1) foi realizada com considerável sucesso e que seus resultados apresentam qualidade superior ao uso do mesmo procedimento sem considerar a estimação simultânea. Atenção especial deve ser dada às Figura 5.5B, Figura 5.5D, Figura 5.5E, Figura 5.5F e Figura 5.5I para as quais o comportamento do modelo matemático empregando as estimativas iniciais não é capaz de acompanhar a tendência dos dados experimentais; para estes casos, as estimativas iniciais dos parâmetros demonstram ser inadequadas.

A Tabela 5.6 apresenta os valores de parâmetros estimados juntamente com as estimativas iniciais e o erro normalizado destes parâmetros (σ_{θ}/θ); o erro normalizado é calculado como a razão entre o desvio padrão e o valor estimado do parâmetro. Como informação importante para a qualidade dos parâmetros identificáveis, o baixo valor obtido para o erro normalizado dos parâmetros indica que tais parâmetros foram estimados com sucesso. Analisando os valores dos parâmetros estimados, verifica-se que a estimativa inicial pôde ser significativamente melhorada. É importante enfatizar que os valores dos parâmetros selecionados podem ser significativamente influenciados pelas estimativas iniciais dos parâmetros não selecionados. Valores mais adequados dos parâmetros podem ser obtidos quando um largo número de experimentos está disponível, permitindo a estimação de todos os parâmetros do modelo matemático.

O procedimento de identificabilidade também atua como uma análise de sensibilidade, uma vez que os parâmetros são ordenados de acordo com a sua influência sobre a predição do modelo. Desta forma, os resultados revelam os parâmetros aparentemente mais influentes do modelo de acordo com o cenário experimental disponível.

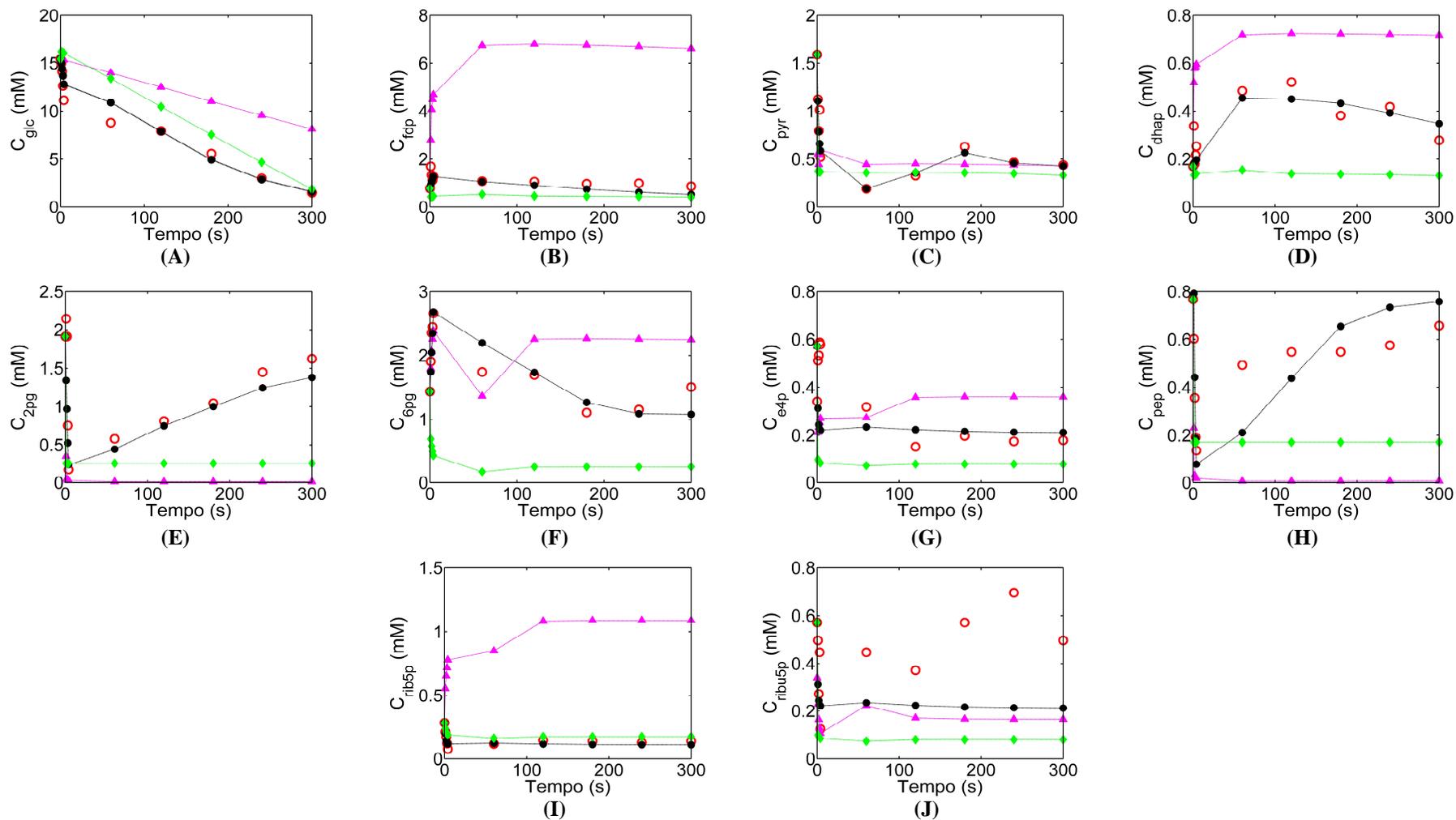


Figura 5.5 - Concentrações de substrato e metabólitos ao longo do tempo: (○) valor experimental, e valor predito com (-▲-) as estimativas iniciais dos parâmetros, (-◆-) o procedimento numérico proposto sem a estimação simultânea, (-●-) o procedimento numérico proposto para avaliação intensiva apresentado pela Figura 5.1.

Tabela 5.6 - Parâmetros identificáveis do modelo matemático da *E. coli* K-12 W3110 empregando a avaliação intensiva.

Parâmetro	Estimativa inicial	Estimado	σ_{Θ}/Θ	Parâmetro	Estimativa inicial	Estimado	σ_{Θ}/Θ
K_{PTS1}	3082,300	3683,300	0,060	$n_{DAHPS,e4p}$	2,600	3,501	0,022
r_{PDH}^{max}	4,010	0,529	0,015	$n_{DAHPS,pep}$	2,200	3,455	0,032
r_{ALDO}^{max}	30,423	1,593	0,081	E'	0,999	1,057	0,054
r_{G6PDH}^{max}	12,873	8,476	0,041	$K_{T,pep}$	0,750	0,630	0,122
$K_{R,pep}$	750,000	0,541	0,051	$K_{GAPDH,nadh}$	1,090	4,752	0,027
r_{GAPDH}^{max}	500,000	0,381	0,021	$K_{PGK,3pg}$	0,473	2,326	0,077
$K_{PEPCxylase,fdp}$	0,700	0,469	0,028	n_{PK}	4,000	1,582	0,030
$K_{R5P,eq}$	1,400	0,647	0,064	h_{hdt}	4,000	0,584	0,136
$n_{PEPCxylase,fdp}$	4,000	8,795	0,041	$K_{TIS,gap}$	0,300	0,567	0,050
r_{PK}^{max}	0,567	0,140	0,069	h_{ign}	2,000	9,169	0,057
n_{PDH}	1,000	7,237	0,028	$n_{G1PAT,fdp}$	2,000	5,993	0,084
$K_{PGI,f6p2}$	0,200	1,271	0,037	E	0,990	0,114	0,078
k_{ign}	1,000	4,152	0,075	$K_{TA,eq}$	1,050	2,583	0,054
$K_{G3PDH,dhap}$	1,000	19,076	0,110	$K_{PGLUMU,3pg}$	0,200	0,084	0,035
r_{TA}^{max}	8,461	4,051	0,051	θ	1,000	1,506	0,065
k_{hdt}	0,100	0,099	0,038	$K_{T,adp}$	1,300	1,145	0,123
$K_{PGI,g6p}$	0,400	0,060	0,101	$K_{TIS,dhap}$	2,800	7,830	0,057
$K_{PGI,f6p}$	0,266	$1,956 \times 10^{-4}$	0,115	r_{PTS}^{max}	82107,310	1,876	0,070
$K_{TKA,eq}$	0,120	6,289	0,050	$K_{PGLUMU,eq}$	0,187	0,626	0,066

$K_{PGLUMU,2pg}$	0,369	1,105	0,039	$n_{PTS,g6p}$	4,000	0,443	0,037
$K_{R5PI,eq}$	4,000	4,182	0,052	$K_{PTS,g6p}$	0,393	0,366	0,080
r_{TKA}^{max}	7,372	1,461	0,069	$K_{PKA,adp}$	0,260	0,208	0,063
r_{R5PI}^{max}	5,182	0,100	0,054	$K_{PKA,adp}$	0,400	0,455	0,056
r_{R5PI}^{max}	3,733	1,366	0,019	$K_{GAPDH,nad}$	0,252	0,093	0,070
r_{PFK}^{max}	1,967	0,382	0,057	$K_{ALDO,gap2}$	1,200	0,240	0,125
$K_{PGI,g6p2}$	0,200	0,272	0,193	$K_{PK,atp}$	22,500	1,483	0,062
L_{PK}	1,000	785,758	0,070	$K_{PGDH,6pg}$	5,449	3,597	0,051
$K_{Synthesis1,pep}$	1,000	0,361	0,053	r_{PGLUMU}^{max}	96,972	0,749	0,063
$K_{PEPCxylase,pep}$	4,070	5,731	0,038	r_{PGDH}^{max}	5,221	0,597	0,036

Para os ajustes observados a partir da Figura 5.5, obtidos com os valores de parâmetros estimados apresentados pela Tabela 5.6 ao empregar a avaliação intensiva, a função objetivo apresentou melhoria de duas ordens de grandezas com relação aos parâmetros empregando a estimativa inicial.

Vale ressaltar que a melhoria da qualidade da predição do modelo deste exemplo foi observada considerando apenas a estimação de aproximadamente a metade dos parâmetros do modelo, sendo o procedimento numérico realizado utilizando valores não conhecidos como condições iniciais das variáveis dependentes deste modelo dinâmico complexo. Desta forma, em cenários semelhantes, o grande mérito do procedimento numérico proposto é selecionar apenas os parâmetros que apresentam maior influência sobre o modelo que podem ser estimados a partir dos dados experimentais disponíveis.

Assim, mesmo sendo necessária a adoção de algumas definições arbitrárias (incertezas relacionadas ao experimento, à modelagem, às concentrações iniciais de alguns metabólitos e às estimativas iniciais dos valores de parâmetros) para informações indisponíveis à identificabilidade de parâmetros, o procedimento de avaliação intensiva demonstra ser possível ajustar e/ou melhorar a predição do modelo. Apesar das fontes de incertezas, agora o modelo é capaz de fornecer informações sobre o que esperar em certas regiões experimentais, permitindo delimitar planejamentos experimentais para investigações de interesse, dentre outras propostas. Naturalmente, para resultados mais acurados, algumas predições devem ser confirmadas com a realização de experimentos adicionais em regiões de grande interesse.

5.2 ACELERANDO O PROCEDIMENTO DE IDENTIFICABILIDADE

Nos procedimentos convencionais de identificabilidade de parâmetros, os parâmetros são incluídos um a um no conjunto de parâmetros selecionados, verificando-se a cada inclusão se este conjunto é estimável. A inclusão do primeiro parâmetro no conjunto de parâmetros selecionados é baseada na sua influência sobre a predição do modelo. Para seleção dos próximos parâmetros, utiliza-se também o grau de correlação com relação aos parâmetros inclusos no conjunto de parâmetros selecionados para re-ordenar os parâmetros restantes. Desta forma, a cada etapa de seleção deve-se avaliar a correlação entre os parâmetros, o que requer operações complexas e custosas com matrizes de grandes dimensões, tais como (KRAVARIS et al., 2012): (i) ortogonalização (YAO et al., 2003; LUND e FOSS, 2008), (ii) índice de correlação (BRUN et al., 2001), (iii) matriz de ganhos relativos, do termo em inglês *Relative Gain Array* – RGA (SANDINK et al., 2001; BOTELHO, 2012), (iv) Análise de componentes principais (LI et al., 2004, SECCHI et al., 2006), (v) Critérios de planejamento de experimentos – D-ótimo, A-ótimo, E-ótimo (WEIJERS e VANROLLEGHEM, 1997; CHU e HAHN, 2007; MACHADO et al., 2009) e (vi) valores singulares de Hankel (SUN e HAHN, 2006). Naturalmente, quanto maior o número de parâmetros, maior a dimensão destas matrizes e também maior é o número de avaliações.

Indubitavelmente, a proposta de selecionar parâmetros avaliando-se um a um é computacionalmente custosa e demorada. Recentemente, CHU e HAHN (2012), apresentaram uma metodologia para identificabilidade de parâmetros baseada em uma heurística para admitir o número de parâmetros que podem ser selecionados a cada iteração. De acordo com estes vetores característicos, um conjunto de parâmetros pode ser incluído no conjunto de parâmetros selecionados a cada iteração, sendo este conjunto determinado por um procedimento de otimização, que também emprega um critério de seleção baseado na ortogonalização da matriz de sensibilidade. Embora bastante atraente, a abordagem do procedimento de identificabilidade de parâmetros baseada em métodos de otimização resulta em um problema combinatorial, em que $NP!/[NPS!(NP-NPS)!]$ representa o número total de combinações para selecionar NPS parâmetros do número total de parâmetros NP (KRAVARIS et al., 2012). De acordo com KRAVARIS et al. (2012), seleção de parâmetros baseada em otimização incluem procedimento de seleção sequencial, técnicas não-determinísticas e heurísticas de redução.

Considerando o típico cenário de identificabilidade de parâmetros no qual se tem um modelo complexo com elevado número de parâmetros, uma importante pergunta a ser respondida é: *qual o menor número de avaliações necessárias para determinar o conjunto de parâmetros identificáveis?* Esta questão parece não ter sido abordada na literatura, mas é de grande relevância para as engenharias química e ambiental que apresentam sistemas complexos, para os quais os experimentos são difíceis, caros e demorados, sendo usuais os modelos dados por estruturas lógico-matemática de grande tamanho e complexidade.

5.3 BUSCA BINÁRIA

A busca binária é uma estratégia empregada para determinar um elemento (e.g. um caractere ou um número) em uma sequência ordenada, sem o uso de informações adicionais. Dado um conjunto de números, $\lambda = \{1, 2, 3, 4, 5, 6, 7, 8\}$, no qual um número seja aleatoriamente selecionado. Quantas perguntas são necessárias para obter o número selecionado respondendo apenas “sim” ou “não”? A forma mais rápida de dar esta resposta sem conhecer qualquer outra informação é perguntar se o número se encontra na primeira metade do conjunto. De acordo com a resposta, pode-se eliminar da busca a metade na qual

o número não está presente. Agora, pergunta se o número se encontra na primeira metade do conjunto restante de números, para que novamente se possa eliminar a metade da qual o número não está presente. Este procedimento caracteriza uma busca em árvore, no qual os ramos que não possuem o número selecionado são abandonados, restringindo rapidamente o conjunto de números a serem investigados. Note que qualquer que seja a posição do número selecionado no conjunto λ , o número de avaliações necessárias é sempre igual a três. Na busca binária o número de avaliações necessárias n_{Aval} é dada por:

$$(E 5.60) \quad n_{Aval} = \log_2 n$$

na qual n é o número de valores possíveis no conjunto original.

A Figura 5.6 apresenta um exemplo ilustrativo sobre o uso da busca binária na solução do problema de seleção do conjunto de número λ .

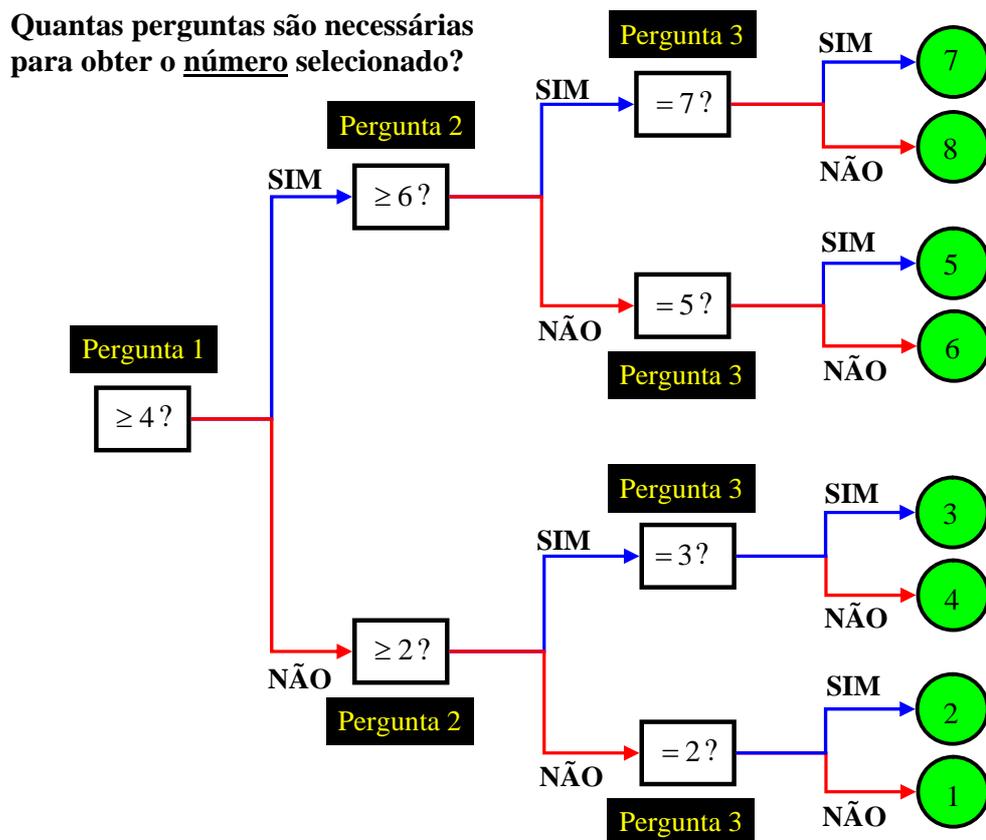


Figura 5.6 - Busca binária para solução do problema de seleção do conjunto de número λ .

5.4 SELEÇÃO CONJUNTO A CONJUNTO DOS PARÂMETROS

Um aspecto importante nos procedimentos de busca é o número de avaliações necessárias para encontrar a “solução”, porque tal procedimento requer tempo e esforço computacional. Nos procedimentos convencionais de identificabilidade de parâmetros, o número de avaliações necessárias pode tornar-se muito elevado, porque os parâmetros são avaliados um a um, sendo que o parâmetro candidato à seleção pode ser incluído e removido no conjunto de parâmetros selecionados por várias vezes ao longo do procedimento. Quando a re-estimação de parâmetros é incluída a cada etapa de seleção do procedimento de identificabilidade de parâmetros, o número de avaliações pode aumentar significativamente. Desta forma, busca-se uma solução para o problema de identificabilidade de parâmetros com um reduzido número de avaliações necessárias, economizando tempo e custo computacional.

Como ideia geral das etapas de um procedimento de identificabilidade de parâmetros simples baseado na busca binária, tem-se que inicialmente os parâmetros são ordenados acordo com sua influência sobre a predição do modelo (e.g., WEIJERS E VANROLLEGHEM, 1997; SANDINK et al., 2001; BRUN *et al.*, 2001; YAO *et al.*, 2003; LI et al., 2004; SECCHI et al., 2006; SUN e HAHN, 2006; CHU e HAHN, 2007; LUND e FOSS, 2008; MACHADO et al., 2009; CHU et al., 2011); esta etapa é usual nos procedimentos de identificabilidade de parâmetros. A busca por parâmetros identificáveis consiste em incluir no conjunto de parâmetros selecionados a primeira metade dos parâmetros mais relevantes da sequência de importância. Ao incluir parâmetros no conjunto de parâmetros selecionados deve-se avaliar o condicionamento do procedimento de estimação de parâmetros a fim de verificar se tais parâmetros podem ser estimados simultaneamente; esta etapa é necessária mesmo quando a seleção de parâmetros é realizada utilizando um critério para avaliar a correlação dos parâmetros, usualmente o condicionamento da matriz *FIM* (YAO et al., 2003; SECCHI et al., 2006), que quando assume um valor elevado também é adotado como o critério de parada em muitos procedimentos.

A principal diferença entre o procedimento de identificabilidade de parâmetros baseado na busca binária, com relação à busca binária propriamente dita, reside no fato de que ao incluir um parâmetro no conjunto de parâmetros selecionados e a matriz *FIM*

continuar inversível, então a ordem dos parâmetros não selecionados pode mudar; na busca binária clássica, a sequência ordenada não muda. Especialmente no caso em que a re-estimação de parâmetros é adotada como parte do procedimento, pois mesmo os parâmetros que já tenham sido incluídos no conjunto de parâmetros selecionados sem sucesso podem tornar a ser incluídos e se obter sucesso, como no caso do procedimento proposto para avaliação intensiva (Figura 5.1). Ocorre que, com a re-estimação, os valores dos parâmetros selecionados são modificados e a matriz *FIM* pode mudar completamente, mantendo-se inversível com a re-inclusão de parâmetros anteriormente descartados. Desta forma, todos os parâmetros não selecionados devem ser re-avaliados de acordo com o critério de identificabilidade escolhido quando um novo conjunto de parâmetros é adicionado ao conjunto de parâmetros selecionado com sucesso. Vale ressaltar que uma vez que tenha sido incluído, o parâmetro não é mais removido do conjunto de parâmetros selecionados.

5.4.1 Procedimento numérico II: Avaliação acelerada dos parâmetros

O procedimento numérico para seleção conjunto a conjunto de parâmetros proposto neste capítulo é apresentado pela Figura 5.7. De acordo com este algoritmo, NPS_s representa o número de parâmetros selecionados com sucesso e deve ser atualizado a cada etapa de seleção bem sucedida. Para cada etapa de seleção bem sucedida, o procedimento busca incluir mais parâmetros no conjunto de parâmetros selecionados $\Theta^{(S)}$, sendo NPS o número de parâmetros temporariamente armazenados neste conjunto. Todavia, se o conjunto $\Theta^{(S)}$ contendo NPS é não identificável, o tamanho deste conjunto será reduzido e NPS é atualizado. No caso de sucessivas falhas, até que $NPS=NPS_s$, indica que o procedimento está no ponto da etapa de seleção já bem sucedida, sendo que neste ponto o conjunto de parâmetros selecionados $\Theta^{(S)}$ não deve ser reduzido e mais parâmetros não podem ser incluídos. Este é o principal critério de parada adotado para o procedimento numérico proposto. Em casos particulares, o procedimento também pode parar quando todos os parâmetros são identificáveis ($\Theta^{(NS)}$ está vazio) ou quando nenhum parâmetro é identificável ($\Theta^{(S)}$ está vazio). Este último caso se dá quando algum problema ocorre para identificar mesmo um parâmetro e possivelmente indica a definição de informações inadequadas para os parâmetros e/ou algum problema de modelagem.

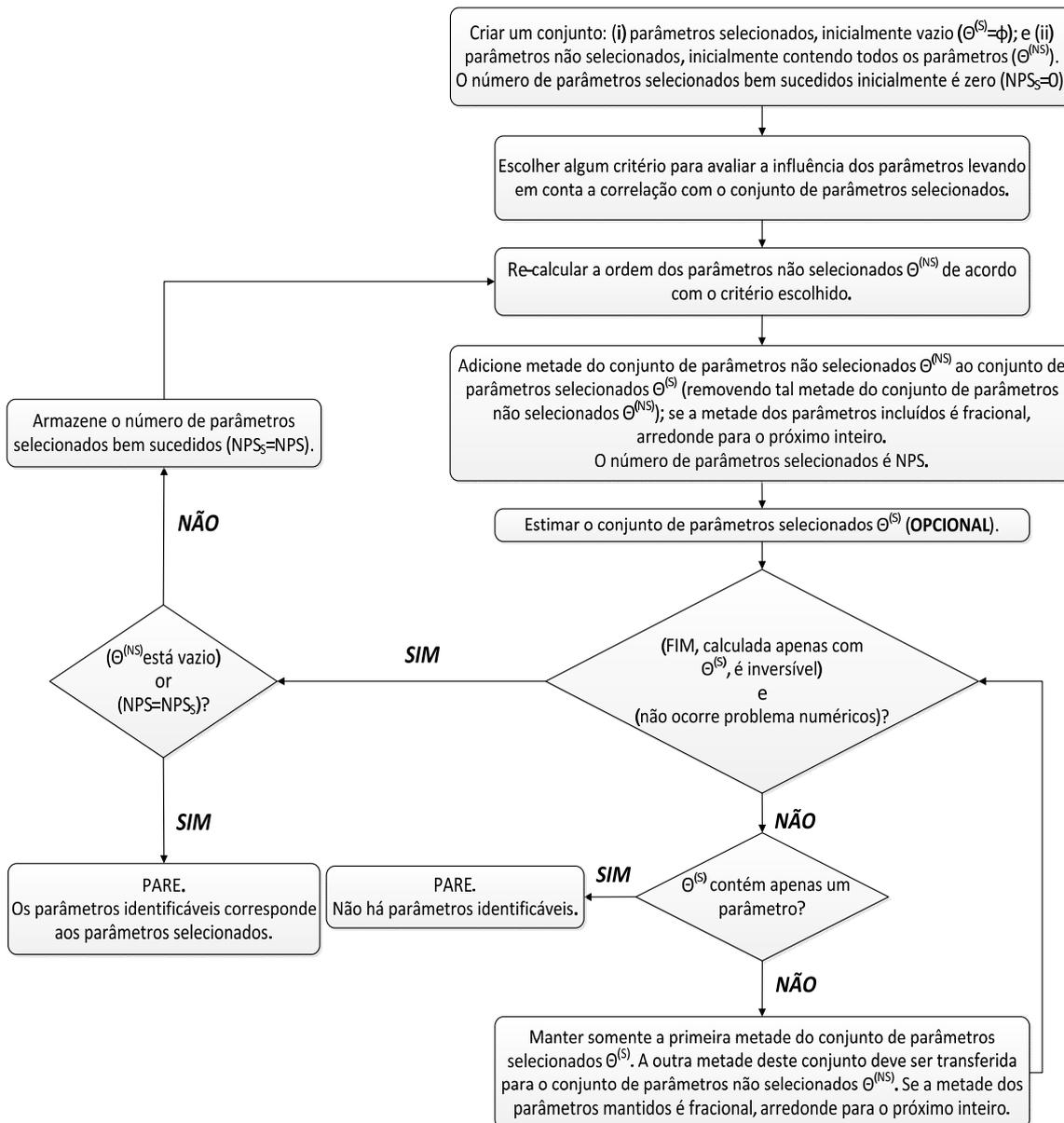


Figura 5.7 - Procedimento numérico proposto para identificabilidade de parâmetros conjunto a conjunto.

Como apresentado na Figura 5.7, os parâmetros são ordenados decrescentemente de acordo com sua importância para a predição do modelo. Este procedimento numérico também pode ser realizado empregando-se uma etapa de estimação de parâmetros a cada conjunto de parâmetros selecionados, presente no algoritmo descrito pela Figura 5.7 como uma etapa opcional. A inclusão desta etapa aumenta a complexidade do procedimento numérico proposto e consequentemente seu custo computacional. A estimação de parâmetros é um procedimento de otimização que inclui: (1) simulação do modelo

matemático (usualmente complexo), (2) verificação dos limites dos parâmetros, (3) cálculo de derivadas, e (4) inversão de matriz. No caso do uso de métodos de otimização não-determinísticos, e.g., Enxame de Partículas (KENNEDY e EBERHART, 1995) ou Algoritmo Genético (GOLDBERG, 1989), o número de simulações do modelo é bastante elevado a cada iteração e o esforço computacional destes métodos é usualmente superior. Adicionalmente, os valores estimados dos parâmetros selecionados e seus limites devem ser atualizados para um novo cálculo da matriz de sensibilidade a cada avaliação. A vantagem da inclusão de uma etapa de estimação de parâmetros no procedimento numérico está relacionada com a qualidade das estimativas iniciais em cada avaliação, que levam a uma melhor seleção.

A fim de esclarecer a metodologia do procedimento proposto, na Figura 5.8 ilustra-se o procedimento para um problema de identificabilidade hipotético contendo 4 parâmetros, apresentando as possíveis combinações para este caso. O primeiro conjunto de parâmetros avaliado no procedimento proposto é selecionado com base no critério de importância dos parâmetros, representados por quadrados preenchidos na Figura 5.8. Usualmente, a correlação paramétrica é avaliada entre o conjunto de parâmetros selecionados $\Theta^{(S)}$ e o conjunto de parâmetros não selecionados $\Theta^{(NS)}$. Esta medida é adotada para assegurar que os parâmetros a serem incluídos no conjunto de parâmetros selecionados $\Theta^{(S)}$ não estão correlacionados com os parâmetros já contidos neste conjunto.

O conhecimento prévio da correlação entre os parâmetros pode ser uma tarefa árdua, porque se trata de um procedimento que requer considerar as combinações entre parâmetros, o que pode levar a uma explosão combinatorial. No procedimento numérico proposto a correlação entre os parâmetros é avaliada a cada etapa de seleção; caso a correlação inviabilize a estimação, o conjunto de parâmetros recém adicionados deve ser reduzido à metade e novamente verifica-se a viabilidade da estimação. O número de avaliações necessárias para selecionar um conjunto de parâmetros identificáveis é diferente do número de avaliações da busca binária (Equação (E 5.60)), uma vez que há a possibilidade de re-ordenar os parâmetros não selecionados quando um novo parâmetro é incluído.

No pior caso do procedimento de identificabilidade baseado na busca binária, somente um parâmetro é incluído por vez e a re-estimação pode tornar possível a inclusão

dos parâmetros que já tenham sido avaliados sem sucesso. Desta forma, pode-se calcular o número máximo de avaliações, que incluem a re-estimação e a inversão da matriz *FIM*, como apresentado nas Equações (E 5.61) e (E 5.62), em que $I_{\text{sup}}(\)$ representa o inteiro superior.

$$(E\ 5.61) \quad N_{Aval} = \underbrace{I_{\text{sup}}(\log_2(NP))}_{\text{Avaliação para incluir somente o primeiro parâmetro}} + \underbrace{I_{\text{sup}}(\log_2(NP-1))}_{\text{Avaliação para incluir somente o segundo parâmetro}} + \dots$$

$$+ \underbrace{I_{\text{sup}}(\log_2(2))}_{\text{Avaliação para incluir somente o penúltimo parâmetro}} + \underbrace{1}_{\text{Avaliação para incluir ou remover o último parâmetro}}$$

$$(E\ 5.62) \quad N_{Aval} = 1 + \sum_{k=0}^{NP-2} I_{\text{sup}}(\log_2(NP-k)) \geq 1 + I_{\text{sup}}(\log_2(NP!)).$$

Do contrário, pode-se comparar o número de avaliações obtidas pelo procedimento numérico proposto com os procedimentos convencionais de identificabilidade de parâmetros, nos quais apenas um parâmetro é selecionado por vez. No pior caso para os procedimentos de identificabilidade convencionais, todos os parâmetros restantes são avaliados e somente na última avaliação um parâmetro é incluído, permitindo a re-avaliação de todos dos demais parâmetros restantes novamente. Desta forma, pode-se calcular o número máximo de avaliações, que incluem a re-estimação e a inversão da matriz *FIM*, como:

$$(E\ 5.63) \quad N_{Aval} = \underbrace{NP}_{\text{Avaliação para incluir somente o primeiro parâmetro}} + \underbrace{(NP-1)}_{\text{Avaliação para incluir somente o segundo parâmetro}} + \dots$$

$$+ \underbrace{(NP-2)}_{\text{Avaliação para incluir somente o penúltimo parâmetro}} + \underbrace{1}_{\text{Avaliação para incluir ou remover o último parâmetro}}$$

$$(E\ 5.64) \quad N_{Aval} = \sum_{k=1}^{NP} k = \frac{NP(NP+1)}{2}.$$

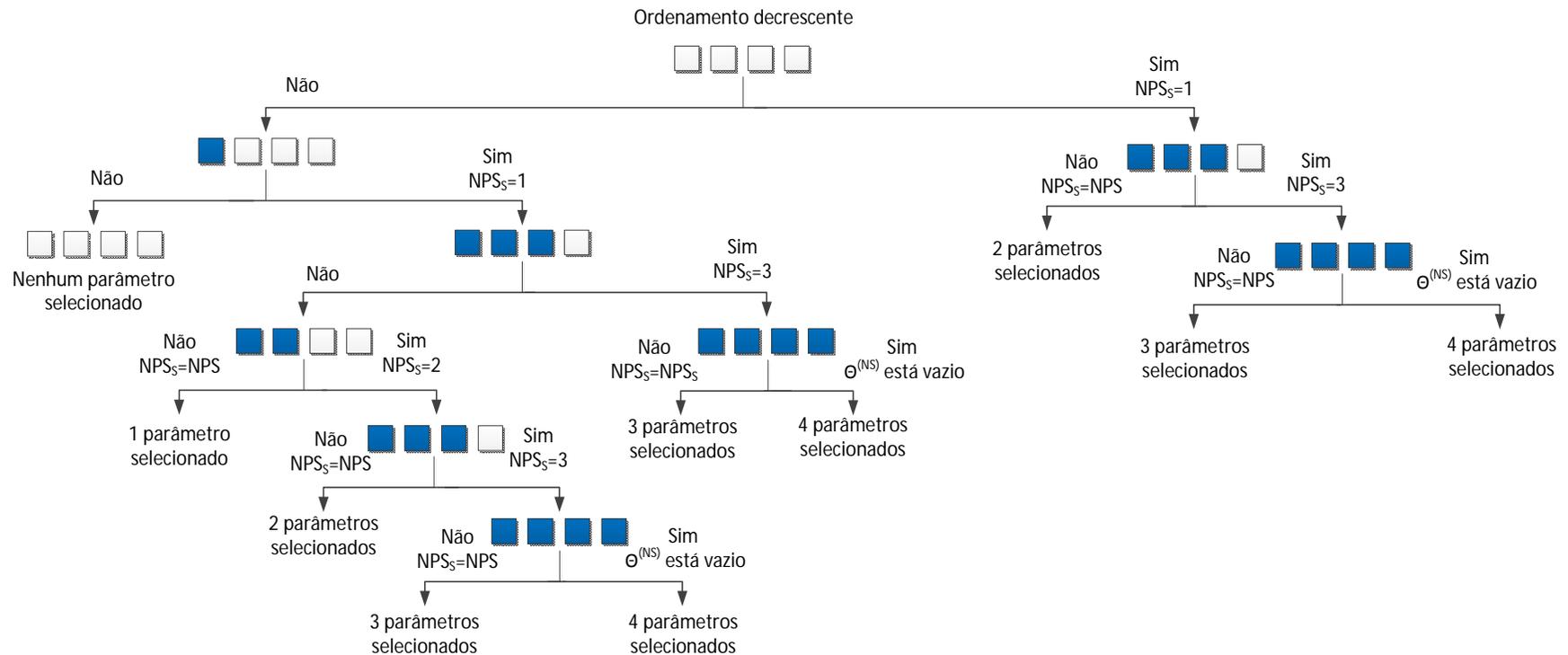


Figura 5.8 - Possíveis combinações empregando a seleção conjunto a conjunto de parâmetros (Figura 5.7) para solução de um problema de identificabilidade hipotético contendo 4 parâmetros; NPS_S e NPS , respectivamente, representam o número de parâmetros selecionados bem sucedidos e o número de parâmetros no conjunto de parâmetros selecionados $\Theta^{(S)}$, sendo $\Theta^{(NS)}$ a representação empregada para o conjunto de parâmetros não selecionados.

Dois estudos de casos são apresentados para ilustrar o procedimento numérico proposto. O primeiro exemplo consiste em um problema de identificabilidade de parâmetro hipotético com solução conhecida, cujo objetivo é verificar se o procedimento é capaz de realizar uma adequada seleção. O segundo exemplo é dado pelo modelo matemático das rotas metabólicas do microorganismo *E. coli* K-12 W3110, descrito na Seção 5.2.2. Em ambos os problemas de identificabilidade, a etapa de re-estimação simultânea foi avaliada.

5.4.2 Estudo de caso 1: modelo hipotético 6

Considere o modelo apresentado na Equação (E 5.65).

$$(E\ 5.65) \quad y = \sum_{k=1}^{NX} x_k \theta_k + \sum_{i=1}^{NX-1} \sum_{j=1}^{NX} x_i x_j \theta_{i,j},$$

em que $y \in \mathfrak{R}^1$, $x \in \mathfrak{R}^{31}$ e $\theta \in \mathfrak{R}^{961}$.

O modelo matemático original é descrito apenas pelos efeitos principais, em que o número de parâmetro, NP , é igual ao número de variáveis independentes, $NX=31$. Os efeitos cruzados são adicionados ao modelo original para criar um modelo super parametrizado e aumentar a complexidade do problema de identificabilidade. É esperado que $31 \times 30 = 930$ parâmetros relacionando os termos de segunda ordem sejam não identificáveis porque estão relacionados com outros parâmetros e não foram empregados para gerar os dados experimentais simulados (y). Desta forma, os efeitos cruzados são confundidos com os efeitos principais.

Os valores das variáveis independentes x foram gerados por meio de experimentos simulados, obtidos utilizando a técnica de planejamento ortogonal de Taguchi, disponível no *software* STATISTICA 6.0. A variável dependente y foi gerada utilizando a saída original do modelo com adição do erro experimental dado pela distribuição normal de probabilidades, com média zero e variância de 0,25 ($\sigma_y^2 = 0,25$). Os valores exatos dos parâmetros formam uma progressão aritmética com o valor do primeiro parâmetro definido igual a cinco, $\theta_1 = 5$, e o valor do último parâmetro definido como cento e cinquenta e cinco, $\theta_{31} = 155$, e a diferença comum entre os

membros sucessivos é igual a cinco. Neste cenário os dados experimentais simulados são ortogonais e a estrutura do modelo é linear.

Dois casos foram investigados, sem e com a etapa de re-estimação. Além disso, uma vez que o procedimento de identificabilidade é fortemente dependente das estimativas inicial dos valores dos parâmetros, também foram avaliados casos considerando valores de estimativa inicial bastante distintas dos valores exatos dos parâmetros. Para modelos lineares nos parâmetros esta dependência é gerada pelo uso da matriz de sensibilidade normalizada B_N e também pelos parâmetros não selecionados quando a etapa de re-estimação é empregada. Em ambos os casos os resultados são apresentado na Tabela 5.7.

Caso a - sem a etapa de re-estimação

O principal objetivo deste caso é demonstrar o baixo tempo e custo computacional do procedimento numérico proposto comparado aos procedimentos de identificabilidade usuais reportados na literatura. De fato, somente 4 avaliações foram necessárias para selecionar 31 parâmetros, enquanto que os procedimentos usuais requerem no mínimo 31 avaliações. Note que, no pior caso, o procedimento numérico proposto requer 8588 avaliações (Equação (E 5.63)), contra 461281 avaliações do procedimento convencional (Equação (E 5.64)); estes números indicam que esforço computacional do procedimento numérico proposto é menor que 2% do esforço requerido pelos procedimentos convencionais para o caso em que todo parâmetro avaliado é removido do conjunto de parâmetros selecionados.

A fim de demonstrar a dependência com relação à estimativa inicial dos parâmetros, dois conjuntos de estimativas iniciais foram investigados:

- (i). Caso a.1 – valores dos parâmetros 1 a 31 em ordem crescente de magnitude, com os parâmetros relacionados aos efeitos cruzados definidos com valor igual a 1.
- (ii). Caso a.2 – valores dos parâmetros 1 a 31 em ordem decrescente de magnitude, com os parâmetros relacionados aos efeitos cruzados definidos com valor igual a 1.

Como esperado, para o caso a.1 os parâmetros foram selecionados em ordem decrescente de acordo com os seus valores, sendo obtido o conjunto de parâmetros

selecionado $\Theta^{[S]} = \{\theta_{31}, \theta_{30}, \dots, \theta_1\}$. Para o caso a.2, os parâmetros foram selecionados de acordo com os seus valores, obtendo-se o conjunto de parâmetros selecionados $\Theta^{[S]} = \{\theta_1, \theta_2, \dots, \theta_{31}\}$. É importante enfatizar que para problemas mais complexos (e.g., dados experimentais escassos e modelos não lineares) os resultados da seleção de parâmetros também são fortemente dependentes do critério de identificabilidade de parâmetros empregado. Apesar da gravidade dos resultados apresentados neste caso, o problema da dependência com as estimativas iniciais também ocorre nos procedimentos convencionais de identificabilidade de parâmetros.

O tempo computacional foi o mesmo para ambos os casos, 3,8 segundos.

Caso b - com a etapa de re-estimação

Semelhante ao caso (a), as investigações foram realizadas considerando diferentes estimativas iniciais:

- (i). Caso b.1 – valores dos parâmetros 1 a 31 em ordem crescente de magnitude, com os parâmetros relacionados aos efeitos cruzados definidos com valor igual a 1.
- (ii). Caso b.2 – valores dos parâmetros 1 a 31 em ordem decrescente de magnitude, com os parâmetros relacionados aos efeitos cruzados definidos com valor igual a 1.

Introduzindo a etapa de re-estimação, 31 parâmetros foram selecionados, sendo a ordem de seleção igual à obtida sem a re-estimação, para ambos os casos. Apresentados pela Tabela 5.7, os resultados dos valores dos parâmetros estimados para ambos os casos apresentam desvios da ordem de grandeza de 1×10^{-2} , que indicam uma boa acurácia dos valores estimados. O tempo computacional do caso b.1 foi de 1,30 minutos e no caso b.2 foi de 1,84 minutos.

Assim como para o caso a, no caso b a ordem da seleção ocorreu de acordo com as estimativas iniciais. Para o caso b, em que a re-estimação é empregada, a diferença entre os valores dos parâmetros estimados e os valores exatos dos parâmetros está principalmente associada aos efeitos cruzados, que apresentam valor igual 0,0 no modelo exato e valor igual a 1,0 como estimativas iniciais, e tais valores não mudam durante a estimação visto que os efeitos cruzados não são incluídos no conjunto de

parâmetros identificáveis. É importante ressaltar, que os valores estimados dos parâmetros também foram afetados pelas estimativas iniciais.

Neste exemplo, a análise *NSRGA* também foi empregada para selecionar o primeiro conjunto de parâmetros, a fim de evitar a seleção de parâmetros correlacionados. Entretanto, o uso desta análise não afetou os resultados.

Tabela 5.7 - Identificabilidade de parâmetros baseada na busca binária para os casos (a) e (b) do modelo simples.

Símbolo	Valor exato	Estimativa inicial		Valor estimado	
		(a.1 e b.1)	(a.2 e b.2)	(b.1)	(b.2)
θ_1	5	1,2	31	14,111 ⁽¹⁾	159,265 ⁽³¹⁾
θ_2	10	2	30	13,868 ⁽²⁾	152,752 ⁽³⁰⁾
θ_3	15	3	29	17,608 ⁽³⁾	144,396 ⁽²⁹⁾
θ_4	20	4	28	23,079 ⁽⁴⁾	145,181 ⁽²⁸⁾
θ_5	25	5	27	26,570 ⁽⁵⁾	138,824 ⁽²⁷⁾
θ_6	30	6	26	31,025 ⁽⁶⁾	134,280 ⁽²⁶⁾
θ_7	35	7	25	39,200 ⁽⁷⁾	129,384 ⁽²⁵⁾
θ_8	40	8	24	45,615 ⁽⁸⁾	120,687 ⁽²⁴⁾
θ_9	45	9	23	46,368 ⁽⁹⁾	123,449 ⁽²³⁾
θ_{10}	50	10	22	57,366 ⁽¹⁰⁾	115,398 ⁽²²⁾
θ_{11}	55	11	21	63,027 ⁽¹¹⁾	108,222 ⁽²¹⁾
θ_{12}	60	12	20	64,825 ⁽¹²⁾	104,095 ⁽²⁰⁾
θ_{13}	65	13	19	72,339 ⁽¹³⁾	103,921 ⁽¹⁹⁾
θ_{14}	70	14	18	68,598 ⁽¹⁴⁾	88,336 ⁽¹⁸⁾
θ_{15}	75	15	17	76,416 ⁽¹⁵⁾	85,012 ⁽¹⁷⁾
θ_{16}	80	16	16	86,799 ⁽¹⁶⁾	86,799 ⁽¹⁶⁾
θ_{17}	85	17	15	85,012 ⁽¹⁷⁾	76,416 ⁽¹⁵⁾
θ_{18}	90	18	14	88,336 ⁽¹⁸⁾	68,598 ⁽¹⁴⁾
θ_{19}	95	19	13	103,921 ⁽¹⁹⁾	72,339 ⁽¹³⁾
θ_{20}	100	20	12	104,095 ⁽²⁰⁾	64,825 ⁽¹²⁾
θ_{21}	105	21	11	108,222 ⁽²¹⁾	63,026 ⁽¹¹⁾
θ_{22}	110	22	10	115,398 ⁽²²⁾	57,366 ⁽¹⁰⁾
θ_{23}	115	23	9	123,449 ⁽²³⁾	46,368 ⁽⁹⁾
θ_{24}	120	24	8	120,687 ⁽²⁴⁾	45,615 ⁽⁸⁾
θ_{25}	125	25	7	129,384 ⁽²⁵⁾	39,200 ⁽⁷⁾
θ_{26}	130	26	6	134,280 ⁽²⁶⁾	31,025 ⁽⁶⁾
θ_{27}	135	27	5	138,824 ⁽²⁷⁾	26,570 ⁽⁵⁾
θ_{28}	140	28	4	145,181 ⁽²⁸⁾	23,079 ⁽⁴⁾
θ_{29}	145	29	3	144,396 ⁽²⁹⁾	17,608 ⁽³⁾
θ_{30}	150	30	2	152,752 ⁽³⁰⁾	13,868 ⁽²⁾
θ_{31}	155	31	1,2	159,265 ⁽³¹⁾	14,111 ⁽¹⁾
$\theta_{1,1}$ à $\theta_{30,31}$	0	1	1	*	*

* parâmetros não selecionados.

⁽ⁱ⁾ indica a posição *i* do parâmetro de acordo com a ordem de seleção.

5.4.3 Estudo de caso 2: modelo das rotas metabólicas da *E. coli* K-12 W3110

Foram utilizados neste exemplo, o modelo (CHASSAGNOLE et al., 2002; Di MAGGIO et al., 2010), os dados experimentais (HOQUE et al., 2005) e estimativas iniciais de parâmetros (Di MAGGIO *et al.*, 2010), que se encontram descritos em maior detalhe na Seção 5.2.2.

Assim como para o modelo hipotético 6 desta seção, dois casos foram investigados, sem e com a etapa de re-estimação a fim de avaliar a redução da dependência do procedimento com a estimativa inicial dos valores de parâmetros. Uma vez que se trata de um problema real de identificabilidade, as estimativas iniciais empregadas foram tomadas da literatura.

Caso a – sem a etapa de re-estimação

Considerando as 5 condições iniciais e os 131 parâmetros do modelo, foram avaliados 136 parâmetros utilizando o procedimento numérico proposto a partir dos dados experimentais medidos dos metabólitos intracelulares. Os resultados são apresentados na Tabela 5.8, em que 20 parâmetros e 5 condições iniciais foram selecionados como identificáveis, sendo requerido apenas 13 avaliações. O elevado número de parâmetros não selecionados pode indicar que vários parâmetros estão correlacionados ou apresentam pouca influência sobre a predição do modelo, e também que os metabólitos intracelulares não medidos armazenam informações relevantes para descrição adequada do comportamento metabólico da *E. coli*. Outro fator importante é a qualidade da estimativa inicial dada aos parâmetros; tais valores podem estar distantes dos valores exatos dos parâmetros.

Entretanto, a estimação simultânea dos parâmetros identificáveis a partir da estimativa inicial, apresentados na Tabela 5.8, não foi possível devido à singularidade da matriz *FIM*.

Tabela 5.8 - Procedimento baseado na busca binária para identificabilidade dos parâmetros das rotas metabólicas da *E. coli* K-12 W3110, sem a etapa de estimação.

Símbolo	Estimativa inicial	Símbolo	Estimativa inicial
<u>Condições iniciais</u>			
C_{xyl5p}	$5,5406 \times 10^{-1}$	C_{3gp}	1,2851
C_{g1p}	$6,7016 \times 10^{-2}$	C_{sed7p}	$4,0874 \times 10^{-2}$
C_{pgp}	$3,7000 \times 10^{-5}$		
<u>Parâmetros</u>			
$K_{PGI,g6p}$	$4,0000 \times 10^{-1}$	$K_{GAPDH,pgp}$	$5,0000 \times 10^{-5}$
r_{PGI}^{max}	$1,0519 \times 10^3$	$K_{G6PDH,nad,ihn}$	$9,0000 \times 10^{-3}$
r_{G6PDH}^{max}	$1,2873 \times 10^1$	$r_{Synthesis2}^{max}$	$5,6900 \times 10^{-1}$
$K_{PGI,g6p2}$	$2,0000 \times 10^{-1}$	r_{PK}^{max}	$5,6700 \times 10^{-1}$
$K_{TIS,eq}$	$1,3900 \times 10^{-1}$	r_{PTS}^{max}	$8,21073 \times 10^4$
$K_{PEPCxylase,fdp}$	$7,0000 \times 10^{-1}$	K_{PTS1}	$3,0823 \times 10^3$
$n_{PEPCxylase,fdp}$	4,0000	$K_{Synthesis2,pyr}$	1,0000
$K_{RUP,eq}$	1,4000	r_{PGDH}^{max}	5,22076
r_{GAPDH}^{max}	50,0000	$K_{PGDH,nadp}$	$5,0600 \times 10^{-2}$
$K_{GAPDH,gap}$	1,2000	$K_{PGDH,nadph}$	$1,3769 \times 10^{-2}$

Caso b – com a etapa de re-estimação

Incluindo a etapa de re-estimação a cada etapa de seleção, o procedimento numérico proposto foi capaz de selecionar 75 parâmetros e 5 condições iniciais demandando apenas 19 iterações. É importante enfatizar que no procedimento convencional, seleção de parâmetros um a um, o número mínimo requerido de iterações para se obter este resultado é 80. Na verdade, a redução do esforço computacional é muito maior visto que o procedimento convencional demanda muitas iterações, para as quais a inclusão de um parâmetro a cada iteração é frequentemente mal sucedida, de forma que vários parâmetros são testados até que um parâmetro seja definitivamente incluído no conjunto de parâmetros selecionados.

O tempo computacional para selecionar os 75 parâmetros e as 5 condições iniciais com o procedimento numérico proposto, seleção de parâmetros conjunto a conjunto, foi de apenas 3,24 minutos. Para comparação, o procedimento numérico para avaliação intensiva apresentado na Seção 5.2.1, Figura 5.1, realiza o procedimento de identificabilidade por meio da seleção um a um, obtendo os resultados em 6,28 horas, com resultados de qualidade semelhante. Os resultados são apresentados pelas Tabela

5.9 e Tabela 5.10, respectivamente, para as condições iniciais e os parâmetros do modelo. De acordo com os resultados apresentados pela Tabela 5.9, os valores calculados para a estimativa inicial das condições iniciais não sofreram mudança significativa após o procedimento de identificabilidade.

Tabela 5.9 - Procedimento baseado na busca binária para identificabilidade dos parâmetros das rotas metabólicas da *E. coli* K-12 W3110, com a etapa de estimação: condição inicial para os metabólitos intracelulares não medidos.

Condições iniciais

Símbolo	Estimativa inicial	Valor estimado	σ_{Θ}/Θ
C_{xy15p}	$5,5406 \times 10^{-1}$	$5,2742 \times 10^{-1}$	$8,9433 \times 10^3$
C_{g1p}	$6,7016 \times 10^{-2}$	$6,4919 \times 10^{-2}$	$9,8674 \times 10^1$
C_{pgp}	$3,7000 \times 10^{-5}$	$3,0857 \times 10^{-3}$	3,3957
C_{3gp}	1,2851	1,7901	3,5516
C_{sed7p}	$4,0874 \times 10^{-2}$	$4,0850 \times 10^{-2}$	$1,4311 \times 10^4$

Tabela 5.10 - Procedimento baseado na busca binária para identificabilidade dos parâmetros das rotas metabólicas da *E. coli* K-12 W3110, com a etapa de estimação de parâmetros.

Símbolo	Valor nominal	Valor estimado	σ_{Θ}/Θ	Símbolo	Valor nominal	Valor estimado	σ_{Θ}/Θ
$K_{PGI,g6p}$	$9,5682 \times 10^3$	$1,0769 \times 10^2$	3,0814	$K_{PGK,adp}$	$1,8500 \times 10^{-1}$	$1,8500 \times 10^{-1}$	$1,9104 \times 10^2$
r_{PGI}^{\max}	$1,0519 \times 10^3$	$5,9234 \times 10^2$	1,2569	$K_{GAPDH,nad}$	$2,5200 \times 10^{-1}$	$2,5213 \times 10^{-1}$	$6,7565 \times 10^1$
r_{G6PDH}^{\max}	7,5600	7,2217	$2,0269 \times 10^1$	$K_{G6PDH,nad,inh}$	$9,0000 \times 10^{-3}$	$8,6536 \times 10^{-3}$	$6,1194 \times 10^1$
$K_{PGI,f6p2}$	$2,0000 \times 10^{-1}$	$1,1109 \times 10^{-7}$	$7,6247 \times 10^5$	r_{RUSP}^{\max}	5,1816	2,2516	$2,5393 \times 10^1$
$K_{PGK,3pg}$	$5,1700 \times 10^{-1}$	$4,5490 \times 10^{-1}$	4,4129	r_{R5PI}^{\max}	3,7333	1,9924	4,6917
$K_{PEPCxylase,fdp}$	$4,6009 \times 10^{-1}$	$4,4571 \times 10^{-1}$	3,7857	r_{ALDO}^{\max}	$3,0423 \times 10^1$	4,0000	$1,7730 \times 10^3$
$n_{PEPCxylase,fdp}$	9,2749	9,2161	2,3282	$K_{ENO,2pg}$	$1,0000 \times 10^{-1}$	$1,0000 \times 10^{-1}$	$1,6760 \times 10^1$
r_{PGDH}^{\max}	$5,7921 \times 10^{-1}$	$9,4885 \times 10^{-1}$	$1,2338 \times 10^3$	$K_{TIS,dhap}$	2,8000	2,5225	$2,2587 \times 10^3$
r_{PFK}^{\max}	1,9670	$1,4669 \times 10^{-1}$	2,5160	$K_{R5PI,eq}$	4,0000	1,4200	$3,5450 \times 10^1$
r_{GAPDH}^{\max}	$5,0000 \times 10^2$	9,4187	9,5864	r_{PDH}^{\max}	4,0104	$6,2336 \times 10^{-1}$	$7,4603 \times 10^{-1}$
$K_{RUSP,eq}$	1,4000	1,4855	$8,9487 \times 10^3$	$K_{PGLUMU,3pg}$	$2,0000 \times 10^{-1}$	$2,0000 \times 10^{-1}$	$6,9664 \times 10^2$
$K_{GAPDH,gap}$	$8,9226 \times 10^{-3}$	$2,2312 \times 10^{-2}$	$1,3122 \times 10^3$	$H_{G6PDH,nadh}$	4,0000	$5,1769 \times 10^{-1}$	$5,1116 \times 10^3$
$K_{GAPDH,pgp}$	$3,4443 \times 10^{-5}$	$8,3295 \times 10^{-5}$	$1,3434 \times 10^3$	$K_{TKa,eq}$	$1,2000 \times 10^{-1}$	$1,1685 \times 10^{-1}$	$2,3249 \times 10^4$
$K_{PGDH,nadp}$	$5,0600 \times 10^{-2}$	1,1772	$1,3215 \times 10^3$	$K_{ENO,pep}$	$1,3500 \times 10^{-1}$	$1,3500 \times 10^{-1}$	$3,9731 \times 10^1$
$K_{PGDH,nadph}$	$1,3769 \times 10^{-2}$	$7,3728 \times 10^{-2}$	$4,0426 \times 10^2$	r_{TIS}^{\max}	$5,2056 \times 10^1$	$8,7199 \times 10^{-2}$	$1,9057 \times 10^3$
$K_{PGDH,aip}$	$2,0800 \times 10^2$	$5,4786 \times 10^1$	$4,2688 \times 10^3$	$n_{DAHPS,pep}$	2,2000	2,4990	$2,6254 \times 10^1$
$K_{PGM,eq}$	$1,5800 \times 10^{-1}$	$1,3716 \times 10^{-1}$	$8,7306 \times 10^1$	$K_{PGLUMU,2pg}$	$3,6900 \times 10^{-1}$	$3,6900 \times 10^{-1}$	$6,3087 \times 10^2$
$K_{ALDO,eq}$	$1,4438 \times 10^{-2}$	$1,4935 \times 10^{-2}$	2,6863	$K_{G6PDH,nadph}$	$3,0000 \times 10^{-2}$	$2,8818 \times 10^{-2}$	$7,5260 \times 10^3$
r_{Synh2}^{\max}	$5,6900 \times 10^{-1}$	$4,2606 \times 10^{-2}$	$2,6446 \times 10^1$	$H_{G6PDH,nadph}$	1,4000	1,4032	$8,6816 \times 10^3$
r_{PK}^{\max}	$3,5309 \times 10^{-2}$	$3,5304 \times 10^{-2}$	$6,0122 \times 10^1$	$K_{R,f6p}$	$8,8572 \times 10^{-3}$	$8,8801 \times 10^{-3}$	$9,6690 \times 10^3$
r_{PTS}^{\max}	1,1118	1,0745	$3,5897 \times 10^3$	$K_{MG6PDH,nadp}$	$2,0000 \times 10^{-2}$	$1,9406 \times 10^{-2}$	$3,7786 \times 10^3$

K_{PTS1}	2,4200x10 ³	2,3386x10 ³	3,6104x10 ³	$K_{TA,eq}$	1,0500	1,0507	1,4314x10 ⁴
$K_{Synthesis2,pyr}$	1,8164x10 ⁻¹	1,8140x10 ⁻¹	1,5514x10 ²	$K_{TIS,eq}$	1,3900x10 ⁻¹	1,3896x10 ⁻¹	2,9680x10 ²
$K_{PK,adp}$	2,6000x10 ⁻¹	2,5944x10 ⁻¹	2,9091x10 ²	$r_{PEPCxylase}^{max}$	9,5000x10 ⁻²	1,0256x10 ⁻²	6,2233x10 ¹
$K_{PGK,pgp}$	4,6800x10 ⁻²	5,4125x10 ⁻²	3,5774	$K_{ALDO,dhap}$	8,8000x10 ⁻²	8,8000x10 ⁻²	1,4290x10 ²
$n_{DAHP,se4p}$	2,6000	3,4499	3,5791	E	9,9000x10 ⁻¹	4,0301x10 ⁻¹	9,6704x10 ³
$K_{PGLUMU,eq}$	9,4504x10 ⁻¹	9,4505x10 ⁻¹	7,0688	r_{PGM}^{max}	5,6180x10 ⁻¹	9,0650x10 ⁻²	5,7009x10 ¹
$K_{PK,pep}$	3,1000x10 ⁻²	3,0968x10 ⁻²	1,4155x10 ²	$K_{PGDH,6pg}$	5,4491	3,6871	1,1154x10 ²
n_{PDH}	1,0000	9,2200	1,5091	$K_{G6PDH,nadh}$	3,0000x10 ⁻¹	3,0070x10 ⁻¹	1,0491x10 ⁴
$K_{PGI,eq}$	4,3000x10 ⁻²	4,2994x10 ⁻²	7,1124x10 ⁻¹	$K_{G6PDH,dm}$	3,0000x10 ⁻¹	3,0613x10 ⁻¹	4,4899x10 ⁴
$K_{PEPCxylase,pep}$	4,0700	8,0500	6,6521x10 ¹	$K_{MG6PDH,g6p}$	7,0000x10 ⁻³	6,9652x10 ⁻³	2,2991x10 ²
$K_{ENO,eq}$	6,7000x10 ⁻¹	7,1033x10 ⁻¹	2,0301	$K_{MG6PDH,g6p}$	7,0000x10 ⁻³	6,9652x10 ⁻³	2,2991x10 ²
r_{TA}^{max}	8,4613	3,2541	1,4314x10 ⁴	$K_{PGK,eq}$	1,8000x10 ³	8,9599x10 ⁻³	2,2371x10 ³
$K_{PGI,g6p2}$	2,0000x10 ⁻¹	5,9448x10 ⁻¹	4,9639	$K_{TIS,gap}$	3,0000x10 ⁻¹	3,4455x10 ⁻¹	3,0763x10 ²
r_{PGK}^{max}	4,2260x10 ³	2,7500x10 ¹	2,3841x10 ³	r_{G3PDH}^{max}	1,0700x10 ⁻²	1,0700x10 ⁻²	5,4517x10 ⁴
r_{ENO}^{max}	2,5274x10 ²	2,5000	1,0603x10 ¹	$K_{G3PDH,dhap}$	1,0000	6,5999	6,1520x10 ⁴
r_{TKa}^{max}	7,3718	5,7700	8,9491x10 ³	$r_{MurSynth}^{max}$	1,5000x10 ⁻³	1,4961x10 ⁻³	1,6909x10 ²
r_{PGLUMU}^{max}	9,6972x10 ¹	7,0600x10 ⁻¹	6,8782x10 ¹	$K_{DAHPS,pep}$	5,3000x10 ⁻³	5,3000x10 ⁻³	1,5787x10 ²

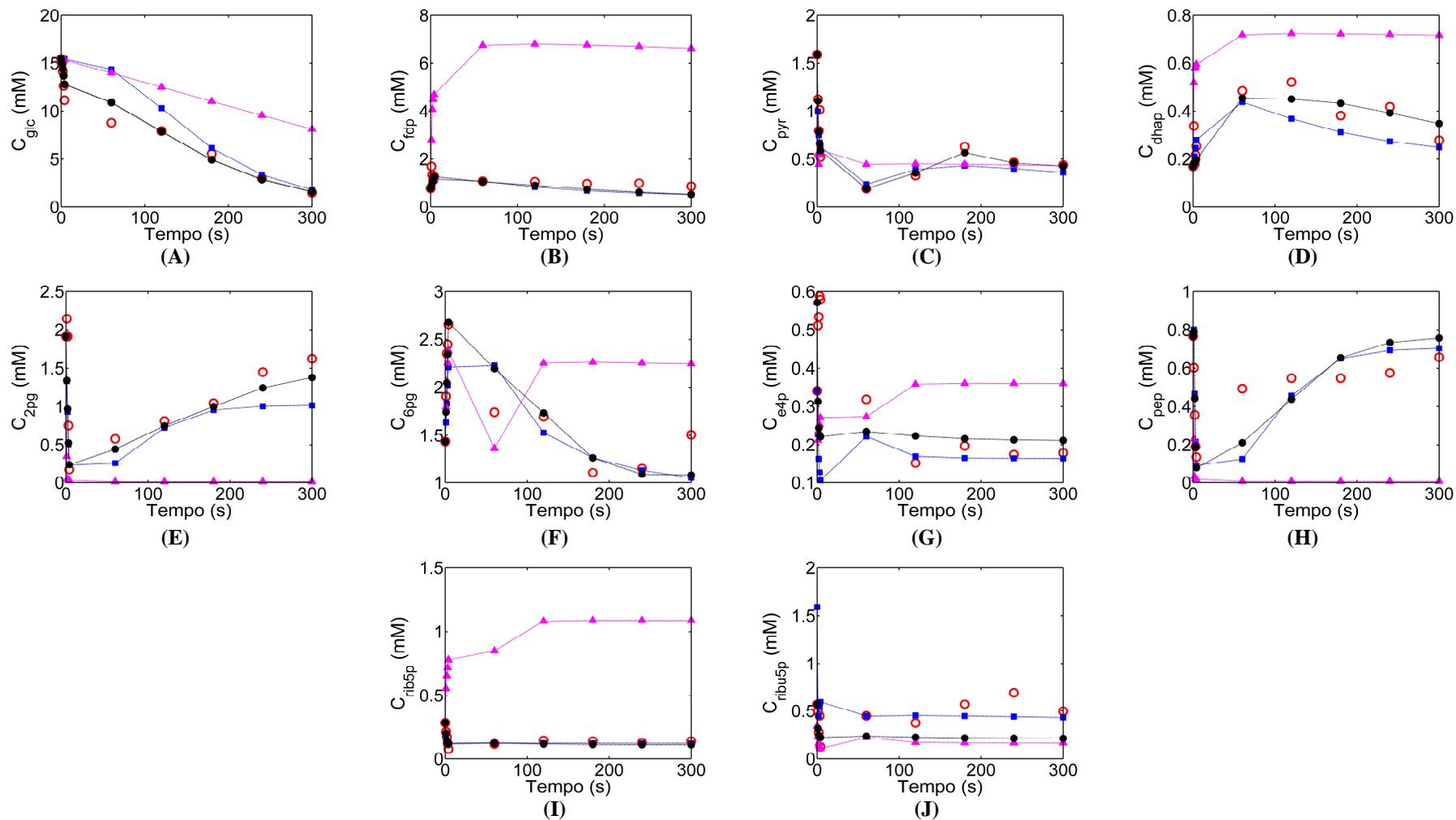


Figura 5.9 - Concentrações de substrato e metabólitos ao longo do tempo: (○) valor experimental, valor predito com (---▲---) as estimativas iniciais (caso (a) - sem a etapa de re-estimação), (---■---) o procedimento de avaliação conjunto a conjunto, descrito pela Figura 5.7 (caso (b) - com a etapa de re-estimação) e, (---●---) o procedimento de avaliação intensiva, descrito pela Figura 5.1

A maioria dos valores estimados para os parâmetros apresentados pela Tabela 5.10 apresenta diferença de pelo menos uma ordem de magnitude em relação aos valores das estimativas iniciais. Esta diferença demonstra que as estimativas iniciais dos parâmetros podem sofrer uma melhora significativa, como observado na Figura 5.9, que apresenta a comparação entre os dados experimentais e a predição do modelo com os valores de parâmetros igual às estimativas iniciais e com os valores de parâmetros estimados. É importante enfatizar que os valores estimados para os parâmetros selecionados podem ser significativamente influenciados pelos valores dos parâmetros não selecionados, que são fixados na estimativa inicial.

O uso da técnica *NSRGA* na etapa de seleção do primeiro conjunto de parâmetros, a fim evitar a inclusão de parâmetros correlacionados neste conjunto, não afeta os resultados. Assim, o uso desta etapa é desaconselhado visto que o esforço adicional introduzido para selecionar o primeiro conjunto de parâmetros é desnecessário.

5.5 CONCLUSÕES PARCIAIS

Dois novos procedimentos numéricos foram propostos a fim de tratar problemas de identificabilidade de parâmetros, com diferentes focos: (i) a avaliação intensiva dos parâmetros - a classificação robusta, (ii) avaliação acelerada dos parâmetros - o mínimo esforço computacional. O primeiro procedimento consiste na avaliação intensiva dos parâmetros, permitindo investigar o potencial de identificabilidade para todos os parâmetros do modelo, mesmo em problemas de estimação mal condicionados. No segundo procedimento busca-se acelerar a convergência dos procedimentos de identificabilidade por meio da busca binária, reduzindo drasticamente o esforço computacional.

Para ilustrar o desempenho dos procedimentos propostos, destaca-se o estudo de caso das rotas metabólicas da *E. coli* K-12 W3110, um problema de identificabilidade real, cujo cenário é composto por um modelo dinâmico complexo, contendo 131 parâmetros e 5 condições iniciais desconhecidas, que devem ser estimados a partir de um conjunto de dados experimentais escasso.

Ambos os procedimentos se mostram aptos à identificabilidade de parâmetros em cenários áridos e podem ser empregados como ferramentas para a análise de sensibilidade e planejamento de experimentos. Como esperado, os resultados obtidos

empregando-se a identificabilidade e a estimação simultâneas superam os resultados obtidos empregando a abordagem clássica. A abordagem da avaliação intensiva parece mais adequada para um estudo pormenorizado do problema de identificabilidade de parâmetros, a que se deseja detalhes para estudo da análise de sensibilidade do modelo e definições de condições experimentais de interesse. Em contrapartida, a abordagem da avaliação acelerada demonstra grande potencial para aplicações em tempo real.

O uso da técnica *NSRGA* para avaliação prévia da correlação paramétrica não afetou os resultados para os exemplos investigados. O mesmo é observado para o uso dos ângulos de correlação. De fato, o resultado obtido por meio destas técnicas é sobreposto pela avaliação da correlação entre os parâmetros realizada a cada etapa de seleção pelo critério de identificabilidade adotado.

Assim, os resultados obtidos neste capítulo corroboram as observações a respeito da forte dependência do procedimento de identificabilidade com as estimativas iniciais dos valores dos parâmetros, reafirmando a importância de se realizar a seleção e a estimação simultaneamente.

CAPÍTULO 6. CONCLUSÕES E SUGESTÕES

A presente tese leva a concluir que a identificabilidade de parâmetros apresenta grande relevância para as engenharias química e ambiental, devido à crescente complexidade dos modernos processos e às dificuldades associadas à obtenção e à disponibilidade de medidas de variáveis cruciais, seja em grandes plantas industriais ou em pequenas bancadas laboratoriais.

A revisão bibliográfica leva a concluir que há um grande panorama para desenvolvimento de novos procedimentos, especialmente com relação a modelos complexos, com lacunas importantes a serem preenchidas como: a baixa frequência de estudos relacionados aos aspectos conceituais do problema de identificabilidade, a ausência do uso de métodos globais para avaliação de parâmetros, o uso da estimação como uma etapa final ou exclusiva à seleção, a ausência da avaliação de todos os parâmetros ante à seleção, a ausência da avaliação dos parâmetros com rápida convergência. Como conclusão final desta etapa tem-se que, apesar da grande variedade de tipos e de termos para a classificação de parâmetros reportada pela literatura, na prática esta classificação é feita adotando-se apenas a seguinte definição: parâmetro identificável, que pode ser estimado, e parâmetro não identificável, que não pode ser estimado.

O estudo e comparação de métodos com diferentes propriedades como a análise de sensibilidade global de SALTELLI (2002), a identificabilidade baseada na ortogonalização de YAO et al. (2003) e a identificabilidade empregando análise dos componentes principais de SECCHI et al. (2006), frente aos estudos de casos propostos, permitem concluir que os critérios de identificabilidade não são sempre concordantes entre si e apresentam grande suscetibilidade com relação às estimativas iniciais dos parâmetros. Outra conclusão importante obtida a partir desta etapa da tese é a verificação de que a seleção baseada apenas na sensibilidade não necessariamente leva aos parâmetros mais estimáveis do modelo, visto que a estimação depende do plano experimental empregado. Também é importante ressaltar que em se tratando de modelos

complexos dinâmicos, a análise de sensibilidade global não permite ampla variação no valor dos parâmetros, ficando localizada para os casos estudados no máximo em 20% em torno da estimativa inicial dos parâmetros. Outra observação importante é o fato de que em alguns casos o conjunto de parâmetros selecionados pode resultar em problemas de estimação mal condicionados quando a estimação não é aplicada simultaneamente à identificabilidade, em especial quando as estimativas iniciais dos parâmetros não são adequadas.

O estudo dos aspectos conceituais da identificabilidade abordou as dificuldades relacionadas com a classificação dos parâmetros e com a caracterização de incertezas, concluindo que o procedimento de fixar um subconjunto de parâmetros e estimar os demais pode levar a resultados desastrosos. A conclusão final desta etapa é que o uso apropriado do procedimento de identificabilidade está sujeito à condição em que as direções características da matriz de informação de Fisher que contém a incerteza paramétrica infinita constituem o espaço nulo da matriz de sensibilidade.

A proposta de novos procedimentos a fim de tratar os problemas numéricos ocasionados pela dependência da seleção com a estimativa inicial dos parâmetros leva à conclusão de que, a estratégia de seleção e estimação simultâneas permite amenizar tais problemas numéricos, bem como assegura que o conjunto de parâmetros selecionados seja identificável. Conclui-se também que o uso de análises adicionais, como a matriz retangular de ganhos relativos e os ângulos de correlação entre os vetores de sensibilidade de parâmetros, são sobrepostos pelo critério de identificabilidade e apenas introduzem esforço adicional ao procedimento. Da comparação entre os dois novos procedimentos propostos, a avaliação intensiva e a avaliação acelerada, ambos permitem o uso de qualquer critério de identificabilidade e são baseados na seleção e estimação simultâneas, mas o esforço computacional é drasticamente reduzido neste segundo, o que demonstra grande potencial do procedimento para aplicações em tempo real. O uso da estratégia de YAO et al. (2003) nos estudos de caso apresentados é ilustrativo e não está baseado em qualquer motivação. Assim como observado nesta tese, espera-se que a avaliação intensiva dos parâmetros forneça um ajuste mais adequado dos modelos aos dados experimentais, quando comparada a seleção conjunto a conjunto, visto que ao selecionar um conjunto de parâmetros, o número de escolhas baseadas na estimativa inicial é muito maior do que considerando apenas a seleção de um parâmetro, para um mesmo cenário.

Assim, esta tese trata do tema da identificabilidade de parâmetros, abordando: (i) o problema, (ii) as metodologias de identificabilidade, (iii) as dificuldades associadas à classificação e à caracterização estatística dos parâmetros identificáveis, e (iv) o desenvolvimento de novos métodos. A conclusão maior consiste na constatação da complexidade do presente tema, sendo que a tarefa de identificar parâmetros ainda apresenta grandes desafios que carecem de estudos à luz da estatística.

Para trabalhos futuros sugerem-se investigações a cerca do desenvolvimento de novos critérios de identificabilidade, da caracterização de incerteza desprezada e desenvolvimento de procedimentos com menor dependência da estimativa inicial dos parâmetros. Acredita-se que a exploração dos tópicos apresentados a seguir é de suma importância para o avanço da identificabilidade e a obtenção de resultados cada vez mais confiáveis com o uso destes procedimentos.

- Uso dos testes estatísticos Chi-quadrado e F-Fisher para a comparação dos valores das funções objetivo obtidos via diferentes procedimentos e/ou diferentes estimativas iniciais;
- Estudo e estabelecimento de critérios que permitam definir estatisticamente, de forma inequívoca, a adequabilidade do procedimento de identificabilidade aplicado a um problema de interesse;
- Estudo e desenvolvimento de critérios de identificabilidade baseados em métodos globais, buscando reduzir a influência das estimativas iniciais sobre a seleção de parâmetros;
- Estudo e desenvolvimento de estratégias para o tratamento prévio das estimativas iniciais via métodos não-determinísticos, e.g. o Exame de Partículas (KENNEDY e EBERHART, 1995) e o Algoritmo Genético (GOLDBERG, 1989);
- Estudo e estabelecimento de critérios de paradas baseados em métodos estatísticos, visando impedir a degradação da predição;
- Estudo e desenvolvimento de procedimentos capazes de contabilizar o volume de incerteza da predição desprezado por meio de métodos de otimização, a fim de reduzir o esforço computacional atualmente requerido para abordar este problema;

- Uso de índices de sensibilidade para fins de cálculo do intervalo de confiança da predição, buscando reduzir o esforço de amostrar o espaço de parâmetros;
- Uso do conceito de volume da região de confiança dos parâmetros para desenvolvimento de critérios de identificabilidade e de parada dos procedimentos;
- Estudo e desenvolvimento de procedimentos baseados em estimadores robustos, visando caracterizar rigorosamente as incertezas e evitando assumir *a priori* as hipóteses do princípio de Máxima Verossimilhança;
- Estudo e desenvolvimento de procedimentos baseados na teoria da decisão e na teoria da informação, visando reduzir o esforço computacional requerido pelos procedimentos clássicos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBERTON, A.L., 2010, *Estimação de Parâmetros e Planejamento de Experimentos: Estudo de Incertezas e Funções de Informação*. Tese de D.Sc., PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ALBERTON, K.P.F., ESTRADA, V.G., SOLEDAD DÍAZ, M., SECCHI, A.R., 2012a., "Indentificabilidade de Parâmetros para Controle de Poluição em Corpos d'água Eutrofizados". In: *Proceeding of the 19^o Congresso Brasileiro de Engenharia Química*. Búzios, Setembro.
- ALBERTON, K.P.F., DI MAGGIO, J.A., SOLEDAD DÍAZ, M., SECCHI, A.R. 2012b. "Modelagem e Estimação de Parâmetros Cinéticos de Rotas Metabólicas em Larga Escala". In: *Proceeding of the 19^o Congresso Brasileiro de Engenharia Química*. Búzios, Setembro.
- ANTON, H., RORRES, C., 2010, *Elementary Linear Algebra: Applications*. 8 ed. New York, John Wiley & Sons.
- ATKINSON, A., DONEV, A., TOBIAS, R., 2007, *Optimum Experimental Design With SAS*. 1 ed. Oxford, Oxford University Press.
- AUDOLY, S., BELLU, G., D'ANGIÒ, L., SACCOMANI, M.P., COBELLI, C., 2001, "Global Identifiability of Nonlinear Models of Biological Systems", *IEEE transactions on bio-medical engineering*. v. 48, n. 1, pp. 55–65.
- BARD, Y., 1974, *Nonlinear Parameter Estimation*. 1 ed. New York, Academic Press.
- BELLU, G., SACCOMANI, M.P., AUDOLY, S., D'ANGIO, L., 2007, "DAISY: A New Software Tool to Test Global Identifiability of Biological and Physiological Systems", *Computer Methods and Programs in Biomedicine*. v. 88, n. 1, pp. 52–61.
- BLANCHET, F.G., LEGENDRE, P., BORCARD, D., 2008, "Forward Selection of Explanatory Variables", *Ecology*. v. 89, n. 9, pp. 2623–2632.

- BOTELHO, V.R., 2012. *Nova Metodologia para Análise de Identificabilidade e Estimação de Parâmetros de Modelos Fenomenológicos*. Dissertação de M.Sc., DEQ/UFRGS, Porto Alegre, Rio Grande do Sul, RS, Brasil.
- BRISTOL, E.H.B.E.H., 1977, "RGA 1977: Dynamic Effects of Interaction". In: *Proceedings of the 1977 IEEE Conference on Decision and Control including the 16th Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory and Applications*, New Orleans, Dezembro.
- BRUN, R., REICHERT, P., KÜNSCH, H.R., 2001, "Practical Identifiability Analysis of Large Environmental Simulation Models", *Water Resources Research*. v. 37, n. 4, pp. 1015–1030.
- CAO, Y.C.Y., ROSSITER, D., OWENS, D., 1997. "Measurement Selection for Partially Controlled Systems". In: *Proceedings of the 36th IEEE Conference on Decision and Control*, San Diego, Dezembro.
- COSTA, C.T.O.G., 2011. *Equilíbrio de Fases em Sistemas com Eletrólitos: Análise de Modelos de Energia Livre de Gibbs em Excesso*. Dissertação de M.Sc., PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- CERQUEIRA, H.S., RAWET, R., PINTO, J.C., 1999, "The Influence of Experimental Errors During Laboratory Evaluation of FCC Catalysts", *Applied Catalysis A: General*. v. 181, n. 1, pp. 209–220.
- CHASSAGNOLE, C., NOISOMMIT-RIZZI, N., SCHMID, J.W., MAUCH, C., REUSS, M., 2002, "Dynamic Modeling of the Central Carbon Metabolism of *Escherichia coli*", *Biotechnology and Bioengineering*. v. 79, n. 1, pp. 53–73.
- CHERNOFF, H., 1953, "Locally Optimal Designs for Estimating Parameters", *The Annals of Mathematical Statistics*. v. 24, n. 4, pp. 586–602.
- CHU, Y., HAHN, J., 2007, "Parameter Set Selection for Estimation of Nonlinear Dynamic Systems". *AIChE Journal*. v. 53, n. 11, pp. 2585–2870.
- CHU, Y., HUANG, Z., HAHN, J., 2009, "Improving Prediction Capabilities of Complex Dynamic Models via Parameter Selection and Estimation", *Chemical Engineering Science*. v. 64, n. 19, pp. 4178–4185.

- CHU, Y., HUANG, Z., HAHN, J., 2011, "Global Sensitivity Analysis Procedure Accounting for Effect of Available Experimental Data", *Industrial & Engineering Chemistry Research*. v. 50, n. 3, pp. 1294–1304.
- CHU, Y., HAHN, J., 2012, "Generalization of a Parameter Set Selection Procedure based upon Orthogonal Projections and the D-Optimality Criterion", *AIChE Journal*. v. 58, n. 7, pp. 2085–2096.
- COBELLI, C., DISTEFANO, J.J., 1980, "Parameter and Structural Identifiability Concepts and Ambiguities: A Critical Review and Analysis", *American Journal of Physiology*. v. 239, n. 1, pp. R7–R24.
- CUKIER, R.I., 1973, "Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. I Theory", *The Journal of Chemical Physics*. v. 59, n. 8, pp. 3873–3878.
- DRAPER, N.R., SMITH, H., 1998, *Applied Regression Analysis*. 3 ed. New York, Wiley.
- EBERHART, R., KENNEDY, J., 1995, "A New Optimizer Using Particle Swarm Theory". In: *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*. Perth, Outubro.
- ESPIE, D.M., MACCHIETTO, S., 1988, "Nonlinear Transformations for Parameter Estimation", *Industrial & Engineering Chemistry Research*. v. 27, pp. 2175–2179.
- FREITAS, K.P., 2006, "Modelagem Empírica do Processo de Deslignificação de Celulose Kraft de Eucalyptus via Redes Neurais", *Principium*, v. 1, pp. 93–101.
- FREITAS, K.P., 2009, *Ferramentas Estatísticas para o Monitoramento de Modelos Inferenciais*. Dissertação de M.Sc., PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- FREITAS, K.P., SECCHI, A.R., 2010, "Emprego da Análise de Sensibilidade Global na Identificabilidade de Parâmetros de Modelos". In: *Proceedings of the 18^o Congresso Brasileiro de Engenharia Química*. Florianópolis, Setembro.

- FREITAS, K.P., SEVERO JR, J.B., SECCHI, A.R., 2011, "Estimabilidade de Parâmetros em Bioprocessos", *In: Proceedings of the 18^o Simpósio Nacional de Bioprocessos*. Caxias do Sul, Julho.
- GOLDBERG, D.E., 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning*. 1 ed. Boston, Addison-Wesley.
- GOLUB, G.H., VAN LOAN, C.F., 1996, *Matrix Computations*. 3 ed. Baltimore, The Johns Hopkins University Press.
- GREWAL, M.S., GLOVER, K., 1976, "Identifiability of Linear and Nonlinear", *IEEE Transactions on Automatic Control*. pp. 833–837.
- HISKENS, I.A., 2001, "Disturbance Measurements", *IEEE Transactions on Power Systems*. v. 16, n. 4, pp. 702–710.
- HOMMA, T., SALTELLI, A., 1996, "Importance Measures in Global Sensitivity Analysis of Nonlinear Models", *Reliability Engineering System Safety*. v. 52, n. 1, pp. 1–17.
- HOQUE, M.A., USHIYAMA, H., TOMITA, M., SHIMIZU, K., 2005, "Dynamic Responses of the Intracellular Metabolite Concentrations of the Wild Type and *pykA* Mutant *Escherichia coli* Against Pulse Addition of Glucose or NH₃ Under Those Limiting Continuous Cultures", *Biochemical Engineering Journal*. v. 26, n. 1, pp. 38–49.
- HORA, S.C., IMAN, R.L., 1990, "Bayesian Modeling of Initiating Event Frequencies at Nuclear Power Plants 1", *Risk Analysis*. v. 10, n. 1, pp. 103–109.
- HUG, S., RAUE, A., HASENAUER, J., BACHMANN, J., KLINGMULLER, U., TIMMER, J., THEIS, F.J., 2013, "High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling", *Mathematical biosciences*. v. 246, pp. 293–304.
- JACQUEZ, J.A., 1987, "Identifiability: the first step in parameter estimation", *Federation Proceedings*. v. 46, n. 8, pp. 2477–2480.
- KIEFER, J., 1959, "Optimal experimental designs", *Journal of the Royal Statistical Society Series B - Statistical Methodology*. v. 21, n. 2, pp. 272–319.
- KIEFER, J., 1975, "Optimal Design: Variation in Structure and Performance Under Change of Criterion", *Biometrika*. v. 62, n. 2, pp. 277–288.

- KIEFER, J., WOLFOWITZ, J., 1960, "The equivalence of two extreme problems", *Canadian Journal of Mathematics*. v. 12, pp. 363–336.
- KOU, B., McAULEY, K.B., HSU, C.C., BACON, D.W., 2005a, "Mathematical Model and Parameter Estimation for Gas-phase Ethylene/Hexene Copolymerization with Supported Metallocene Catalyst", *Macromolecular Materials and Engineering*. v. 290, pp. 537–557.
- KOU, B., McAULEY, K.B., HSU, C.C., BACON, D.W., YAO, K.Z., 2005b, "Mathematical Model and Parameter Estimation for Gas-phase Ethylene Homopolymerization with Supported Metallocene Catalyst", *Industrial & Engineering Chemistry Research*. v. 44, pp. 2428–2442.
- KREUTZ, C., RAUE, A., KASCHEK, D., TIMMER, J., 2013, "Profile Likelihood in Systems Biology", *The FEBS journal*. v. 280, n. 11, pp. 2564–71.
- LANGMUIR, I., 1916, "The Constitution and Fundamental Properties of Solids and Liquids. Part I. Solids", *Journal of the American Chemical Society*. v. 38, n. 11, pp. 2221–2295.
- LARENTIS, A.L., BENTES JR., A.M.P., DE RESENDE, N.S., SALIM, V.M., PINTO, J.C., 2003, "Analysis of experimental errors in catalytic tests for production of synthesis gas", *Applied Catalysis A: General*. v. 242, n. 2, pp. 365–379.
- LEE, C.H., HERGET, C.J., 1975, "Regions of constrained maximum likelihood parameter identifiability", *IEEE Transactions on Automatic Control*. v. 21, n. 4.
- LI, R., HENSON, M.A., KURTZ, M.J., 2004, "Selection of Model Parameters for Off-Line Parameter Estimation", *IEEE Transactions on Control Systems Technology*. v. 12, n. 3, pp. 402–412.
- LONGUI, L.G.S., LUZZIVETO, D.J., FERREIRA, L.S., RECH, R., AYUB, M.A.Z., SECCHI, A.R., 2004, "A Kinetic Model for the *Kluyveromyces marxianus* growth using cheese whey as substrate", *Journal of Industrial Microbiology and Biotechnology*. v. 31, n. 1, pp. 35–40.
- LJUNG, L., GLAD, T., 1994, "On Global Identifiability for Arbitrary Model Parametrizations", *Automatica*. v. 30, n. 2, pp. 265–276.
- LUND, B.F., FOSS, B.A., 2008, "Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models", *Automatica*. v. 44, n. 1, pp. 278–281.

- LUZIVETTO, D.J., CARDOZO, N.S.M., SECCHI, A.R., 2006. "Nonlinear Model Reduction Using Identifiability Measures". *In: Proceedings of the Workshop on Solving Industrial Control and Optimization Problems*. Gramado, Abril.
- MACHADO, V.C., TAPIA, G., GABRIEL, D., LAFUENTE, J., BAEZA, J.A., 2009, "Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model", *Environmental Modelling Software*. v. 24, n. 11, pp. 1274–1284.
- MARGARIA, G., RICCOMAGNO, E., CHAPPELL, M.J., WYNN, H., 2001, "Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences", *Mathematical Biosciences*. v. 174, n. 1, pp. 1–26.
- MARIMONT, R.B., SHAPIRO, M.B., 1979, "Nearest Neighbour Searches and the Curse of Dimensionality", *IMA Journal of Applied Mathematics*. v. 24, n. 1, pp. 59–70.
- MCLEAN, K.A.P., WU, S., MCAULEY, K.B., 2012, "Mean-Squared-Error Methods for Selecting Optimal Parameter Subsets for Estimation", *Industrial & Engineering Chemistry Research*. v. 51, n. 17, pp. 6105–6115.
- MICHAELIS, L., MENTEN, M.L., 1913, "Die Kinetik der Invertinwirkung", *Biochemische Zeitschrift*. v. 49, pp. 333–369.
- MONOD, J., 1942, *Recherches sur la Croissance des Cultures Bactériennes*. Paris, Hermann.
- MORAD, K., SVRCEK, W.Y., MCKAY, I., 1999, "A Robust Direct Approach for Calculating Measurement Error Covariance Matrix". *In: Proceedings of the 1999 American Control Conference*. San Diego, Junho.
- MORRIS, M.D., 1991, "Factorial Sampling Plans for Preliminary Computational Experiments", *Technometrics*. v. 33, n. 2, pp. 161–174.
- MURPHY, S.A.; VAN DER VAART, A.W., 2000, "On Profile Likelihood", *Journal of the American Statistical Association*. v. 95, n. 450, pp. 449–465.
- NGUYEN, V.V, WOOD, E.F., 1982, "Review and Unification of Linear Identifiability Concepts", *SIAM Review*. v. 24, n. 1, pp. 34–51.

- OMLIN, M., BRUN, R., REICHERT, P., 2001, "Biogeochemical Model of Lake Zürich: Sensitivity, Identifiability and Uncertainty Analysis", *Ecological Modelling*. v. 141, pp. 105–123.
- OMLIN, M., REICHERT, P., 1999, "A Comparison of Techniques for the Estimation of Model Prediction Uncertainty", *Ecological Modelling*. v. 115, n. 1, pp. 45–59.
- PETZOLD, L.R., 1982, "A description of Dassl: A Differential/Algebraic System Solver". In: *Proceedings of the 10th International Association for Mathematics and Computers in Simulation World Congress*. Montreal, Agosto.
- PINTO, J.C., LOBÃO, M.W., MONTEIRO, J.L., 1990, "Sequential Experimental Design for Parameter Estimation: A Different Approach", *Chemical Engineering Science*. v. 45, n. 4, pp. 883–892.
- POHJANPALO, H., 1978, "System Identifiability Based on the Power Series Expansion of the Solution", *Mathematical Biosciences*. v. 41, n. 1-2, pp. 21–33.
- PRATA, D.M., 2009, *Reconciliação robusta de dados para monitoramento em tempo real*. Tese de doutorado. Tese de D.Sc., PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- RABITZ, H., 1989, "Systems Analysis at the Molecular Scale", *Science*. v. 246, n. 4927, pp. 221–226.
- RAUE, A., KREUTZ, C., MAIWALD, T., BACHMANN, J., SCHILLING, M., KLINGMULLER, U., TIMMER, J., 2009, "Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood", *Bioinformatics*. v. 25, n. 15, pp. 1923–9.
- RAWET, R., CERQUEIRA, H.S., PINTO, J.C., 2001, "The Influence of Covariances During Laboratory Evaluation of FCC Catalysts", *Applied Catalysis A: General*. v. 207, n. 1-2, pp. 199–209.
- REID, D., 1979, "An Algorithm for Tracking Multiple Targets", *IEEE Transactions on Automatic Control*. v. 24, n. 6, pp. 843–854.
- ROMAGNOLI, J.A., SANCHEZ, M.C., 1999, *Data Processing and Reconciliation for Chemical Process Operations*. New York, Academic Press.

- RUSNAK, I., GUEZ, A., BAR-KANA, I., 1983, "Simultaneous State Observability and Parameters Identifiability of Discrete Stochastic Linear Systems". In: *Proceedings of the American Control Conference*. San Francisco, Junho.
- SACCOMANI, M.P., AUDOLY, S., BELLU, G., 1998, "Global Identifiability of Nonlinear Model Parameters". In: *Proceedings of the American Control Conference*. Kitakyushu, Julho.
- SACCOMANI, M.P., AUDOLY, S., BELLU, G., D'ANGIO, L., 2010, "Examples of Testing Global Identifiability of Biological and Biomedical Models with the DAISY Software", *Computers in biology and medicine*. v. 40, n. 4, pp. 402–7.
- SACCOMANI, M.P., AUDOLY, S., D'ANGIÒ, L., 2003, "Parameter identifiability of nonlinear systems: the role of initial conditions", *Automatica*. v. 39, n. 4, pp. 619–632.
- SALTELLI, A., 1999, "Sensitivity Analysis: Could Better Methods be Used?", *Journal of Geophysical Research*. v. 104, n. D3, pp. 3789–3793.
- SALTELLI, A., 2002, "Making Best Use of Model Evaluations to Compute Sensitivity Indices", *Computer Physics Communications*. v. 145, n. 2, pp. 280–297.
- SALTELLI, A., RATTO, M., ANDRES, T., CAMPOLONGO, F., CARIBONI, J., GATELLI, D., SAISANA, M., TARANTOLA, S., 2008, *Global Sensitivity Analysis: The Primer*. New York, John Wiley & Sons.
- SALTELLI, A., TARANTOLA, S., CAMPOLONGO, F., RATTO, M., 2004, *Sensitivity analysis in practice: A guide to assessing scientific models*. West Sussex, John Wiley & Sons.
- SANDINK, C.A., MCAULEY, K.B., MCLELLAN, P.J., 2001, "Selection of Parameters for Updating in On-line Models", *Industrial & Engineering Chemistry Research*. v. 40, n. 18, pp. 3936–3950.
- SANTOS, T.J.; PINTO, J.C., 1998, "Taking Variable Correlation into Consideration During Parameter Estimation", *Brazilian Journal of Chemical Engineering*. v. 15, pp. 1–20.
- SCHWAAB, M., 2005, *Avaliação de Algoritmos Heurísticos de Otimização em Problemas de Estimção de Parâmetros*. Dissertação de M.Sc., PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

- SCHWAAB, M., 2007, *Desenvolvimento e implementação de novas técnicas de estimação de parâmetros e planejamento sequencial de experimentos*. Tese de D.Sc., PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- SCHWAAB, M., ALBERTON, A.L., PINTO, J.C., 2011, *ESTIMA&PLANEJA: Pacote computacional para estimação de parâmetros e planejamento de experimentos para discriminação de modelos*. Relatório Técnico, PEQ-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- SCHWAAB, M., BISCAIA, JR., E.C., MONTEIRO, J.L., PINTO, J.C., 2008, "Nonlinear parameter estimation through particle swarm optimization", *Chemical Engineering Science*. v. 63, n. 6, pp. 1542–1552.
- SCHWAAB, M., PINTO, J.C., 2007, *Análise de Dados Experimentais: I. Fundamentos de Estatística e Estimação de Parâmetros*. 1 ed. Rio de Janeiro, e-papers.
- SECCHI, A.R., ALMEIDA NETO, E., FINKLER, T.F., 2006, "An Algorithm for Automatic Selection and Estimation of Model". In: *Proceedings of the International Symposium on Advanced Control of Chemical Processes*. Gramado, Abril.
- SOBOL, I.M., 1994, *A Primer for the Monte Carlo Method*. ed. 1. Boca Raton, CRC Press LLC.
- SOBOL, I.M., 1990, "On Sensitivity Estimation for Nonlinear Mathematical Models", *Matematicheskoe Modelirovanie*. v. 2, n. 1, pp. 112-118.
- SUN, C., HAHN, J., 2006, "Parameter Reduction for Stable Dynamical Systems Based on Hankel Singular Values and Sensitivity Analysis", *Chemical Engineering Science*. v. 61, n. 16, pp. 5393–5403.
- THOMPSON, D.E., MCAULEY, K.B., MCLELLAN, P.J., 2009, "Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst", *Macromolecular Reaction Engineering*. v. 3, n. 4, pp. 160–177.
- TOJAL, T.S., ALBERTON, K.P.F., BISCAIA JR, E.C., SECCHI, A.R., 2013, "Seleção e Estimação Automática de Parâmetros de um Modelo de Polimerização Radicalar Controlada de Estireno com MMS e MMA Mediada por Nitróxido em Reator Tubular". In: *Proceedings of the 12^o Congresso Brasileiro de Polímeros*. Florianópolis, Setembro.

- VAJDA, S., RABITZ, H., WALTER, E., LECOURTIER, Y., 1989, "Qualitative and Quantitative Identifiability Analysis of Nonlinear Chemical Kinetic Models", *Chemical Engineering Communications*. v. 83, n. 1, pp. 191–219.
- VENZON, D.J., MOOLGAVKAR, S.H., 1988, "A Method for Computing Profile-Likelihood-Based Confidence Intervals", *Applied Statistics*. v. 37, n. 1, pp. 87–94.
- VERGHESE, GEORGE C.; VÉLEZ-REYES, M., 1999, "Subset Selection for Improved Parameter Estimation in On-Line Identification of a Synchronous Generator", *IEEE Transactions on Power Systems*. v. 14, n. 1, pp. 218–225.
- WATTS, D.G., 1994, "Estimating Parameters in Nonlinear Rate Equations", *Canadian Journal of Chemical Engineering*. v. 72, pp. 701–710.
- WEIJERS, S.R., VANROLLEGHEM, P.A., 1997, "A Procedure for Selecting Best Identifiable Parameters in Calibrating Activated Sludge Model no.1 to Full-scale Plant Data", *Water Science and Technology*. v. 36, n. 5, pp. 69–79.
- WHITE, L.J., EVANS, N.D., LAM, T.J., SCHUKKEN, Y.H., MEDLEY, G.F., GODFREY, K.R., CHAPPELL, M.J., 2001, "The Structural Identifiability and Parameter Estimation of a Multispecies Model for the Transmission of Mastitis in Dairy Cows", *Mathematical Biosciences*. v. 174, n. 2, pp. 77–90.
- WU, S., HARRIS, T.J., MCAULEY, K.B., 2007, "The Use of Simplified or Misspecified Models: Linear Case", *Canadian Journal of Chemical Engineering*. v. 85, n. 4, pp. 386–398.
- WU, S., MCAULEY, K.B., HARRIS, T.J., 2011, "Selection of Simplified Models: I. Analysis of Model-selection Criteria Using Mean-squared Error", *The Canadian Journal of Chemical Engineering*. v. 89, n. 1, pp. 148–158.
- YAO, K.Z., SHAW, B.M., KOU, B., MCAULEY, K.B., BACON, D., 2003, "Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design", *Polymer Reaction Engineering*. v. 11, n. 3, pp. 563–588.
- YU, W., HARRIS, T.J., 2009, "Parameter Uncertainty Effects on Variance-based Sensitivity Analysis", *Reliability Engineering & System Safety*. v. 94, n. 2, pp. 596–603.

YUE, H., BROWN, M., KNOWLES, J., WANG, H., BROOMHEAD, D.S., KELL, D.B., 2006, "Insights into the Behaviour of Systems Biology Models from Dynamic Sensitivity and Identifiability Analysis: A Case Study of an NF-Kappab Signalling Pathway", *Molecular Biosystems*. v. 2, n. 12, pp. 640–9.