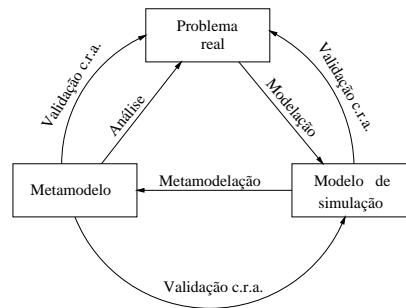


UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO



**Construção de Metamodelos de
Regressão Não Linear para
Simulação de Acontecimentos Discretos**

Maria Isabel da Conceição Santos Reis dos Santos
(MESTRE)

Dissertação para a obtenção do Grau de Doutor em
Engenharia e Gestão Industrial

Orientador: Doutor Acácio Manuel de Oliveira Porta Nova

Presidente: Reitor da Universidade Técnica de Lisboa

Vogais: Doutora Maria Teresa Ferreira Soares Mendes
Doutora Maria Fernanda Neto Ramalhoto
Doutor Luís António Tadeu dos Santos Almeida
Doutor António José Lopes Rodrigues
Doutor Acácio Manuel de Oliveira Porta Nova

Dezembro, 2002

Título: Construção de Metamodelos de Regressão Não Linear para Simulação de Acontecimentos Discretos

Nome: Maria Isabel da Conceição Santos Reis dos Santos

Doutoramento em: Engenharia e Gestão Industrial

Orientador: Acácio Manuel de Oliveira Porta Nova

Provas concluídas em:

Resumo

Neste trabalho, propõe-se uma metodologia para construir metamodelos não lineares de regressão no contexto da simulação de acontecimentos discretos. Esta metodologia inclui um procedimento detalhado para obtenção de resultados para metamodelação, selecção do tipo de metamodelo, estimação (estimação de parâmetros, construção de regiões de confiança e testes de hipóteses) e validação de metamodelos. Tanto na parte da estimação, como na da validação, são propostas técnicas estatísticas da análise de regressão (algumas das quais tiveram de ser adaptadas ao caso dos metamodelos não lineares). O uso desta metodologia conduz, geralmente, a funções não lineares simples que aproximam adequadamente o comportamento de certos sistemas, que os modelos de regressão polinomial não conseguem aproximar convenientemente. Isto é confirmado através de dois exemplos: um centro de inspecção e reparação de automóveis e um modelo de filas de espera $M/M/1$.

Para melhorar a eficiência do processo de estimação, reduzindo a variância dos parâmetros estimados, utiliza-se a técnica das variáveis de controlo. Nesta perspectiva, apresentam-se procedimentos para estimação em que algumas das variáveis independentes do metamodelo são variáveis de controlo. Assumindo a normalidade conjunta da resposta e das variáveis de controlo, obtêm-se regiões de confiança e testes de hipóteses. Também se apresentam medidas para avaliar a eficiência obtida ao utilizar variáveis de controlo.

Palavras-chaves: Simulação de acontecimentos discretos; metamodelos não lineares; validação de metamodelos; técnicas de redução de variância; variáveis de controlo.

Abstract

This thesis proposes a methodology for building nonlinear regression metamodels in discrete event simulation experiments. This methodology includes detailed procedures for analyzing the simulation model output, selecting the metamodel type and estimating and validating the nonlinear simulation metamodel. The procedures for estimating metamodel parameters, building confidence regions, testing hypotheses and validating the estimated metamodel use statistical techniques from regression analysis, some of which had to be adapted to the nonlinear metamodel case. The proposed methodology produces, in general, simple nonlinear functions that adequately approximate the behaviour of systems for which polynomial regression models are inadequate. This is emphasized through two illustrative examples: an automobile inspection and repair centre and an M/M/1 queueing model.

In order to improve the efficiency of the estimation process, control variables are used to reduce the variance of the estimated parameters, taking advantage of readily available concomitant output variables. Under the assumption of joint normality of the response and the controls, confidence regions and hypotheses tests are obtained. A set of measures to quantify the efficiency achievable by the use of control variates is also included.

Keywords: Discrete event simulation; nonlinear metamodels; nonlinear regression; metamodel validation; variance reduction techniques; control variables.

Agradecimentos

Pela contribuição dada para a realização deste trabalho, pretendo expressar o meu reconhecimento muito particular a algumas pessoas.

Ao Prof. Acácio Porta Nova na qualidade de meu orientador científico e, muito especialmente, pela sua constante disponibilidade e encorajamento. O seu acompanhamento empenhado tem sido fundamental para a minha formação científica e profissional.

Aos Professores João Cunha e Serra e Joaquim Sampaio Cabral pelo incentivo e alento dispensados, muito em particular nos momentos mais delicados deste trabalho.

Ao Prof. James Wilson pelos comentários e sugestões numa parte deste trabalho.

Ao Prof. Guilherme Arroz, à Andreia e ao Natalino por todo o apoio e encorajamento que me deram durante a realização deste trabalho.

Por fim, uma referência muito especial àqueles que comigo padeceram e pacientemente me apoiaram e a quem dedico esta dissertação, ao Pedro, ao Gonçalo e ao Henrique.

Lisboa, Dezembro de 2002

Conteúdo

1	Introdução	1
1.1	Enquadramento em simulação	2
1.2	Métodos para análise de resultados	5
1.3	Motivação e objectivos	12
1.4	Organização da dissertação	13
2	Estudos importantes sobre metamodelos	17
2.1	Metamodelos de simulação	18
2.1.1	Análise comparativa	27
2.2	Validação de metamodelos de regressão	32
2.3	Técnicas de redução de variância	42
2.3.1	Análise comparativa	48
2.4	Método das variáveis de controlo	50

3	Metodologia para metamodelação	59
3.1	Processo de metamodelação	63
3.2	Estimação de metamodelos	74
3.2.1	Estimador dos mínimos quadrados	75
3.2.2	Regiões de confiança	81
3.2.3	Testes de hipóteses	84
4	Validação de metamodelos	85
4.1	Adequação do metamodelo	86
4.1.1	Hipóteses distribucionais	87
4.1.2	Falta de ajustamento	88
4.1.3	Capacidade previsiva	92
4.2	Validação com respeito ao modelo de simulação	109
4.3	Validação com respeito ao problema real	112
5	Variáveis de controlo	113
5.1	Metamodelos com variáveis de controlo	114
5.2	Resultados gerais sobre estimação	117
5.2.1	Razão mínima entre variâncias generalizadas	117
5.2.2	Estimador controlado dos mínimos quadrados	120
5.3	Resultados para metamodelos normais	121
5.3.1	Distribuição do estimador controlado	122
5.3.2	Razão entre variâncias e factor de perda	126
5.3.3	Regiões de confiança	128
5.3.4	Testes de hipóteses	130

5.4	Razão entre variâncias e factor de perda experimentais	131
6	Aplicação	135
6.1	Centro de inspecção e reparação de automóveis	136
6.2	Modelo de filas de espera $M/M/1$	146
6.2.1	Método das variáveis de controlo	154
7	Conclusões e recomendações	161
7.1	Sumário de investigação	161
7.2	Trabalho futuro	164
	APÊNDICES	167
A	Catálogo de curvas	167
	Bibliografia	177

Notação

Nesta dissertação, iremos ter de lidar com *escalares*, *vectores* e *matrizes aleatórios* e *determinísticos*, assim como com os seus *valores observados* que são aleatórios. Além disso, também iremos ter *estimadores* de *parâmetros*. A notação que iremos utilizar é a seguinte:

- Os *escalares* serão denotados com letras minúsculas (latinas ou gregas). Uma exceção a esta regra é o caso das *variáveis aleatórias*, que serão sempre identificadas com letras latinas maiúsculas, quer sejam escalares, vectores ou matrizes.
- Os *vectores* serão representados por letras minúsculas a **negrito**.
- As *matrizes* serão identificadas por letras maiúsculas a **negrito**.
- As *variáveis aleatórias* serão sempre denotadas com letras latinas maiúsculas. Os vectores e matrizes aleatórios irão estar a **negrito**. Iremos distinguir entre valores de vectores e matrizes aleatórios por menção ou contexto.
- A distinção entre os *valores observados* das variáveis aleatórias e as próprias variáveis aleatórias irão ser por contexto ou menção explícita.
- Os *parâmetros* nos modelos serão, geralmente, representados por letras gregas minúsculas, minúsculas a **negrito** e maiúsculas a **negrito** consoante sejam respectivamente *escalares*, *vectores* ou *matrizes*.
- Os *estimadores* de cada um dos parâmetros serão denotados pelas letras gregas correspondentes escoradas superiormente por um acento circunflexo, ou um til, ou um ponto.
- Ao longo de toda a dissertação utilizam-se os seguintes símbolos tal como descrito na tabela abaixo:

símbolo	significado
n	número de pontos experimentais
r	número de réplicas em cada ponto experimental
N	número total de execuções do programa de simulação ($N = n \times r$)
m	número de parâmetros desconhecidos do metamodelo (dimensão de θ)
q	número de variáveis de controlo
d	número de variáveis de decisão

- As distribuições de probabilidade irão ser representadas pelos símbolos usuais (χ^2 , N , F , W , etc.). Se X e Y são variáveis aleatórias, então os símbolos $E[X]$, $\text{Var}[X]$ e $\text{Cov}[X]$ representam o valor esperado, a variância e a covariância, respectivamente. Se \mathbf{X} e \mathbf{Y} forem vectores aleatórios, então tem-se analogamente $\mathcal{E}[\mathbf{X}]$ para o valor esperado do vector \mathbf{X} , $\mathcal{D}[\mathbf{X}]$ para a matriz de dispersão, ou de variância-covariância, de \mathbf{X} e $\mathcal{C}[\mathbf{X}, \mathbf{Y}]$ para a covariância entre os dois vectores \mathbf{X} e \mathbf{Y} . A covariância entre uma variável aleatória X e um vector \mathbf{Y} representa-se por $\mathcal{C}[X, \mathbf{Y}]$.

- A notação para entidades relacionadas com uma matriz \mathbf{A} é definida como se segue:

A_{ij} – elemento da matriz que ocupa a i -ésima linha e a j -ésima coluna;

$\mathbf{A}_{.i}$ – i -ésima linha de \mathbf{A} ;

$\mathbf{A}_{.j}$ – j -ésima coluna de \mathbf{A} .

Sejam \mathbf{A} e \mathbf{B} matrizes $n \times r$ e $r \times l$ respectivamente, e seja \mathbf{AB} o seu produto. Então $[\mathbf{AB}]_{ij}$ representa o elemento da matriz \mathbf{AB} que ocupa a i -ésima linha e a j -ésima coluna.

A *transposta* de uma matriz \mathbf{A} é representada por \mathbf{A}^T e os seus elementos são dados por $A_{ij}^T = A_{ji}$. Se a matriz \mathbf{A} for quadrada, então o seu *determinante* representa-se por $|\mathbf{A}|$ e se \mathbf{A} é não singular $|\mathbf{A}| \neq 0$. A sua matriz *inversa* designa-se por \mathbf{A}^{-1} . Uma matriz $r \times r$ com elementos na diagonal principal d_i , $i = 1, \dots, r$, e com os outros elementos nulos, representa-se por $\mathbf{A} = \text{diag}[d_1, \dots, d_r]$; quando $d_i = 1$, $i = 1, \dots, r$, então temos a matriz identidade \mathbf{I}_r .

- O *produto de Kronecker* à direita da matriz \mathbf{A} , $n \times r$, pela matriz \mathbf{B} , $m \times q$, define-se por

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & \dots & A_{1r}\mathbf{B} \\ \vdots & & \vdots \\ A_{n1}\mathbf{B} & \dots & A_{nr}\mathbf{B} \end{bmatrix}. \quad (1)$$

- Suponha-se $f(\mathbf{X}_{i.}, \boldsymbol{\theta})$ uma função com valores reais e com argumento m -dimensional $\boldsymbol{\theta}$. O vector gradiente de $f(\mathbf{X}_{i.}, \boldsymbol{\theta})$ é denotado por

$$\mathbf{f}_i(\boldsymbol{\theta}) = \left(\frac{\partial f}{\partial \theta_1}(\mathbf{X}_{i.}, \boldsymbol{\theta}), \dots, \frac{\partial f}{\partial \theta_m}(\mathbf{X}_{i.}, \boldsymbol{\theta}) \right)^T$$

e é um vector $m \times 1$. O elemento

$$\frac{\partial f}{\partial \theta_j}(\mathbf{X}_{i.}, \boldsymbol{\theta})$$

é a derivada parcial de $f(\mathbf{X}_{i.}, \boldsymbol{\theta})$ em relação ao j -ésimo parâmetro do vector $\boldsymbol{\theta}$ ($j = 1, \dots, m$). O símbolo T na definição de $\mathbf{f}_i(\boldsymbol{\theta})$ significa *transposto*.

Seja $\mathbf{f}(\boldsymbol{\theta})$ o vector $n \times 1$

$$\mathbf{f}(\boldsymbol{\theta}) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}))^T.$$

A matriz Jacobiana de \mathbf{f} no ponto $\boldsymbol{\theta}$ denota-se por

$$\mathbf{F}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial f}{\partial \theta_1}(\mathbf{X}_{1.}, \boldsymbol{\theta}) & \frac{\partial f}{\partial \theta_2}(\mathbf{X}_{1.}, \boldsymbol{\theta}) & \dots & \frac{\partial f}{\partial \theta_m}(\mathbf{X}_{1.}, \boldsymbol{\theta}) \\ \frac{\partial f}{\partial \theta_1}(\mathbf{X}_{2.}, \boldsymbol{\theta}) & \frac{\partial f}{\partial \theta_2}(\mathbf{X}_{2.}, \boldsymbol{\theta}) & \dots & \frac{\partial f}{\partial \theta_m}(\mathbf{X}_{2.}, \boldsymbol{\theta}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial \theta_1}(\mathbf{X}_{n.}, \boldsymbol{\theta}) & \frac{\partial f}{\partial \theta_2}(\mathbf{X}_{n.}, \boldsymbol{\theta}) & \dots & \frac{\partial f}{\partial \theta_m}(\mathbf{X}_{n.}, \boldsymbol{\theta}) \end{bmatrix}.$$

Nesta notação podemos escrever

$$\mathbf{F}(\boldsymbol{\theta}) = (\mathbf{f}_1^T(\boldsymbol{\theta}), \dots, \mathbf{f}_n^T(\boldsymbol{\theta}))^T.$$

A matriz que se obtém ao eliminar o vector $\mathbf{f}_i(\boldsymbol{\theta})$ à matriz Jacobiana $\mathbf{F}(\boldsymbol{\theta})$ denota-se por $\mathbf{F}_{(-i)}(\boldsymbol{\theta})$. Representa-se por $\mathbf{f}_{(-i)}(\boldsymbol{\theta})$ o vector que se obtém quando se elimina a componente $f(\mathbf{X}_{i.}, \boldsymbol{\theta})$ ao vector $\mathbf{f}(\boldsymbol{\theta})$. De uma forma geral denota-se por $\mathbf{x}_{(-i)}$ o vector que se obtém ao eliminar a componente x_i ao vector $\mathbf{x} = (x_1, \dots, x_n)$.

Lista de Figuras

2.1	Conceito de variável de controlo.	51
3.1	Metamodelo, modelo de simulação e problema real.	61
3.2	Modelo de Simulação.	62
3.3	Processo de metamodelação.	64
3.4	Cálculo da sequência de médias móveis com $w = 2$ e baseado em r réplicas.	67
3.5	Médias móveis para vários valores da janela de Welch: $w = 10, 30, 50, 100$	68
3.6	Algumas relações funcionais não lineares.	71
6.1	Centro de inspeção e reparação de automóveis.	137
6.2	Modelo na linguagem SLAM II (centro de inspeção e reparação de automóveis).	138
6.3	Para $\mu = 10$, média móvel com janela de Welch $w = 50$ baseada em 20 réplicas, onde cada réplica contém 2000 observações.	139
6.4	Resultados da simulação do centro de inspeção e reparação de automóveis.	140
6.5	Amostras da resposta em papel de probabilidade normal: quantil normal $\Phi^{-1}[(j - 0.5)/r]$ versus a j -ésima estatística de ordem $Y_{i(j)}$ da resposta (pontos experimentais 2, 6, 10 e 14).	143
6.6	Gráficos dos resíduos ponderados (metamodelo 'MMF').	144

6.7	Papel de probabilidade normal (metamodelo 'MMF'): $\Phi^{-1}[(j-0.5)/r]$ versus o j -ésimo resíduo ponderado ordenado.	144
6.8	Resultados da simulação do sistema $M/M/1$	147
6.9	Monómio com $\theta_1 = 25$ e $\theta_2 = 5$	148
6.10	Hipérbole com $\theta_1 = 1$ e $\theta_2 = -1$	148
6.11	Papel de probabilidade normal (hipérbole): $\Phi^{-1}[(j-0.5)/r]$ versus o j -ésimo resíduo ponderado ordenado.	152
6.12	Gráficos dos resíduos ponderados (hipérbole).	152
A.1	Curvas exponenciais.	169
A.2	Curvas com potência de x desconhecida.	170
A.3	Curvas com x como potência.	170
A.4	Curvas com funções trigonométricas.	171
A.5	Algumas funções racionais.	172
A.6	Curvas de crescimento de Sigmóide.	173
A.7	Curvas de crescimento de Sigmóide (continuação).	174
A.8	Curvas de densidade de produção	175
A.9	Função logarítmica e algumas curvas combinadas.	175

Lista de Tabelas

3.1	Disposição dos resultados para estimação de metamodelos.	69
5.1	Disposição dos resultados para estimação de metamodelos, com variáveis de controlo.	115
6.1	Detecção do viés inicial.	140
6.2	Curvas não lineares candidatas.	141
6.3	Diagnósticos do metamodelo (centro de inspeção e reparação de automóveis).	142
6.4	Teste de Rao.	144
6.5	Teste de validação cruzada dupla, para o metamodelo 'MMF'.	145
6.6	Intervalos de confiança de 95% para os parâmetros do metamodelo 'MMF' (centro de inspeção e reparação de automóveis).	145
6.7	Detecção do viés inicial.	146
6.8	Curvas candidatas.	147
6.9	Resultados da estimação.	150
6.10	Diagnósticos do metamodelo (sistema $M/M/1$).	150
6.11	Um método de validação cruzada.	151
6.12	Teste de falta de ajustamento (sistema $M/M/1$).	151

6.13	Teste de validação cruzada dupla, para o metamodelo hiperbólico.	152
6.14	Solução analítica <i>versus</i> metamodelo hiperbólico.	153
6.15	Intervalos de confiança de 95% para os parâmetros (sistema $M/M/1$).	154
6.16	Coeficientes de regressão estimados (sistema $M/M/1$).	155
6.17	Teste de falta de ajustamento (sistema $M/M/1$, com variáveis de controlo).	156
6.18	Teste de validação cruzada dupla, para o metamodelo controlado.	156
6.19	Solução analítica e metamodelo controlado: uma comparação.	157
6.20	Intervalos de confiança de 95% para os parâmetros do metamodelo hiperbólico (sistema $M/M/1$, com variáveis de controlo).	158
6.21	Estimativas das razões entre variâncias e do factor de perda.	159
A.1	Catálogo de curvas.	168

Capítulo 1

Introdução

Os modelos de simulação digital são frequentemente utilizados para tomar decisões sobre modificações a fazer no desenho de sistemas reais. Eles são utilizados como substitutos do sistema real porque, ou é impraticável construir múltiplas versões protótipo do sistema real, ou têm custos elevados, ou têm outras limitações proibitivas de experimentação com o sistema real. Contudo, um modelo de simulação, embora seja mais simples que o sistema real, pode ser bastante complicado e, por conseguinte, é comum construírem-se modelos ainda mais simples para os aproximar — metamodelos de simulação. A forma simples de um metamodelo pode revelar as características gerais importantes do comportamento do modelo de simulação. Foi Kleijnen [60] quem começou, em 1987, a utilizar o termo 'metamodelo' mas os metamodelos têm sido utilizados em simulação ao longo dos últimos trinta anos. Actualmente continua a existir um grande interesse em investigação nesta área (ver, por exemplo, [19], [8] e [65]). A experimentação com o programa de simulação permite estimar e validar um metamodelo de simulação para que ele possa ser utilizado com fiabilidade. Contudo, simulação directa de um sistema complexo pode ser caro, exigindo comprimentos grandes das simulações para obter precisões aceitáveis nos estimadores dos parâmetros desconhecidos do metamodelo. Um meio de melhorar a precisão desses estimadores é utilizar técnicas de redução de variância (ver [110], [27] e [102]).

1.1 Enquadramento em simulação

Um dos pioneiros dos conceitos da simulação foi John Von Neumann. Nos anos 1940's ele concebeu a ideia de executar múltiplas repetições de um modelo, reunir resultados estatísticos e concluir sobre o comportamento do sistema real, baseado nesse modelo. Mais tarde isto veio a ser conhecido como método de Monte Carlo devido à utilização de variáveis geradas aleatoriamente para representar comportamentos que não podem ser modelados exactamente, mas podem ser caracterizados estatisticamente. Von Neumann utilizou este método para estudar as acções aleatórias dos neutrões e a eficácia dos bombardeamentos de aeronaves. Não passou muito tempo e foram encontradas aplicações civis deste método na representação de fábricas com o objectivo de determinar a produtividade potencial máxima.

Actualmente a palavra *simulação* é utilizada com uma vasta gama de significados, que vão desde *imitação* ou *fingimento* até *resolução de equações diferenciais*. Nesta dissertação, entende-se simulação como sendo *a construção de um modelo matemático e lógico de um sistema e sua experimentação num computador digital* [96]. Mais precisamente, vai-se tratar de simulações com acontecimentos discretos, porque todos os modelos de simulação considerados serão discretos (o estado do sistema muda apenas num conjunto finito, ou numerável, de instantes no tempo, denominados acontecimentos discretos), dinâmicos (vai ser estudada a sua evolução no tempo) e de natureza estocástica (de um modo geral, eles contêm uma ou mais variáveis aleatórias).

A simulação tem tido uma crescente aceitação e utilização em investigação operacional, bem como em muitos outros ramos da ciência, devido a algumas razões que são discriminadas em seguida.

1. A maior parte dos problemas reais são muito complicados não sendo possível tratá-los através da utilização de métodos analíticos, ou mesmo de métodos numéricos.
2. Os computadores são cada vez mais rápidos, têm cada vez mais memória e existem pacotes de programas com mais e melhores realizações de métodos numéricos. Isto torna possível levar a cabo estudos de alguns sistemas em larga escala, que há algum

tempo atrás não seriam possíveis. No caso de estudos que já se faziam há alguns anos, torna-se possível obter resultados mais precisos e mais rapidamente do que anteriormente.

3. O desenvolvimento de diversas linguagens de simulação tem ajudado o trabalho dos utilizadores na construção dos modelos de simulação e sua experimentação.

Realizar execuções do programa de simulação para um grande número de combinações de variáveis de entrada pode ser, contudo, muito caro e a interpretação da quantidade de resultados produzidos ser muito difícil. Uma forma de ultrapassar este inconveniente é, sempre que possível, construir uma relação matemática simples que relaciona os dados e os resultados do modelo de simulação, isto é, um modelo do modelo de simulação, ou **metamodelo** [9]. Existem diversas razões, que em seguida são descritas, para que o interesse na área de metamodelos de simulação se tenha mantido ao longo dos últimos trinta anos.

Evidenciar. Os metamodelos podem expor, mais claramente que o modelo de simulação, a natureza fundamental das relações entre os dados e resultados do sistema. A sua forma simples pode evidenciar as características gerais do comportamento do modelo de simulação.

Verificar e validar. A essência fornecida por um modelo mais simples (metamodelo) pode ser utilizada para verificação e validação do modelo mais complexo que lhe deu origem (modelo de simulação).

Identificar. Os metamodelos podem ser utilizados para identificar os parâmetros do sistema que mais afectam o seu desempenho.

Experimentar. Podem ser muito úteis para responder a questões do tipo 'o que acontece se?' porque utilizam menos recursos computacionais que os modelos de simulação digital, que em geral se traduzem por custos elevados e grandes demoras na sua execução. Este tipo de estudo experimental é particularmente importante quando a resposta com interesse é uma quantidade aleatória e o objectivo é otimizar o desenho do sistema.

Compor. Inserção do metamodelo noutra modelo é também uma vantagem importante quando o modelo original é apenas uma componente de um modelo de um sistema mais complexo. Nesta situação, o modelo do sistema pode ser impraticavelmente lento e/ou grande se não se utilizarem metamodelos para algumas ou todas as componentes.

Inverter. Um metamodelo permite responder a questões 'inversas' que seriam de difícil resposta sem o seu auxílio. Responder a problemas inversos sem o auxílio de um metamodelo exige tempo de computação excessivo, porque é necessário fazer diversas tentativas e cada tentativa corresponde a uma execução do programa de simulação.

É crucial para todo o tipo de vantagem referida anteriormente, a estimação precisa dos parâmetros desconhecidos do metamodelo. Existem dois grandes desafios na área de melhorar a precisão da estimação de metamodelos: detecção do viés inicial e técnicas de redução de variância.

(i) **Detecção do viés inicial.** Um modelo de simulação digital estocástico necessita da especificação completa das condições iniciais, em cada execução do programa. Estas condições iniciais podem ser escolhidas aleatoriamente de acordo com a probabilidade conjunta da sua ocorrência. Contudo, isso é tipicamente impossível ou impraticável. Em geral, as condições iniciais representam estados extremamente raros. Como consequência, após começar a execução do programa, o processo encontra-se numa fase não estacionária. Se o processo for estável, ele vai evoluindo até atingir o estado estacionário (equilíbrio estatístico). É de extrema importância a recolha de resultados fora do estado transiente inicial, para que eles possam caracterizar o estado de equilíbrio (ver as páginas 138 e 146). Mais precisamente, é necessário deitar fora as observações correspondentes a esse estado transiente antes da análise estatística, para que se possa tirar conclusões fiáveis em relação ao sistema no seu estado de equilíbrio. O ignorar deste período inicial pode resultar na obtenção de enviesamento significativo nos resultados. Por outro lado, não ter em conta parte dos resultados faz com que a variância dos estimadores aumente. Têm sido desenvolvidos muitos trabalhos nesta área, tanto em relação à escolha de condições iniciais, como em relação à detecção do estado estacionário (ver, por exemplo, [38], [76], [120] [126] e [82]).

(ii) **Técnicas de redução de variância.** Os modelos de simulação digital são frequentemente utilizados para estimar metamodelos. Neste caso, quando é elaborada a experiência de simulação, o objectivo final é determinar os estimadores dos parâmetros desconhecidos do metamodelo. Estas medidas são variáveis aleatórias e, por conseguinte, possuem variâncias que necessitam de ser estimadas, com o objectivo de construir intervalos de confiança, em torno dos parâmetros de interesse, e fornecer medidas de erro associadas às estimativas. Quanto menor for a variância, maior será a precisão das estatísticas geradas com o auxílio da experimentação com o modelo de simulação: necessitamos, portanto, de menor número de execuções, para obter a mesma precisão, ou, obtemos maior precisão, com o mesmo número de execuções. Por conseguinte, as técnicas de redução de variância são mecanismos eficazes para melhorar a precisão das estatísticas produzidas na simulação digital (ver o Capítulo 5). Neste sentido, têm sido desenvolvidos diversos trabalhos sobre este assunto ao longo dos últimos anos e continua a ser uma área de grande actualidade (ver, por exemplo, [122], [16], [27] e [102]).

1.2 Métodos para análise de resultados

Desde há muitos anos que a simulação estocástica digital tem sido uma ferramenta frequentemente utilizada na avaliação do desempenho de diversos sistemas. Acontece porém que, muitas vezes, os resultados dos estudos de simulação têm pouca credibilidade, porque são apresentados sem ter em conta a sua natureza estocástica e a necessidade de uma análise estatística de resultados correcta. Por exemplo, em sistemas de filas de espera, os acontecimentos modelados são controlados por números aleatórios e, por conseguinte, os resultados produzidos não são mais que amostras aleatórias (ver, por exemplo, [118, secção 6.1] ou [73]). Kleijnen defende que, se se ignorar a natureza aleatória dos resultados, em vez de utilizar um modelo de simulação dispendioso, é preferível atirar uma moeda ao ar [56]. É absolutamente necessária a inferência estatística em qualquer situação em que o mesmo programa (correcto) produza diferentes

resultados em cada execução do programa.

Na realidade, um estudo de simulação não pode ser visto como um mero exercício de programação. Por conseguinte, após ter-se gasto muito esforço intelectual e tempo na construção do modelo de simulação, têm de ser utilizadas técnicas estatísticas apropriadas para escolher os dados e analisar os resultados das experiências de simulação. Além disso, tem de se ter em conta que as técnicas estatísticas clássicas, baseadas em observações independentes e identicamente distribuídas, não podem ser directamente aplicadas porque os resultados de, praticamente, todas as experiências de simulação não são estacionários (as distribuições das sucessivas observações variam ao longo do tempo) e são auto-correlacionados (as observações estão correlacionadas umas com as outras). Por exemplo, nas simulações de sistemas de filas de espera (ver a Secção 6.2), após começar o processo, ele está numa fase não estacionária (tempo de aquecimento). Se o processo for estável, move-se assintoticamente para o estado estacionário (equilíbrio estatístico). Note-se porém que, em geral, parâmetros diferentes tendem para o estado estacionário com taxas diferentes. É preciso ter cuidado na análise de resultados porque, se eles forem analisados nos períodos transientes iniciais, não caracterizam o estado estacionário. A ideia natural é deitar fora essas observações antes da análise estatística. O ignorar deste período inicial pode resultar em obtenção de enviesamento significativo nos resultados. Por outro lado, não ter em conta parte dos resultados faz com que a variância dos estimadores aumente.

Têm sido propostos diversos métodos para análise de resultados com o objectivo de ultrapassar os problemas teóricos que surgem da natureza correlacionada das observações, recolhidas durante uma simulação no estado estacionário (ver, por exemplo, [72], [95], [108] e [1]). Em geral, estes métodos impõem determinados requisitos na forma de análise e pré-processamento de resultados, que dependem da maneira pela qual eles tentam enfraquecer, ou até mesmo remover, as dependências estatísticas entre observações ou a forma como as tomam em consideração.

1. Método das réplicas independentes O problema da auto-correlação dos resultados originais é ultrapassado da seguinte forma: a simulação é repetida um determinado número de vezes, de cada vez usando uma sequência diferente de números aleatórios independentes; seguidamente é calculado o valor médio das observações recolhidas em cada execução do programa de simulação. Deste modo estas médias são independentes e identicamente distribuídas e, pelo teorema do limite central, são também normalmente distribuídas. O estimador da média da população obtém-se calculando a média destas médias. Welch [118] apresenta e discute diversos estimadores baseados neste método para análise de resultados.

Os restantes métodos, para análise de resultados, foram desenvolvidos para obter estimadores no estado estacionário, a partir de uma única execução do programa de simulação.

2. Método das médias de lotes Este método foi mencionado, pela primeira vez, por Blackman e Tuckey [13] e por Conway, Johnson e Maxwell [23] e consiste no seguinte: o programa de simulação é executado uma única vez, deixando passar muito tempo (simulado), e são recolhidas N observações originais. Em seguida, divide-se este grupo de observações em lotes que não se sobrepõem e calcula-se a média das observações dentro de cada lote. Estes lotes têm de ser suficientemente grandes para assegurar independência aproximada das estatísticas calculadas a partir de cada um deles. A ideia, que está por trás deste método, é que observações mais separadas no tempo são menos correlacionadas. Por conseguinte, com lotes suficientemente grandes, conseguem-se médias baseadas nestes lotes não correlacionadas (ou quase). Também aqui não devem ser tomadas em conta as observações correspondentes ao estado transiente inicial.

3. Método regenerativo Este método foi sugerido por Cox e Smith [24], e depois desenvolvido independentemente por Fishman [33, 34] e Crane e Iglehart [25, 26]. Primeiro identificam-se os sucessivos instantes, no tempo, em que o processo simulado recomeça (no sentido probabilístico), isto é, no qual as futuras transições de estado não

dependem do passado. Por exemplo, quando um cliente encontra a fila de espera vazia e o servidor desocupado, ele inicia o ciclo – servidor ocupado – para o servidor. A partir deste momento nenhum acontecimento passado influencia a futura evolução do sistema. Na teoria dos processos regenerativos, tais instantes no tempo são denominados *pontos regenerativos*. O comportamento do sistema entre pontos regenerativos consecutivos constitui um *ciclo regenerativo*. O facto de, em cada ponto regenerativo, o processo renascer faz com que os lotes de observações, recolhidas durante diferentes ciclos regenerativos (isto é, dentro de períodos de tempo limitados por pontos regenerativos consecutivos), sejam estatística e independentemente distribuídas. Note-se que, neste método, também existem lotes mas o seu comprimento é aleatório, ao contrário dos lotes no método das médias de lotes. Não esquecer que, o programa de simulação é executado até que um número adequado de ciclos regenerativos assegurem a validade das técnicas utilizadas neste método.

4. Método espectral Este método apareceu pela primeira vez em 1967 [30, 36]. O método pressupõe que as observações representam uma sucessão estacionária em covariância. A análise é feita no domínio das frequências, aplicando a transformada de Fourier a uma determinada função de auto-correlação, resultando na função de densidade espectral. Observe-se que, nas restantes técnicas para análise de resultados, lida-se directamente com o próprio processo estocástico, enquanto que aqui se trabalha com a densidade espectral do processo no domínio das frequências. Esta técnica evita os problemas que surgem da dependência entre resultados. A função de densidade espectral no ponto zero é, para uma dimensão de amostra suficientemente grande, proporcional à variância procurada. Note-se que, o espectro de frequências tem natureza aleatória uma vez que as observações recolhidas também a têm. Têm sido propostas diversas técnicas com o objectivo de obter bons estimadores para o espectro de frequências. Muitas delas seguem procedimentos clássicos de estimação espectral baseada no conceito de janelas espectrais (funções com pesos especiais introduzidas para diminuir o enviesamento final dos estimadores).

5. Método auto-regressivo Este método foi desenvolvido por Fishman [31, 32, 35] e assume que o processo, constituído pelos resultados de simulação, tem covariância estacionária e pode ser representado através de um modelo auto-regressivo. A maior dificuldade em utilizar este método é a validade do modelo auto-regressivo, o que limita a sua aplicabilidade.

6. Método de séries temporais padronizadas Este método foi proposto originalmente por Schruben [104, 103] e depende da convergência dos processos aleatórios padronizados, para um processo aleatório de Wiener, com incrementos independentes (processo de ponte Browniana). É uma aplicação dos processos aleatórios dependentes e o seu teorema do limite central funcional, que é uma generalização no conhecido teorema do limite central (escalar). De acordo com esta aproximação, analisa-se uma sequência de observações dividindo-a em lotes e transforma-se cada uma delas na sua forma padrão, exigida pelo teorema do limite central funcional. Em seguida, pode-se analisar diversas funções da sequência transformada para construir um intervalo de confiança para a média. O método exige que o processo analisado seja estacionário. Por conseguinte, tem de se deitar fora as observações iniciais, referentes ao período não estacionário transiente, antes da divisão por lotes (não sobrepostos).

Para finalizar, faz-se uma revisão breve dos diferentes métodos para análise de resultados salientando as suas vantagens e inconvenientes.

A análise espectral pode produzir, por vezes, resultados finais bastante precisos mas nesses casos a implementação pode não ser fácil devido à sua sofisticada natureza (ver, por exemplo, [72] e [81]). Além disso, não existe nenhuma regra geral para escolher determinados parâmetros necessários à aplicação deste método. Contudo, este método parece ser promissor mas tem de ser feito muito trabalho nesta área.

A principal desvantagem da representação auto-regressiva é a exigência da existência de uma representação auto-regressiva, do processo simulado, o que torna o método muito restritivo [20]. Além disso, Law e Kelton [72], após estudos comparativos de

diferentes métodos, com dimensão de amostra fixo, também descobriram que a aproximação auto-regressiva não oferece melhores resultados que outros métodos, para análise de resultados, computacionalmente mais simples. Além disso, se for preciso um número grande de observações, a determinação dos parâmetros necessários à aplicação do método pode ser computacionalmente dispendiosa.

A análise regenerativa necessita de amostras de dimensão elevada (não menos de 100 ciclos), para obter intervalos de confiança válidos. Outra desvantagem é que exige uma boa escolha de estados regenerativos. A selecção de estados regenerativos que ocorrem mais frequentemente não garante a melhor qualidade do estimador [15]. Além disso, para sistemas de grande dimensão, pode ser difícil ou até mesmo impossível identificar instantes regenerativos. Mesmo quando são encontrados, os estados regenerativos podem ocorrer com frequência extremamente baixa. Para tentar ultrapassar esta dificuldade, Andradóttir, Calvin e Glynn [4] apresentaram uma definição de ciclos regenerativos que envolve *splitting* e que permite identificar mais facilmente instantes regenerativos. Henderson and Glynn [47] realçam a essência da dificuldade em identificar instantes regenerativos em simulações com acontecimentos discretos. Outra desvantagem é que, este método só é válido sob determinadas condições, o que o torna ainda mais restritivo [41]. Consequentemente, a análise regenerativa não é uma técnica para utilização geral.

Em muitas situações, o método das médias de lotes comporta-se melhor do que o método das réplicas independentes e por essa razão é muito utilizado na prática [71]. Contudo, tem a desvantagem de ainda não ter sido descoberta uma técnica geral satisfatória para escolher os lotes [22] e o facto de não se conseguir eliminar totalmente a correlação entre observações [77]. Têm surgido diversos métodos que resultam de modificações no método das médias de lotes como, por exemplo, os métodos das médias de lotes sobrepostos (existem observações que pertencem a dois lotes distintos), amostragem não correlacionada (apenas uma única observação de cada lote é recolhida) e médias de lotes ponderadas (atribuem-se pesos às observações dentro de cada lote). Acontece porém que, tanto o método das médias de lotes, como os seus métodos derivados, não têm aplicação geral; por exemplo, o método das médias de lotes ponderadas [12] só produz

melhores resultados se o processo apresentar correlação forte e estiver disponível um número reduzido de observações. Além disso, pode-se pretender estimar outras variáveis que não são médias (Welch mencionou variâncias, quantis e probabilidades [118]) e a independência aproximada, como foi definida para as médias de lotes, não parece ser generalizável para incluir outros quantis definidos em médias separadas.

Sob determinadas condições, o método das séries temporais padronizadas consegue ser mais vantajoso que o método das médias de lotes não sobrepostos. Contudo, apesar dos estimadores terem formas numéricas simples, as técnicas estatísticas envolvidas são sofisticadas. Com o objectivo de melhorar a qualidade dos estimadores, obtidos através deste método, tem sido feito trabalho nesta área (ver, por exemplo, [114]). Infelizmente, não existe nenhuma regra simples para escolher o tamanho dos lotes e têm sido feitas poucas implementações deste método. Além disso, este método precisa, em geral, de lotes maiores que os métodos de médias de lotes (sobrepostos ou não).

O trabalho desenvolvido no âmbito desta dissertação situa-se no contexto do método das réplicas independentes. Apesar desta técnica ser mais sensível à não estacionariedade das observações recolhidas durante o período transiente inicial (cada nova réplica começa com um novo período inicial), comparando com os outros métodos baseados numa única execução do programa de simulação, ela pode ser utilizada em qualquer modelo de simulação, para estimar qualquer tipo de medida pretendido. Também convém salientar que, actualmente, existem bons métodos para controlar ou eliminar o enviesamento inicial, o que faz com que o método das réplicas independentes seja perfeitamente viável. Além disso, na maior parte das situações, pode-se obter melhor precisão do estimador pontual, medido pelo respectivo valor do erro quadrático médio (ver a página 43), se o programa de simulação for repetido diversas vezes, em vez de ser executado uma só vez (ver [52], [115] e [121]).

1.3 Motivação e objectivos

A maior parte dos investigadores de simulação prefere os metamodelos de regressão (ver, por exemplo, [61], [29], [94] e [21]). A utilização deste tipo de modelos auxiliares permite o uso de técnicas estatísticas conhecidas da análise de regressão. Dentro da classe dos metamodelos de regressão, os metamodelos polinomiais têm recebido atenção considerável devido à sua simplicidade (ver, por exemplo, [67]). Contudo, os polinómios não têm a capacidade de se ajustarem à forma de funções regulares arbitrárias, com base em valores arbitrários que podem ser perturbados por componentes estocásticas com distribuição geral. Também, na maior parte das situações, a previsão fora da região experimental não é satisfatória. Além disso, é habitual a não linearidade em problemas da vida real e, por conseguinte, a aproximação utilizando polinómios torna-se irrealista. Como resultado, é frequente neste tipo de problemas não obter bons ajustes quando se utilizam metamodelos polinomiais (ver, por exemplo, problemas envolvendo sistemas de filas de espera [37]). Outra desvantagem, nos estudos com metamodelos em simulação, é que os metamodelos são utilizados de uma maneira *ad hoc*.

Sendo assim, existe uma necessidade urgente de desenvolver uma metodologia para metamodelação que permita obter ajustes globais melhores e mais realistas. Neste trabalho, oferece-se uma maneira sistemática de construir metamodelos não lineares de regressão e propõem-se ferramentas estatísticas para levar a cabo cada passo da metodologia (ver [100] e [101]). Os métodos estatísticos utilizados são uma aplicação e adaptação, à simulação de acontecimentos discretos, de técnicas estatísticas conhecidas na área da regressão. Sendo assim, este estudo situa-se numa área, que continua a ser de grande actualidade, denominada técnicas estatísticas em simulação. Foram escolhidos os metamodelos não lineares de regressão porque não tem as desvantagens dos metamodelos polinomiais e permitem:

(i) captar a essência do modelo que lhe deu origem através da sua forma simples;

(ii) uma vez que se podem utilizar as ferramentas estatísticas de análise de regressão não linear, é possível caracterizar a sua precisão através da construção de intervalos de confiança, etc;

(iii) prever valores longe da região experimental e esses valores serem fiáveis;

(iv) na fase da estimação de metamodelos, utilizar algum software já existente em pacotes de programas de estatística e de análise numérica.

A metodologia desenvolvida utiliza a simulação de acontecimentos discretos para estimar e validar um metamodelo de regressão não linear geral, em que a resposta do sistema depende de determinadas variáveis explicativas relevantes. Esta metodologia salienta a importância da validação de metamodelos, porque na realidade um metamodelo só pode ser utilizado para previsão e análise de sensibilidade se ele aproxima adequadamente o comportamento dos resultados numéricos obtidos pelo programa de simulação. Tanto a estimação (estimação de parâmetros, construção de regiões de confiança e testes de hipóteses) como a validação são elaboradas utilizando e adaptando técnicas estatísticas de regressão. Com o objectivo de reduzir o erro (variância) dos estimadores dos parâmetros do metamodelo, recorre-se ao uso do método das variáveis de controlo. Comparativamente com outras técnicas de redução de variância, o método das variáveis de controlo não exige que o processo estocástico seja alterado de forma alguma. Além disso, a sua fácil realização e grande generalidade faz com que este método seja atractivo para obter redução de variância.

1.4 Organização da dissertação

No capítulo 'Estudos importantes sobre metamodelos', enquadra-se o trabalho desenvolvido com aspectos relacionados de simulação, incluindo os trabalhos mais recentes sobre estimação e validação de metamodelos e técnicas de redução de variância. Esta pesquisa também inclui os diversos tipos de metamodelos que surgem na literatura corrente e a razão pela qual se escolheram metamodelos não lineares de regressão.

No capítulo 'Metodologia para metamodelação', descreve-se uma metodologia para construção de metamodelos de regressão não linear para simulação de acontecimentos discretos, apresentando procedimentos adequados para levar a cabo cada passo dessa metodologia. Um dos passos importantes da metodologia é a estimação de metamodelos. Por esse facto, reserva-se uma secção à apresentação de resultados assintóticos sobre os estimadores dos parâmetros desconhecidos do metamodelo, que incluem a expressão para esses estimadores, regiões de confiança e testes de hipóteses.

A validação do metamodelo estimado é um passo muito importante da metodologia sendo tratada no capítulo 'Validação de metamodelos'. A validação de um metamodelo consiste em três passos (verificação da adequação do metamodelo, validação do metamodelo com respeito ao modelo de simulação e validação do metamodelo com respeito ao problema real) e são fornecidos procedimentos numéricos e estatísticos para efectuar cada um deles.

No capítulo 'Variáveis de controlo', é descrito um procedimento para melhorar a precisão da estimação de metamodelos, recorrendo à utilização da técnica das variáveis de controlo. Aqui são apresentadas expressões para os estimadores controlados, regiões de confiança e testes de hipóteses. Também é tratado o problema de avaliação do desempenho da técnica das variáveis de controlo através de determinadas medidas conhecidas no caso de metamodelos lineares de regressão.

No capítulo 'Aplicação', aplica-se a metodologia a dois problemas distintos: um exemplo é um problema actual que consiste num centro para inspecção e manutenção de automóveis (CIMA) e o outro é a situação clássica do sistema $M/M/1$. No primeiro exemplo, consideram-se diversos metamodelos candidatos, incluindo lineares e não lineares e a metodologia permitiu seleccionar um metamodelo não linear que aproxima bem o sistema em estudo. No segundo exemplo, a metodologia permitiu eleger dois metamodelos não lineares: um sem variáveis de controlo e outro com variáveis de controlo. Em seguida, neste segundo exemplo, estudou-se o efeito produzido pela utilização da técnica das variáveis de controlo.

Finalmente, no capítulo 'Conclusões e recomendações', apresentam-se as conclusões e referem-se os resultados mais importantes da investigação. Também se dão perspectivas de trabalho futuro, com base nas conclusões e resultados obtidos.

No 'Apêndice A', inclui-se um catálogo de curvas muito importante para a selecção do tipo de metamodelo (passo 3 da metodologia). Na bibliografia, podem ser encontradas referências a livros e publicações consultados na execução deste trabalho e que representam um complemento para a compreensão dos problemas e soluções que surgem na construção de metamodelos de regressão.

Capítulo 2

Estudos importantes sobre metamodelos

Os metamodelos de simulação têm sido utilizados ao longo dos últimos trinta anos. Os que surgem mais frequentemente na literatura são os metamodelos polinomiais devido à sua simplicidade. Contudo, existem outros tipos de metamodelos que têm sido estudados. Na Secção 2.1, apresentam-se os diversos tipos de metamodelos que surgem mais frequentemente na literatura, assim como os trabalhos desenvolvidos mais recentemente sobre eles (ver, por exemplo, [9] ou [10]). A Secção 2.2 trata um assunto particular da área de metamodelos que tem grande relevo no trabalho que foi desenvolvido: validação de metamodelos de regressão.

Um objectivo frequente das experiências de simulação desenhadas estatisticamente é estimar um metamodelo, por exemplo um metamodelo de regressão. Qualquer que seja o propósito com que o metamodelo é desenvolvido, é de extrema importância a estimação precisa dos parâmetros desconhecidos do metamodelo. Em particular, podem ser necessárias amostras de dimensão excessivamente grande para obter precisão aceitável nos estimadores dos parâmetros do metamodelo. A melhoria da eficiência estatística dos estimadores referidos é obtida tipicamente (mas não sempre) através da redução de variância. Na Secção 2.3, apresentam-se as diversas técnicas de redução de variância salientando as suas vantagens e inconvenientes e trabalhos relevantes na área. Na Secção 2.4, explica-se em pormenor a técnica eleita para utilizar neste trabalho — a técnica das

variáveis de controlo — e apresentam-se trabalhos relevantes nesta área, directamente relacionados com o trabalho desenvolvido.

2.1 Metamodelos de simulação

O modelo (ou programa) de simulação representa-se através da função matemática, ϕ :

$$\mathbf{Y} = \phi(\mathbf{Z}, \mathbf{r}), \quad (2.1)$$

onde \mathbf{Y} é a resposta do sistema, $\mathbf{Z} = (Z_1, \dots, Z_k)$ são os parâmetros e variáveis de entrada e \mathbf{r} é ou um vector de números aleatórios ou um valor, ou semente, inicial escolhido do gerador de números pseudo-aleatórios; note-se que se introduz aleatoriedade em simulação estocástica através de números pseudo-aleatórios. O vector \mathbf{Z} , na simulação de um processo de fabrico, pode incluir as seguintes componentes: o número de máquinas, os tempos de processamento das máquinas e a distribuição dos tempos entre falhas das máquinas. O vector resposta \mathbf{Y} pode incluir o média do trabalho em processamento ou a média diária do custo de operação. Contudo, na prática corrente, utilizam-se os metamodelos com uma única resposta: é desenvolvido um metamodelo para cada resposta individual. Por conseguinte, para cada resposta, os metamodelos (funções aproximadoras do programa de simulação) que vão ser considerados são da forma:

$$Y = f(\mathbf{X}, \boldsymbol{\theta}) + \epsilon, \quad (2.2)$$

onde as variáveis explicativas $\mathbf{X} = (X_1, \dots, X_d)$ pertencem a um subconjunto de \mathbb{R}^d , $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T \in \Theta \subseteq \mathbb{R}^m$ é um vector de parâmetros desconhecidos, ϵ representa o erro e f é uma função desconhecida. O erro ϵ inclui, tanto efeitos devido à inadequação de f como representante de ϕ , como efeitos intrínsecos, sempre presentes em qualquer modelo de simulação estocástica (dependem de \mathbf{r} em (2.1)). Nalguns casos X_i , em (2.2), é idêntico ao parâmetro ou variável de simulação Z_j , em (2.1). Noutras situações, X_i pode ser uma transformação de um ou mais Z_j 's; por exemplo, no caso específico dos

sistemas das filas de espera, $X_1 = Z_1/Z_2$ pode ser uma melhor variável explicativa (se considerarmos $\rho = \lambda/\mu$). Por conseguinte, os parâmetros e variáveis de entrada \mathbf{Z} do modelo de simulação (2.1) determinam as variáveis independentes \mathbf{X} do metamodelo de simulação (2.2). Os coeficientes θ em (2.2) são designados por parâmetros do metamodelo e têm de ser estimados.

As áreas mais importantes em modelação com metamodelos de simulação incluem:

- (i) a escolha da forma funcional de f ;
- (ii) o desenho de experiências, isto é, a selecção de um conjunto de pontos experimentais \mathbf{X} para cada observação Y , com vista à estimação dos parâmetros desconhecidos θ , a atribuição das sementes de números pseudo-aleatórios, o comprimento das execuções do programa de simulação, etc; e
- (iii) a avaliação da precisão do metamodelo ajustado que consiste na construção de intervalos de confiança, testes de hipóteses, falta de ajustamento e outros diagnósticos.

A forma funcional é geralmente descrita por uma combinação linear de funções base pertencentes a uma família paramétrica. Por conseguinte, é preciso escolher a família (por exemplo, polinómios, funções seno, polinómios por intervalos, ondas, etc) e depois escolher a forma para eleger o membro da família escolhida. O desenho experimental e a avaliação da precisão estão relacionados entre si, porque a selecção do desenho experimental irá ser determinado, em parte, pela sua influência na avaliação da precisão.

Existem diversos tipos de metamodelos na literatura corrente:

1. modelos de regressão;
2. modelos de Taguchi;
3. modelo linear generalizado;
4. *splines*;
5. funções base radiais;
6. núcleos regulares;

7. modelos de correlação espacial;
8. aproximações no domínio da frequência (Fourier e pequenas ondas);
9. redes neurais artificiais.

Uma vez que existem diversos tipos de metamodelos disponíveis é preciso saber como é que se pode fazer uma escolha entre eles. Existem muitos critérios que podem ser considerados. Alguns deles apresentam-se em seguida:

1. a capacidade de captar a essência do problema a partir da forma do metamodelo;
2. a capacidade de adaptação à forma de diversas funções regulares arbitrárias com base em valores observados, que podem ser perturbados por componentes estocásticas com distribuição geral;
3. a capacidade de caracterizar a precisão através de intervalos de confiança, testes de hipóteses, etc;
4. a robustez da previsão longe dos valores observados;
5. a facilidade de cálculo da função aproximadora f ;
6. a estabilidade numérica dos cálculos e, conseqüentemente, robustez das previsões para pequenas mudanças nos parâmetros que definem a função f ;
7. A existência de software para calcular o metamodelo, caracterizar o seu ajuste e usá-lo para previsão.

Em seguida, são apresentados em sequência os metamodelos que foram enumerados anteriormente. A apresentação consiste em focar a natureza das funções aproximadoras e as suas vantagens e inconvenientes com base nos critérios de 1 a 7.

1. Modelos de regressão Os metamodelos polinomiais de regressão têm sido uma escolha preferencial entre os investigadores de simulação (ver, por exemplo, [74], [55], [58], [51], [62], [61], [29] e [21]). Neste caso, a função aproximadora tem a forma

$$f(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^m \theta_i p_i(\mathbf{X}), \quad (2.3)$$

onde $p_1(\mathbf{X}), \dots, p_m(\mathbf{X})$ é um conjunto de m funções conhecidas que dependem das variáveis de decisão incluindo, por exemplo, os termos de primeira ordem X_1, X_2, \dots alguns termos de interacção de segunda ordem X_1X_2, X_1X_3, \dots e algumas interacções de ordens superiores $X_1X_2X_3, \dots$. O vector de coeficientes $\boldsymbol{\theta}$ pode ser estimado, a partir dos resultados de uma experiência de simulação através do método dos mínimos quadrados. O estimador dos mínimos quadrados de $\boldsymbol{\theta}$ é o estimador que minimiza a soma dos quadrados dos erros entre as respostas observadas e os valores previstos ou ajustados. Este estimador é um vector aleatório, porque depende das observações aleatórias recolhidas na experiência de simulação. Suponha-se, no contexto das réplicas independentes, uma experiência de simulação com n pontos experimentais e r réplicas em cada ponto experimental; então, a relação da resposta Y_{ij} para a j -ésima réplica do i -ésimo ponto experimental tem a forma

$$Y_{ij} = f(\mathbf{X}_i, \boldsymbol{\theta}) + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, r,$$

onde ϵ_{ij} é o erro correspondente à j -ésima réplica do i -ésimo ponto experimental. Por hipótese, considera-se que

$$\epsilon_{ij} \sim N(0, \sigma_i^2), \quad (2.4)$$

para todo o $i = 1, \dots, n$ e $j = 1, \dots, r$. A literatura tradicional considera variâncias constantes $\sigma_i^2 = \sigma^2$. Contudo, em muitos casos, a hipótese de variâncias constantes é irrealista (por exemplo, a variância pode crescer ao longo dos pontos experimentais). Por conseguinte, estas duas situações são estudadas separadamente e associadas a elas estão estimadores diferentes:

(i) *Variância dos erros constante.*

Quando a variância dos erros não varia ao longo dos pontos experimentais, utiliza-se o *estimador dos mínimos quadrados usuais*, o qual tem a forma

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{Y}},$$

onde

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1m} \\ 1 & X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix}$$

para polinômios de primeiro grau e inclui produtos, termo a termo, das colunas desta matriz para polinômios de ordem superior. Além disso, $\bar{\mathbf{Y}} = (\bar{Y}_1, \dots, \bar{Y}_n)^T$ com $\bar{Y}_i = \sum_{j=1}^r Y_{ij}/r$.

Pode ser demonstrado que $\hat{\boldsymbol{\theta}}$ é o estimador que possui menor variância dentro da classe dos estimadores lineares não enviesados. As variâncias das componentes do vector $\hat{\boldsymbol{\theta}}$ são os elementos da diagonal principal da matriz

$$\mathcal{D}[\hat{\boldsymbol{\theta}}] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (2.5)$$

Para estimar σ^2 , utiliza-se o seguinte estimador (independente do metamodelo):

$$\hat{\sigma}^2 = \frac{1}{N - n} \sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2,$$

onde $N = nr$. Substituindo este estimador em (2.5), obtém-se a matriz de variância-covariância estimada $\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}] = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}$.

(ii) *Variância dos erros heterogénea.*

Quando a variância dos erros não é constante, ao longo de todos os pontos experimentais, o melhor estimador linear não enviesado é obtido através do método dos mínimos quadrados ponderados. O *estimador dos mínimos quadrados ponderados* é dado por

$$\hat{\boldsymbol{\theta}} = \left\{ \mathbf{X}^T (\mathcal{D}[\bar{\mathbf{Y}}])^{-1} \mathbf{X} \right\}^{-1} \mathbf{X}^T (\mathcal{D}[\bar{\mathbf{Y}}])^{-1} \bar{\mathbf{Y}},$$

onde $\bar{\mathbf{Y}} = (\bar{Y}_1, \dots, \bar{Y}_n)^T$ e se assume que $\mathcal{D}[\bar{\mathbf{Y}}]$ é não singular. Uma vez que as observações são independentes com variâncias heterogêneas tem-se $\mathcal{D}[\bar{\mathbf{Y}}] = \text{diag}[\sigma_1^2, \dots, \sigma_n^2]/r$, onde $\sigma_i^2 = \text{Var}[Y_{ij}]$, $i = 1, \dots, n$, $j = 1, \dots, r$.

Infelizmente, o estimador dos mínimos quadrados ponderados não resolve grande parte dos problemas porque, na prática, as variâncias σ_i^2 são desconhecidas. Neste caso, utiliza-se o *estimador dos mínimos quadrados ponderados estimados*

$$\tilde{\boldsymbol{\theta}} = (\mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \left(\frac{\hat{\boldsymbol{\Sigma}}}{r} \right)^{-1} \bar{\mathbf{Y}}, \quad (2.6)$$

com

$$\hat{\boldsymbol{\Sigma}} = \text{diag}[\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2], \quad (2.7)$$

onde $\hat{\sigma}_i^2$ é o estimador clássico da variância σ_i^2 [64, página 151]

$$\hat{\sigma}_i^2 = \frac{1}{r-1} \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2, \quad i = 1, \dots, n. \quad (2.8)$$

O estimador (2.6) é não linear uma vez que a transformação de $\bar{\mathbf{Y}}$ envolve as variáveis aleatórias $\hat{\boldsymbol{\Sigma}}$. Sob determinadas condições de regularidade não restritivas, pode ser provado que, este estimador tem matriz de variância-covariância assintótica dada por

$$\mathcal{D}[\tilde{\boldsymbol{\theta}}] = \left(\mathbf{X}^T \left(\frac{\boldsymbol{\Sigma}}{r} \right)^{-1} \mathbf{X} \right)^{-1}$$

para N suficientemente grande. Kleijnen utilizou o método de Monte Carlo para mostrar que, se os erros satisfizerem a condição (2.4), então o estimador dos mínimos quadrados ponderados estimados é não enviesado [61]. Para amostras de pequena dimensão utiliza-se por vezes o *estimador dos mínimos quadrados corrigidos*. Este estimador é não enviesado e coincide com o estimador dos mínimos quadrados usuais mas a matriz de variância-covariância estimada tem agora a forma

$$\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \left(\frac{\hat{\boldsymbol{\Sigma}}}{r} \right)^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}.$$

2. Modelos de Taguchi Os metamodelos de Taguchi gozam da propriedade de permitirem que a variância dos erros ϵ dependa do vector \mathbf{X} . A aproximação de Taguchi proporciona duas modificações do modelo (2.3). A primeira consiste em considerar os erros independentes mas a variância é modelada explicitamente a depender de \mathbf{X} , isto é, $\text{Var}[\epsilon_i] = \sigma^2(\mathbf{X})$, $i = 1, \dots, n$. Esta situação pode ser tratada com modelos polinomiais, embora não seja de uma maneira inteiramente satisfatória, através de: (i) transformação de dados; (ii) utilizando mínimos quadrados ponderados; ou (iii) variando inversamente o comprimento da experiência de simulação, em cada ponto experimental, em relação ao valor da variância $\text{Var}[\epsilon_i]$. A segunda modificação consiste em combinar o metamodelo que descreve a resposta Y , sendo este da forma (2.3), com o modelo que caracteriza σ^2 , através de

$$10 \log (E[Y^2]/\sigma^2(\mathbf{X})) = \sum_{i=1}^m \gamma_i p_i(\mathbf{X}). \quad (2.9)$$

Este modelo, resultante da referida combinação, envolve apenas termos lineares em \mathbf{X} . O objectivo de Taguchi é encontrar algumas componentes em \mathbf{X} que tem valores pequenos de γ_i , em (2.9), mas grandes valores de θ_i em (2.3).

3. Modelo linear generalizado O modelo (2.3) é vulgarmente denominado de modelo linear geral. O modelo linear generalizado é uma generalização do modelo linear geral em duas direcções. A primeira é que a distribuição dos erros pode pertencer a uma família exponencial que não seja forçosamente a distribuição normal. Por exemplo, distribuição de Poisson, binomial, gamma e distribuições gaussianas inversas. A segunda generalização tem a ver com a forma do modelo que substitui (2.3). Para o modelo clássico em (2.3), tem-se $E[Y] = f(\mathbf{X}, \boldsymbol{\theta})$. Neste caso, esta relação é mais flexível e é feita através da função diferenciável ligação, ou a sua inversa, g : $E[Y] = g(f(\mathbf{X}, \boldsymbol{\theta}))$.

4. Splines As funções *spline* são baseadas em funções base que são polinómios por intervalos. A construção deste tipo de funções é motivada pelo facto de algumas dificuldades dos modelos polinomiais serem ultrapassadas se: (i) os polinómios forem

aplicados em pequenas regiões; e (ii) apenas polinómios de graus baixos forem utilizados na aproximação. As funções *spline* são caracterizadas por exigirem continuidade da função aproximadora até uma determinada ordem que depende do grau do *spline*. O metamodelo é, neste caso, o seguinte

$$f(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^m \theta_i B_i(\mathbf{X}),$$

onde, mais frequentemente, $B_i(\mathbf{X})$ são funções base polinomiais de graus dois e três. No caso da simulação estocástica, os parâmetros desconhecidos são obtidos com base num compromisso explícito entre a precisão da aproximação nos pontos experimentais e a regularidade do metamodelo resultante. As funções *spline* surgem através da resolução de um problema de optimização, onde a importância relativa da precisão *versus* regularidade é controlada através de um determinado parâmetro. Consoante o grau da função *spline* e a escolha do parâmetro de medida da precisão *versus* regularidade, assim surgem três diferentes métodos: *splines* regulares, interpolação com *splines* e *splines* de regressão. As funções *spline* univariadas são a base das funções *spline* multivariadas, ou seja, funções para aproximar problemas multivariados.

5. Funções base radiais As funções base radiais são uma alternativa para modelar problemas multivariados. Um exemplo são as funções base multiquadráticas

$$f(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^n a_i \|\mathbf{X} - \mathbf{X}_i\|,$$

onde $\|\cdot\|$ representa a norma euclidiana. Os coeficientes a_i são calculados através da interpolação da função f com a verdadeira função nos pontos do desenho experimental.

6. Núcleos regulares Todos os métodos descritos anteriormente produzem valores previstos que são funções lineares dos valores observados com coeficientes determinados pelas funções base e seus coeficientes. O metamodelo núcleo regular utiliza esta representação explicitamente, sem desenvolver uma representação explícita para f em termos das funções base. Um valor de f é calculado directamente como uma soma ponderada dos valores observados Y_i , onde os pesos são determinados pela função núcleo.

Existem muitas formas desta soma ponderada ou função núcleo e existem muitas maneiras para calcular a função f . Também existe a extensão deste tipo de funções ao caso multivariado.

7. Modelos de correlação espacial Os modelos de correlação espacial consistem numa aproximação de modelação paramétrica de regressão que partilha de algumas particularidades com as funções *spline* regulares e núcleos regulares. A regularidade esperada da função é captada numa função de correlação espacial. A hipótese do modelo é

$$Y = f(\mathbf{X}, \boldsymbol{\theta}) + Z(\mathbf{X}),$$

onde Z é um processo estocástico Gaussiano com função de correlação espacial que depende de determinados parâmetros. Este tipo de modelos tem sido usado em simulação determinística mas, existe a sugestão de adicionar um termo estocástico para poder ser utilizado em simulação estocástica.

8. Aproximações no domínio da frequência Pensando nas variações de f ao longo do seu domínio, em termos da correlação espacial, surge a ideia das funções base de Fourier para representar uma aproximação do programa de simulação. Sempre que seja possível essa aproximação, a decomposição de Fourier é baseada em funções base com suporte global. Por conseguinte, a precisão do ajuste depende muito do cancelamento das oscilações das funções base.

9. Redes neuronais artificiais As redes neuronais artificiais são um ramo da inteligência artificial. Elas representam uma ligação de elementos simples de processamento capazes de processar informação em resposta a dados exteriores (ver, por exemplo, [50] e [8]). O elemento de processamento é referido como sendo um neurónio. Cada neurónio é ligado a, pelo menos, outro neurónio numa rede como se fosse um tecido. Uma rede é constituída por células que estão organizadas em camadas. Existem três tipos de camadas distintas numa rede neuronal: a camada de dados, a ou as camadas ocultas e a camada de resultados. A ligação das células ao longo das camadas representa

a transmissão de informação entre células. As redes neuronais são baseadas nalguma forma de modelo de aprendizagem e são treinadas dando repetidos exemplos de dados e respectivos resultados.

A aprendizagem pode ser de três tipos: supervisionada, não supervisionada e reforçada. Na aprendizagem supervisionada, cada resposta é guiada por parâmetros específicos. O computador é instruído para comparar os dados com as respectivas respostas ideais e são registadas todas as discrepâncias entre os dados e as respostas ideais. Em seguida, a rede usa esta informação para tentar adivinhar até que ponto as respostas geradas por novos dados são semelhantes ou diferentes das respostas ideais. Consoante o tipo do conjunto, que contem os dados e os resultados, assim a rede supervisionada pode ser de regressão, de classificação ou de regressão multi-modal.

Na aprendizagem não supervisionada, são calculados os dados e os respectivos resultados e, depois, guardados como padrões. A rede neuronal calcula um conjunto de padrões e identifica semelhanças ou diferenças entre eles. Em seguida, são atribuídos valores a esses padrões para que a rede possa decidir acerca de novos dados e resultados respectivos. Neste caso, as comparações são feitas em relação a novos resultados, em vez de a resultados ideais. Frequentemente, as redes não supervisionadas são utilizadas para descobrir padrões onde uma dada resposta não é conhecida antecipadamente.

A aprendizagem reforçada é um compromisso entre a aprendizagem supervisionada e a aprendizagem não supervisionada. Neste caso, à medida que o sistema está a aprender não lhe é fornecida informação contínua sobre se ele está ou não a aprender bem ou mal (como no caso das redes não supervisionadas). A aproximação supervisionada é apenas feita nos casos em que o sistema falha.

2.1.1 Análise comparativa

O objectivo desta secção é fazer uma revisão sobre as vantagens e inconvenientes dos metamodelos descritos anteriormente.

Nos metamodelos polinomiais de primeira e segunda ordens os coeficientes desconhecidos têm uma interpretação padrão. Um valor numérico grande do coeficiente do termo linear indica que a correspondente componente de \mathbf{X} tem uma influência significativa na resposta Y que está a ser modelada. Um valor numérico grande do coeficiente do termo quadrático indica uma resposta não linear. Um valor grande para o coeficiente do termo do produto cruzado $X_i X_j$ é interpretado como sendo a mudança na influência (linear) de X_i como função do valor de X_j (e de forma semelhante para X_j em relação a X_i). Os modelos polinomiais têm êxito no que diz respeito aos critérios 3, 5, 6 e 7. O cálculo dos estimadores dos parâmetros é simples e numericamente estável para qualquer desenho experimental razoável. A construção de intervalos de confiança, testes de falta de ajustamento e outros diagnósticos é feita utilizando técnicas estatísticas clássicas conhecidas [64]. Em relação ao critério 4, para polinómios de ordem baixa, a precisão de um valor previsto não se degrada rapidamente à medida que se afasta de uma observação experimental arbitrária. Contudo, o mesmo não se verifica para polinómios de ordem elevada. Enquanto que a interpretação destes metamodelos é extremamente simples e directa, o conjunto de formas que eles podem aproximar é limitado. Para optimização de superfícies de resposta isto em geral não é problema, porque os modelos são repetidamente construídos para aproximação em pequenas regiões experimentais e depois postos de parte à medida que avança a procura do óptimo. Contudo, quando existem múltiplos objectivos, as aproximações polinomiais podem ser uma má escolha. O modelo clássico polinomial de regressão exige hipóteses restritivas sobre o erro ϵ em (2.2), ou seja, requer que as respostas provenientes do modelo de simulação sejam independentes e identicamente distribuídas com distribuição normal.

Os modelos de Taguchi e linear generalizado são menos exigentes em relação às hipóteses sobre o erro ϵ em (2.2) e partilham de muitas propriedades com os modelos polinomiais. Os estimadores dos parâmetros desconhecidos, nos modelos de Taguchi, calculam-se da mesma forma que no caso dos modelos polinomiais e partilham do mesmo bom comportamento numérico. A forma polinomial de ordem baixa para os modelos (2.3) e (2.9) torna as interpretações bastante parecidas. A estrutura de (2.9) faz com que os procedimentos estatísticos simples para averiguar a precisão dos mo-

delos polinomiais não possam ser utilizados em estudos com os modelos de Taguchi. Além disso, as limitações inerentes aos polinómios também se verificam nos modelos de Taguchi e podem ser ajustados e estudados eficazmente utilizando as capacidades de modelação linear, presentes nos pacotes de programas estatísticos.

O modelo linear generalizado fornece um enquadramento formal, para metamodelos de simulação, quando o teorema do limite central não pode fornecer uma aproximação adequada para os resultados do modelo de simulação, considerando comprimentos de execução do programa de simulação que sejam práticos. Este modelo tem sugerido uma alternativa para analisar experiências de Taguchi. Ele proporciona uma fundamentação estatística forte e uma forma mais flexível. Contudo, esta flexibilidade não é conseguida sem uma desvantagem. As ferramentas disponíveis para avaliar a precisão do ajuste (critério 3) são baseadas na teoria assintótica dos estimadores de máxima verosimilhança. Estas são mais fracas que as correspondentes para os modelos lineares clássicos da forma (2.3). Além disso, algumas técnicas estatísticas para por exemplo desenho de experiências, testes de hipóteses e análise de variância não estão bem desenvolvidas. Em relação ao critério 7, existe software disponível para fazer o ajuste.

Os *splines* univariados são fáceis de descrever e compreender mas não são computacionalmente robustos. Surgem problemas numéricos quando existem muitos intervalos e, conseqüentemente, muitos parâmetros desconhecidos para estimar. Também são muito sensíveis, se for alterado um valor observado. Os *splines* produto de tensores exigem muitas execuções do programa de computação numa rede rectangular completa e têm dificuldades numéricas em calcular os coeficientes do metamodelo quando existem muitos dados. Em relação aos *splines* de interacção e aos *splines* adaptativos de regressão, eles são computacionalmente robustos e têm a capacidade de fornecer aproximações precisas ao longo de todo o domínio dos dados. Também existe algum software disponível. Um grande inconveniente é a falta de capacidade para caracterizar a precisão do ajuste através de intervalos de confiança, etc.

As funções base radiais são uma aproximação inviável porque o número de condição deteriora-se rapidamente, à medida que a dimensão e o número de valores observa-

dos aumenta. Além disso, uma vez que este método é um método de interpolação, a sua aplicação directa a experiências de simulação é limitada. Os metamodelos de núcleo regular são funções simples das observações, logo é possível construir intervalos de confiança para a verdadeira função. Contudo, este método é fraco em relação aos critérios considerados para escolher o tipo de metamodelo. Os modelos de correlação espacial têm sido utilizados para modelar simulações determinísticas, embora exista uma extensão para simulação estocástica. Não são particularmente bons em relação aos critérios 1 a 7. As aproximações no domínio da frequência têm a desvantagem associada a funções base globais. Por conseguinte, falham no que diz respeito aos critérios 4 e 6. Também, o software e aplicações são limitados até ao presente momento. Além disso, também não são bons em relação aos outros critérios.

Geralmente as redes neuronais são providas de precisão suficientemente boa, comparativamente com a precisão dos exemplos de aprendizagem, para tomar decisões sobre o modelo. Além disso, são uma alternativa a técnicas convencionais que, em geral, estão limitadas por hipóteses restritivas de normalidade, linearidade, independência de variáveis, etc. Contudo, não é boa a capacidade de caracterizar a precisão do ajuste através de intervalos de confiança, etc. Além disso, o tamanho da rede usada para modelar a resposta da simulação pode ser muito grande e pode acontecer ter de se usar muitos exemplos de aprendizagem, o que torna o estudo bastante lento.

Um tipo de metamodelos que goza das vantagens dos metamodelos polinomiais e não tem as desvantagens destes é a classe dos metamodelos não lineares de regressão. Como são modelos não lineares têm a capacidade de captar a essência do problema a partir da sua forma, com um reduzido número de parâmetros desconhecidos, e são robustos nas previsões fora da região experimental. Têm a grande vantagem de, ao contrário dos modelos polinomiais, se adaptarem a funções regulares arbitrárias com base em valores observados (que podem ser perturbados por componentes estocásticas com distribuição geral). A capacidade de caracterizar a precisão, através da construção de intervalos de confiança, testes de hipóteses e outros diagnósticos, é boa porque os metamodelos não lineares de regressão utilizam ferramentas conhecidas e muito estudadas da teoria estatística da regressão não linear. O cálculo da função f depende do problema específico

mas, em geral, não é preciso uma função muito complicada para explicar comportamentos de resultados que, à partida, podem parecer difíceis de compreender. Os métodos numéricos que utiliza estão suficientemente desenvolvidos e existem boas realizações desses métodos (métodos estáveis que não precisam de introdução de valor inicial para a sua execução, etc.) em pacotes de programas disponíveis, tanto para os especialistas de simulação, como para outro tipo de analista. Este tipo de metamodelos foi o escolhido porque se adapta particularmente bem à simulação de acontecimentos discretos e satisfaz plenamente os critérios 1 a 7.

Existe um tipo de redes neuronais que é treinada utilizando regressão não linear. Estas redes são conhecidas por redes neuronais supervisionadas de regressão. Por exemplo, uma rede com duas camadas e uma única variável de resposta pode ter a forma

$$\begin{aligned} \text{Camada escondida:} \quad h_j &= f^{(1)} \left(\sum_{k=1}^d w_{jk}^{(1)} X_k + \eta_j^{(1)} \right) \\ \text{Camada de resultados:} \quad Y &= f^{(2)} \left(\sum_{j=1}^u w_j^{(2)} h_j + \eta^{(2)} \right) \end{aligned}$$

onde d é o número de variáveis de entrada, e u é o número de unidades escondidas e, por exemplo, $f^{(1)}(a) = \tanh(a)$ e $f^{(2)}(a) = a$. Os "pesos" w e o enviesamento η juntos dão origem ao vector de parâmetros desconhecidos do metamodelo, θ . Esta rede é treinada minimizando a soma dos quadrados dos erros

$$\sum_{i=1}^n [Y_i - f(\mathbf{X}_i, \theta)]^2$$

onde

$$f(\mathbf{X}_i, \theta) = \sum_{j=1}^u w_j^{(2)} \tanh \left(\sum_{k=1}^d w_{jk}^{(1)} X_{ik} + \eta_j^{(1)} \right) + \eta^{(2)}.$$

Quanto mais camadas a rede apresentar, maior será a instabilidade numérica, na minimização do erro, e mais difícil será o cálculo da função aproximadora f . Se considerarmos apenas uma camada e uma única variável de resposta, então a rede neuronal supervisionada de regressão reduz-se ao metamodelo de regressão não linear com que iremos trabalhar. As redes neuronais supervisionadas de regressão são mais vantajosas que os metamodelos não lineares de regressão, no que diz respeito à capacidade

de adaptação à forma de diversas funções arbitrárias, com base em valores observados (critério 2). Contudo, elas são menos vantajosas em relação aos critérios 1, 4, 5 e 6.

2.2 Validação de metamodelos de regressão

Após a estimação dos parâmetros desconhecidos do metamodelo, este apenas pode ser utilizado para determinado fim (por exemplo, aproximação de uma superfície resposta, previsão de respostas no sistema real ou optimização do desempenho do sistema) se ele aproximar suficientemente bem o programa de simulação. Por conseguinte, após o metamodelo ter sido estimado, é de extrema importância averiguar se o hipotético metamodelo é, de facto, uma representação precisa do modelo de simulação. Com este objectivo e no contexto dos metamodelos lineares de regressão em simulação, têm sido propostas diversas técnicas estatísticas de validação utilizadas na teoria clássica da regressão linear.

Kleijnen propôs um teste de falta de ajustamento para validar metamodelos de regressão linear em simulação digital [59]. Este procedimento utiliza informação adicional, frequentemente disponível em simulação, para obter estimativas independentes da variância da resposta em cada ponto experimental. Ele considera uma única resposta e diversos factores de entrada. O contexto é qualquer técnica de análise de resultados em simulação em que o programa de simulação seja executado uma única vez tal como os métodos das médias de lotes, espectral, etc. O procedimento investigado neste artigo é baseado na seguinte aproximação, geralmente utilizada em ciência quando da validação de um modelo:

- (1) Estimar o modelo;
- (2) Usar o modelo estimado para prever novas observações;
- (3) Comparar os valores previstos com as observações novas reais. Entenda-se por observações novas reais uma ou mais novas observações não usadas na estimação dos

parâmetros do modelo. Para esta comparação, Kleijnen propôs a familiar estatística t de Student. Esta estatística inclui a variabilidade inerente das observações (de simulação) prevista e nova.

A estatística t é combinada com um 'truque' para obter muitas observações 'novas' para validação, chamado validação cruzada. Kleijnen esclarece que este truque é o mesmo que a validação cruzada discutida em muitos artigos estatísticos como, por exemplo, Allen [2] e Stone [112]. O que é novo, neste artigo, é a combinação que Kleijnen propõe entre a validação cruzada e o teste t . Para manter a taxa de erro válida ao longo de toda a experiência ¹, ele recorre à utilização da desigualdade de Bonferroni.

São realizadas diversas experiências de Monte Carlo para estudar os erros α e β do procedimento de validação proposto. Os resultados experimentais sugerem que este procedimento vale a pena, no sentido em que é suficientemente interessante para justificar o esforço que foi dispendido no seu estudo.

Para testar a validade do metamodelo hipotético, Kleijnen propôs a estatística t seguinte:

$$t = \frac{Y - \hat{Y}}{\hat{\sigma}_{Y-\hat{Y}}} \quad (2.10)$$

onde \hat{Y} é a previsão da regressão e $\hat{\sigma}_{Y-\hat{Y}}$ denota o desvio padrão estimado de $Y - \hat{Y}$. Este artigo investiga em detalhe o comportamento estatístico da estatística t proposta utilizando uma aproximação de Monte Carlo.

Passemos agora a explicar o método, de validação cruzada, proposto por Kleijnen. Suponhamos n execuções do programa de simulação produzindo

$$\{X_{ij}, Y_i, \hat{\sigma}_i : i = 1, \dots, n, j = 1, \dots, d\} \quad (2.11)$$

onde $\hat{\sigma}_i$ denota o desvio padrão estimado da resposta Y_i . O valor $\hat{\sigma}_i$ é baseado na execução i do programa de simulação utilizando uma técnica tal como o método das médias de lotes, método espectral, etc. Se cada execução do programa de simulação usar uma semente r diferente, então sabemos que as respostas Y_i são independentes. Por

¹*experimentwise error rate*

consequente, a matriz de variância-covariância $\Sigma_{\mathbf{Y}}$ de $\mathbf{Y}^T = (Y_1, \dots, Y_n)$ é dada por $\Sigma_{\mathbf{Y}} = \text{diag}[\sigma_1^2, \dots, \sigma_n^2]$, onde $\sigma_i^2 = E[\hat{\sigma}_i^2]$ com $i = 1, \dots, n$.

A previsão através do metamodelo de regressão, baseada em resultados de regressão padrão, é

$$\hat{Y}_i = \mathbf{X}_i \hat{\boldsymbol{\theta}} \quad (2.12)$$

onde $\hat{\boldsymbol{\theta}}$ é o estimador dos mínimos quadrados usuais

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (2.13)$$

A matriz de variância-covariância de $\hat{\boldsymbol{\theta}}$, $\Sigma_{\hat{\boldsymbol{\theta}}}$, é a seguinte:

$$\Sigma_{\hat{\boldsymbol{\theta}}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma_{\mathbf{Y}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \quad (2.14)$$

Por conseguinte

$$\text{Var}[\hat{Y}_i] = \mathbf{X}_i \Sigma_{\hat{\boldsymbol{\theta}}} \mathbf{X}_i^T. \quad (2.15)$$

Logo o denominador da equação (2.10) é

$$\hat{\sigma}_{Y-\hat{Y}} = \left(\hat{\sigma}_i^2 + \widehat{\text{Var}}[\hat{Y}_i] \right)^{1/2} \quad (2.16)$$

onde

$$\widehat{\text{Var}}[\hat{Y}_i] = \mathbf{X}_i \hat{\Sigma}_{\hat{\boldsymbol{\theta}}} \mathbf{X}_i^T, \quad (2.17)$$

com

$$\hat{\Sigma}_{\hat{\boldsymbol{\theta}}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\Sigma}_{\mathbf{Y}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1},$$

sendo $\hat{\Sigma}_{\mathbf{Y}} = \text{diag}[\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2]$ (ver (2.11), (2.14) e (2.15)). A equação (2.16) assume que Y e \hat{Y} são independentes. Esta exigência estatística é automaticamente satisfeita se o modelo de regressão é validado pelo seguinte método de validação cruzada:

(1) Eliminamos uma observação, por exemplo a última n , o que resulta num conjunto de $n - 1$ observações multidimensionais, denotado por

$$\left\{ \mathbf{X}_{(-n)}, \mathbf{Y}_{(-n)}, \hat{\Sigma}_{\mathbf{Y}_{(-n)}} \right\} \quad (2.18)$$

onde

$$\mathbf{X}_{(-n)} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1d} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n-1,1} & X_{n-1,2} & \dots & X_{n-1,d} \end{bmatrix}. \quad (2.19)$$

Além disso,

$$\mathbf{Y}_{(-n)}^T = (Y_1, \dots, Y_{n-1}) \quad (2.20)$$

e $\hat{\Sigma}_{\mathbf{Y}_{(-n)}}$ é uma matriz diagonal $(n-1)$ por $(n-1)$ obtida de $\hat{\Sigma}_{\mathbf{Y}}$ eliminando as linha e coluna n .

(2) Estimamos θ a partir das restantes observações $(n-1)$:

$$\hat{\theta}_{(-n)} = (\mathbf{X}_{(-n)}^T \mathbf{X}_{(-n)})^{-1} \mathbf{X}_{(-n)}^T \mathbf{Y}_{(-n)}, \quad (2.21)$$

sendo

$$\hat{\Sigma}_{\hat{\theta}_{(-n)}} = (\mathbf{X}_{(-n)}^T \mathbf{X}_{(-n)})^{-1} \mathbf{X}_{(-n)}^T \hat{\Sigma}_{\mathbf{Y}_{(-n)}} \mathbf{X}_{(-n)} (\mathbf{X}_{(-n)}^T \mathbf{X}_{(-n)})^{-1}. \quad (2.22)$$

(3) Prevemos a observação eliminada. Da equação (2.12) vemos que Y_n é prevista por

$$\hat{Y}_{n(-n)} = \mathbf{X}_n \hat{\theta}_{(-n)} \quad (2.23)$$

onde $\mathbf{X}_n = (X_{n1}, \dots, X_{nd})$. A variância estimada da previsão $\hat{Y}_{n(-n)}$ segue de (2.17):

$$\widehat{\text{Var}}[\hat{Y}_{n(-n)}] = \mathbf{X}_n \hat{\Sigma}_{\hat{\theta}_{(-n)}} \mathbf{X}_n^T. \quad (2.24)$$

Observe-se que Y_n e $\hat{Y}_{n(-n)}$ são estatisticamente independentes porque $\hat{Y}_{n(-n)}$, depende de $\hat{\theta}_{(-n)}$ e este não depende dos dados relativos ao ponto experimental n .

(4) A seriedade do erro previsto é medido por

$$t_n = \frac{Y_n - \hat{Y}_{n(-n)}}{\left(\hat{\sigma}_n^2 + \widehat{\text{Var}}[\hat{Y}_{n(-n)}] \right)^{1/2}}. \quad (2.25)$$

(5) Em seguida, a regra da observação n – eliminada no passo 1 – é aplicada para cada observação i' ($i' = 1, \dots, n-1$). Daqui resultam n observações dependentes para t definidas em (2.25); por exemplo, suponha-se que existe uma observação 'selvagem' y_3 , então ela afecta tanto $\hat{\theta}_{(-1)}$ como $\hat{\theta}_{(-2)}$, e torna t_1 e t_2 dependentes. Uma vez que o

modelo de regressão postulado deverá ser válido para todos os n pontos experimentais, o modelo de regressão é rejeitado sempre que o valor de alguma das n observações em t seja significativo. Por conseguinte, define-se a hipótese nula

$$H_0 : E[\hat{Y}_{i(-i)}] = E[Y_i] \quad i = 1, \dots, n \quad (2.26)$$

e rejeita-se H_0 se

$$\max_{i=1, \dots, n} |t_i| > t^{\alpha_c/2} \quad (2.27)$$

onde $t^{\alpha_c/2}$ é definido por

$$P(t > t^{\alpha_c/2}) = 1 - P(t < t^{\alpha_c/2}), \quad (2.28)$$

ou seja, é o valor t crítico ajustado para comparações múltiplas usando a desigualdade de Bonferroni. Isto é, para uma dada taxa de erro válida ao longo de toda a experiência, α_E , cada um dos n valores t_i é testado a um nível de significância α_c onde $\alpha_c = \alpha_E/n$. Por exemplo se $n = 8$ e $\alpha_E = 20\%$, então $\alpha_c = 2.5\%$ e $\alpha_c/2 = 1.25\%$. Em resumo

$$\alpha_c = \frac{\alpha_E}{n}. \quad (2.29)$$

Kleijnen salienta que se as n observações em t fossem independentes, então a desigualdade de Bonferroni não seria necessária. A desigualdade de Bonferroni é conservadora, porque garante que α_E é igual ou inferior a $n\alpha_c$. Para n estatísticas independentes t , é fácil garantir exactamente a taxa de erro válida ao longo de toda a experiência através da resolução de $(1 - \alpha_E) = (1 - \alpha_c)^n$. Além disso afirma que, para estatísticas t independentes, existem alternativas para $\max |t_i|$ (mas não necessariamente melhores), por exemplo, $\sum |t_i|$ (ver Reynolds e Deaton [98]).

É importante salientar que, se o número de observações é muito grande, então a validação cruzada pode ser restrita a menos que a permutação completa. Por exemplo, o metamodelo pode ser validado num conjunto de pontos experimentais seleccionados aleatoriamente. Além disso, ganhando tempo computacional e tempo de análise, a aproximação de Bonferroni torna-se menos conservadora. Contudo, é ignorada informação potencial.

Kleijnen observou que uma desvantagem do seu procedimento é a necessidade de fazer um número grande de regressões. Panis, Myers and Houck [94] estenderam o trabalho de Kleijnen [59] e demonstraram que é possível utilizar o teste de validação cruzada de Kleijnen fazendo uma única regressão. Além disso, também mostraram que a estatística, na qual o procedimento de Kleijnen se baseia, é análoga à estatística de R de Student utilizada frequentemente em diagnósticos de regressão. Embora os autores sigam a formulação do problema descrita em [59], uma resposta e vários factores, o estudo é feito no contexto do método das réplicas independentes. Outro resultado importante é a extensão do procedimento de Kleijnen para incluir mínimos quadrados ponderados (neste caso, eles também conseguem utilizar o procedimento com uma única regressão). É possível obter este resultado porque Kleijnen, Brent e Brouwers [63] descobriram que em simulação os mínimos quadrados ponderados, usando os pesos $1/\hat{\sigma}_i^2$, dão estimadores mais precisos para θ . Por conseguinte, a estatística t pode ser baseada em mínimos quadrados ponderados em vez de mínimos quadrados usuais. Tanto para os mínimos quadrados ponderados como para os usuais, são desenvolvidas extensões das ferramentas de diagnóstico de regressão modernas, para complementar as estratégias tradicionais de inferência para construção de modelos.

Considere-se o problema proposto por Kleijnen em [59], mas em que o estudo é feito com base no método das réplicas independentes. Como consequência, a resposta de interesse é tipicamente a média \bar{Y}_i . O metamodelo utilizado é um modelo linear

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}.$$

O procedimento de validação, sugerido por Kleijnen, necessita de n operações de eliminação de pontos experimentais. Embora fazer as n regressões seja relativamente barato do ponto de vista computacional, é um processo pouco eficiente e demorado, especialmente quando o número de observações é grande. Kleijnen sugere que, quando uma situação deste tipo acontece, pode-se considerar a permutação num subconjunto dos n pontos experimentais. Contudo, ele salienta que, uma desvantagem desta estratégia é a potencial perda de informação disponível. Felizmente, Pannis, Myers e Houck [94] conseguiram ultrapassar esta dificuldade demonstrando que a estatística (2.25) pode

ser escrita na forma

$$t_i = \frac{\bar{Y}_i - \hat{Y}_i}{\hat{\sigma}_i(1 - h_{ii})^{1/2}}, \quad (2.30)$$

onde h_{ii} é o i -ésimo elemento da diagonal principal da matriz "chapéu" $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ (ver [87, página 209]). Por conseguinte, o analista não tem a necessidade de fazer n regressões distintas.

Os autores salientam que, observando este resultado importante, conclui-se que a estatística- t é uma simples padronização do resíduo usual, $e_i = \bar{Y}_i - \hat{Y}_i$, calculado a partir do conjunto de observações completo. Com a exceção de $\hat{\sigma}_i$, a expressão anterior é a estatística R de Student usada em diagnósticos de regressão. Sob a hipótese de homogeneidade de variância $\hat{\sigma}_i$, para R de Student, é a raiz quadrada do erro quadrático médio calculado sem utilizar a i -ésima observação, enquanto que a estatística t de Kleijnen utiliza informação adicional, apenas presente na modelação no contexto da simulação, para produzir este estimador. Se as r observações, nas quais este estimador $\hat{\sigma}_i$ se baseia, forem independentes e normalmente distribuídas, então (2.30) segue uma distribuição t com $r - 1$ graus de liberdade, desde que a hipótese nula (2.26) do teste de Kleijnen se verifique.

Considere-se o caso em que se utiliza a estimação através dos mínimos quadrados usuais embora as variâncias dos erros sejam desiguais para observações correspondentes a pontos experimentais distintos. Nesta situação, ainda podemos utilizar o resultado (2.30), uma vez que os h_{ii} continuam a ser os elementos da diagonal principal da matriz "chapéu" $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Isto é, não precisamos fazer n regressões para calcular t_i , $i = 1, \dots, n$.

Contudo, em caso de heterogeneidade da variância nos pontos experimentais, é mais usual estimar a matriz de variância-covariância $\Sigma_{\mathbf{Y}}$. O estimador que se obtém, ao substituir $\Sigma_{\mathbf{Y}}$ por $\hat{\Sigma}_{\mathbf{Y}} = \text{diag}[\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_n^2]$ no estimador dos mínimos quadrados ponderados

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y},$$

designa-se por estimador dos mínimos quadrados ponderados estimados e é dado por:

$$\tilde{\boldsymbol{\theta}} = \left(\mathbf{X}^T \hat{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \hat{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{Y}. \quad (2.31)$$

Definindo

$$\hat{\mathbf{W}} = \text{diag}[1/\hat{\sigma}_1, 1/\hat{\sigma}_2, \dots, 1/\hat{\sigma}_n],$$

Pannis, Myers e Houck [94] adaptam, da seguinte forma, o procedimento de validação cruzada de Kleijnen:

(1) Calcular o estimador dos mínimos quadrados ponderados estimados (2.31).

(2) Construir a matriz "chapéu" modificada

$$\mathbf{H}^* = \hat{\mathbf{W}}\mathbf{X} \left(\mathbf{X}^T \hat{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \hat{\mathbf{W}}^T.$$

(3) Calcular os valores

$$t_i^* = \frac{\bar{Y}_i - \hat{Y}_i}{\hat{\sigma}_i (1 - h_{ii}^*)^{1/2}}, \quad i = 1, 2, \dots, n.$$

O teste de Kleijnen é realizado com base em (2.27), com t_i substituído por t_i^* .

Em [94] salienta-se que, as simplificações computacionais que são introduzidas para t_i , no caso do método dos mínimos quadrados usuais, e t_i^* no caso dos mínimos quadrados ponderados, necessitam que a matriz de dispersão das respostas $\Sigma_{\mathbf{Y}}$, ou a sua matriz estimada $\hat{\Sigma}_{\mathbf{Y}}$, seja diagonal. Como consequência, estas técnicas computacionais apenas podem ser aplicadas a sequências independentes.

Os autores consideram particularmente importante compreender o papel desempenhado pelas técnicas de diagnóstico na construção de metamodelos de simulação. Referem que Myers [83] observa que estes diagnósticos são um suplemento inestimável para testes de hipóteses formais e técnicas de ajuste de modelos. Isto deve-se ao facto de outras ferramentas falharem frequentemente na construção de modelos utilizáveis. Quando rejeitamos um modelo num sentido tradicional, os procedimentos padrão, tais como testes de hipóteses, falham no que diz respeito a dar informação de porque é que o modelo falha ou a sugerir representações alternativas. Mesmo quando o modelo hipotético não é rejeitado, podemos sempre argumentar que é necessária avaliação adicional. Eles crêem que as ferramentas de diagnóstico permitem preencher o espaço deixado vazio, com muito pouco esforço computacional.

Uma alternativa, para testar se o metamodelo de regressão é válido, é o teste F de falta de ajustamento ² que compara dois estimadores com a mesma variância, σ^2 (ver, por exemplo, [66] [57] e [61]). Este teste, na literatura clássica sobre projecto de experiências, assume que $\mathcal{D}[\mathbf{Y}] = \sigma^2\mathbf{I}$. Kleijnen [61] levou a cabo uma experiência de Monte Carlo para estudar alguns exemplos em que esta condição se verifica. O teste F clássico compara os resíduos estimados (reflectindo falta de ajustamento) com o estimador de variância $\hat{\sigma}^2$:

$$F_{n-m, n(r-1)} = \frac{m}{(n-m)\hat{\sigma}^2} (\bar{\mathbf{Y}} - \mathbf{X}\hat{\boldsymbol{\theta}})^T (\bar{\mathbf{Y}} - \mathbf{X}\hat{\boldsymbol{\theta}}), \quad (2.32)$$

onde

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_i^2.$$

Rao [97] generalizou este teste para poder ser utilizado com mínimos quadrados generalizados:

$$F_{n-m, r-n+m} = \frac{r(r-n+m)}{(n-m)(r-1)} (\bar{\mathbf{Y}} - \mathbf{X}\hat{\boldsymbol{\theta}})^T \hat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{Y}} - \mathbf{X}\hat{\boldsymbol{\theta}}), \quad (2.33)$$

onde $\hat{\boldsymbol{\Sigma}}$ é dado por (2.7). Neste teste, assume-se que $n > m = \text{rank}(\mathbf{X})$ e $r > n$, para que $\hat{\boldsymbol{\Sigma}}$ seja não singular. Qualquer que seja o aspecto da matriz de variância-covariância, um ajuste perfeito (ou seja, $\bar{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\theta}}$) implica $F = 0$ e, por conseguinte, o metamodelo não é rejeitado. Além disso, quando r tende para infinito, tanto (2.32) como (2.33) convergem em distribuição para $\chi_{n-m}^2/(n-m)$ se o metamodelo assumido for correcto. Nos casos em que a variância dos erros não varia ao longo dos pontos experimentais, é preferível utilizar a estatística (2.32), em vez de (2.33), uma vez que (2.32) tem maior potência de teste: $n(r-1) > r-n+m$. Uma exigência técnica para ambos os testes é que $n > m$.

Kleijnen utilizou o teste de Rao no contexto dos números aleatórios comuns em simulação [61] e provou, através de uma experiência exaustiva de Monte Carlo, que este teste é melhor que o teste de validação cruzada [59] se as respostas do modelo de simulação tiverem distribuição simétrica, por exemplo, distribuição normal ou uniforme; se, por outro lado, as respostas possuírem uma distribuição assimétrica, por exemplo log-normal, então é melhor utilizar o teste de validação cruzada. Além disso, apresentou

²lack-of-fit

e comparou diversos procedimentos para construção de intervalos de confiança. No contexto da estratégia de Schruben-Margolin e no caso em que a distribuição dos erros é normal multivariada, mas a estrutura de covariância Σ é desconhecida, Tew e Wilson [113] também propõem o teste de Rao de falta de ajustamento. Embora estes autores não tenham realizado um estudo tão exaustivo quanto o de Kleijnen, eles confirmaram as conclusões gerais nos vários estudos limitados que fizeram (ver também [111]).

Para terminar esta secção, referimos um artigo que salienta a utilidade de desenvolver um metamodelo não linear, como modelo auxiliar do modelo de simulação, e que dá ênfase à importância de validar o metamodelo com o objectivo de determinar se, de facto, ele aproxima convenientemente bem a relação entre os dados e os resultados de um programa de simulação [37]. O modelo (ou programa) de simulação possui uma única resposta, diversos factores e é estudado no contexto das réplicas independentes. O exemplo utilizado para ilustração é o sistema M/M/s, com 1 e 2 servidores e uma única fila de espera. Os clientes chegam ao sistema de acordo com um processo de Poisson com taxa de chegadas constante λ e os tempos de serviço seguem uma distribuição exponencial com média $1/\mu$. A medida de desempenho que nos interessa analisar é a média do comprimento da fila $E[CF]$ (número de clientes na fila que esperam atendimento). A duração das simulações foi de quinze semanas e foram simuladas seis configurações diferentes do sistema, utilizando diversos valores de λ , μ e s . Foram realizadas 20 réplicas de cada configuração do sistema (ponto experimental), resultando em 120 observações.

A contribuição mais importante, no contexto deste trabalho, é a utilização de um metamodelo não linear para ajustar um metamodelo de simulação. Esse metamodelo, embora sendo não linear, é linearizável. Por conseguinte, o estudo é feito a partir do metamodelo linearizado sendo depois aplicados procedimentos de regressão linear. Também é importante a discussão e ilustração do uso de duas técnicas de validação, uma em que um terço da amostra é posta de parte e o chamado método de validação cruzada dupla.

Numa primeira abordagem, os autores tentaram um ajuste através de um metamodelo linear de regressão, com $E[CF]$ como variável dependente e, como variáveis indepen-

dentos, os três efeitos λ , μ e s e alguns efeitos de interacção. Através do teste F de falta de ajustamento, chegaram à conclusão que o metamodelo hipotético não aproximava de forma alguma a relação entre os dados e os resultados do programa de simulação. Pensaram então que a presença de um termo de falta de ajustamento significativo poderia ser devido a uma selecção inadequada do metamodelo, por exemplo o uso de um metamodelo linear em vez de não linear, ou a omissão de uma ou mais variáveis cruciais. O erro puro deve-se a flutuações aleatórias da variável resposta. Os autores salientam que este tipo de mau ajustamento é frequente quando se ajustam resultados de simulação provenientes de sistemas de filas de espera onde λ , μ e s se relacionam de forma não linear. Por conseguinte, propuseram a seguinte relação não linear:

$$E[CF]_i = \alpha \frac{\lambda_i^{\theta_2}}{\mu_i^{\theta_3} s_i^{\theta_4}} \nu_i, \quad i = 1, \dots, n, \quad (2.34)$$

onde ν_i é o erro e α , θ_2 , θ_3 e θ_4 são os parâmetros do modelo. A linearização do modelo anterior é obtida aplicando logaritmos em ambos os membros:

$$\ln E[CF]_i = \ln \alpha + \theta_2 \ln \lambda_i - \theta_3 \ln \mu_i - \theta_4 \ln s_i + \ln \nu_i.$$

Este modelo linearizado pode ser escrito na forma

$$Y_i = \theta_1 + \theta_2 X_{i2} - \theta_3 X_{i3} - \theta_4 X_{i4} + \epsilon_i, \quad (2.35)$$

onde foi feita a mudança de variáveis: $Y_i = \ln E[CF]_i$, $\theta_1 = \ln \alpha$, $X_{i2} = \ln \lambda_i$, $X_{i3} = \ln \mu_i$, $X_{i4} = \ln s_i$ e $\epsilon_i = \ln \nu_i$. O modelo (2.34) não foi rejeitado. Para tomar esta decisão, foram realizados o teste F (para testar a falta de ajustamento) e os testes com um terço da amostra posta de parte e de dupla validação cruzada (para testar a capacidade previsiva do metamodelo).

2.3 Técnicas de redução de variância

Originalmente, as técnicas de redução de variância foram desenvolvidas, nos primeiros anos dos computadores, com o objectivo de reduzir o erro em problemas de Monte

Carlo, tais como o cálculo de integrais múltiplos e o estudo de distribuições amostrais [46]. Contudo, muitas destas técnicas originais de redução de variância não podem ser aplicadas directamente a experiências de simulação de sistemas dinâmicos complexos. Nesta secção, é dada uma descrição geral sobre as técnicas de redução de variância que foram desenvolvidas para, ou adaptadas à, simulação digital. Além disso, são dadas referências e discutem-se os trabalhos mais recentes ou importantes nesta área (de acordo com o ponto de vista e conhecimento do autor). Estudos e descrições mais aprofundadas encontram-se, por exemplo, nos trabalhos de L'Ecuyer[75], Nelson e Schmeiser [86] e Wilson [123]. Estes trabalhos contêm vastas bibliografias e, em particular, os dois últimos fornecem boas taxonomias de técnicas de redução de variância. Também outras boas introduções nesta área podem ser encontradas, por exemplo, nos livros de Law e Kelton [73, Capítulo 11] e [54, Capítulo 3], entre outros.

Frequentemente, o objectivo de uma experiência de simulação estocástica é estimar a média da resposta Y ($\mu_Y = E[Y]$). A maneira mais simples de estimar μ_Y , a partir da sequência de observações recolhidas Y_1, Y_2, \dots, Y_n , é calcular a média aritmética:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

O *enviesamento*, a *variância* e o *erro quadrático médio* (MSE) do estimador \bar{Y} são respectivamente,

$$\begin{aligned} \text{Enviesamento}[\bar{Y}] &= E[\bar{Y} - \mu_Y] \\ \sigma^2[\bar{Y}] &= E[(\bar{Y} - E[\bar{Y}])^2] \\ \text{MSE}[\bar{Y}] &= E[(\bar{Y} - \mu_Y)^2]. \end{aligned}$$

Note-se que, a partir destas definições,

$$\text{MSE}[\bar{Y}] = (\text{Enviesamento}[\bar{Y}])^2 + \sigma^2[\bar{Y}].$$

Assume-se que o custo (isto é, tempo de CPU) para calcular \bar{Y} é também uma variável aleatória $C(\bar{Y})$. Definimos *eficiência* de \bar{Y} por (ver, por exemplo, [75])

$$\text{Eff}(\bar{Y}) = \frac{1}{\text{MSE}[\bar{Y}] \cdot E[C(\bar{Y})]}. \quad (2.36)$$

Neste contexto, para dois estimadores \bar{Y} e $\bar{Y}(\delta)$, dizemos que $\bar{Y}(\delta)$ é *mais eficiente* que \bar{Y} se $\text{Eff}(\bar{Y}) < \text{Eff}(\bar{Y}(\delta))$. Melhorar a eficiência significa encontrar um outro estimador, $\bar{Y}(\delta)$, que seja mais eficiente que o utilizado mais vulgarmente \bar{Y} . Em geral, ambos os estimadores são não enviesados e têm aproximadamente os mesmos custos computacionais. Por conseguinte, melhorar a eficiência é equivalente a reduzir a variância. Contudo, a eficiência pode por vezes ser melhorada através do aumento de variância [75], mas este tema está fora do contexto da nossa investigação.

O critério de eficiência (2.36) é uma das muitas possibilidades de medidas de eficiência, mas é um dos mais utilizados na literatura especializada [43]. Ele está em concordância com a hipótese de não enviesamento dos estimadores. Mais precisamente, na ausência de enviesamento, este critério é invariante em relação ao número de réplicas para calcular o estimador; se se duplicar o número de réplicas então cai para metade o valor da variância e duplica o esforço computacional, logo o valor de Eff não é alterado.

As técnicas de redução de variância podem ser agrupadas em três classes principais, as quais são descritas nas três secções seguintes: métodos de correlação, métodos de importância e métodos de substituição; é também usual combinar vários destes métodos simultaneamente.

1. Métodos de correlação Os métodos de correlação incluem três técnicas que tiram proveito da correlação linear entre as respostas do modelo de simulação, para assim se conseguir um aumento de eficiência. As técnicas dos **números aleatórios comuns** e das **variáveis antitéticas** necessitam que o experimentador induza correlação positiva e negativa respectivamente, dentro dos blocos das execuções do programa de simulação, por forma a forçar uma dependência funcional ou estocástica apropriada entre as variáveis de entrada utilizadas em cada execução do programa de simulação. Por esse motivo, são conhecidas como métodos de correlação induzida. O método dos números aleatórios comuns utiliza-se para comparar um ou mais sistemas relacionados, enquanto que o segundo, tal como todos os outros métodos de redução de variância, é utilizado para estimar uma medida de desempenho de um sistema individual. Em

alternativa à indução de correlação, a técnica das **variáveis de controlo** utiliza métodos de regressão para explorar uma determinada correlação inerente entre a resposta e um determinado vector de variáveis aleatórias concomitantes, com média conhecida, que é observado em cada execução do programa de simulação.

O método dos números aleatórios comuns é normalmente a técnica mais utilizada para comparar dois ou mais sistemas alternativos, sujeitando-os ao mesmo ambiente experimental (ver, por exemplo, [125]). A ideia é que se deve comparar configurações alternativas sob condições experimentais semelhantes de forma a que se possa ter mais confiança que as diferenças observadas no desempenho são devidas a diferenças na configuração do sistema e não a flutuações das condições experimentais. Em simulação digital, estas condições experimentais são consequência das variáveis aleatórias geradas para conduzir as simulações. O nome deste método vem da possibilidade de utilizar, em muitas situações, os mesmos números aleatórios (com distribuição uniforme entre zero e um) para simular ao longo do tempo cada uma das configurações alternativas. Embora esta técnica nem sempre resulte, Glasserman e Yao [40] estudaram um tipo de situação no qual ela funciona extremamente bem.

A ideia do método das variáveis antitéticas é construir pares de execuções do programa de simulação, de tal forma que uma observação, com valor baixo numa das execuções do par, tende a ser compensada por outra, com um valor elevado na outra execução. Neste caso, as duas observações estão correlacionadas negativamente. Assim, ao usar a média das duas observações, como um ponto para análise, este tenderá a estar mais perto da média, que se pretende estimar, que cada uma das observações individuais. Este método tenta induzir correlação negativa, gerando números aleatórios complementares para conduzir as simulações dentro de cada par. Crawford e Gallwey [27] descrevem este método, incluindo bibliografia sobre trabalhos importantes, desde a altura do seu aparecimento em 1956.

O método das variáveis de controlo foi referenciado, pela primeira vez, em 1937 [117]. Como já foi referido anteriormente, a redução de variância obtém-se utilizando técnicas estatísticas de análise de regressão. O método das variáveis de controlo é uma das

mais populares técnicas de redução de variância, porque as variáveis de controlo concomitantes estão sempre disponíveis em simulação digital (ver, por exemplo, [70], [116], [89], [85], [90] e [48]). Por conseguinte, não existe a necessidade de alterar o plano de amostragem. É uma técnica *a posteriori* para reduzir o erro dos estimadores e permite-nos observar o comportamento natural (não forçado) do sistema ao longo do tempo (simulado).

Também é possível obter redução de variância através da utilização de variáveis de controlo externas, ou seja, variáveis observadas em experiências de simulação auxiliares. Este caso exige experimentação com versões simplificadas do sistema. A grande desvantagem é que a simplificação pode perturbar a relação entre as variáveis de controlo e as respostas do modelo original. Além disso, torna a técnica mais dispendiosa.

2. Métodos de importância Esta classe de métodos inclui, a **amostragem de importância**, a **amostragem estratificada**, a **amostragem sistemática**, o **Monte Carlo condicional** e a **amostragem hipercúbica latina**; pode-se encontrar uma discussão detalhada sobre estes métodos em [123]. Numa simulação, a maior parte da contribuição para o valor global do parâmetro de interesse vem, de uma maneira geral, de um ou mais subconjuntos *importantes* do domínio dos dados (espaço amostral). A ideia destes métodos é concentrar a amostragem nestes subconjuntos (mudando a(s) lei(s) de probabilidade) e, por conseguinte, obter estimadores mais precisos. Existem diversos métodos pertencentes à classe da amostragem de importância, mas a **roleta russa** e a **repartição** são os que melhor se adaptam à simulação com acontecimentos discretos, porque são aplicados dinamicamente ao longo do tempo de simulação [80]. Enquanto que a técnica das variáveis de controlo é eficaz quando existe uma relação linear forte entre a resposta e as variáveis de controlo (com média conhecida), a amostragem estratificada explora o total conhecimento *a priori* das distribuições das variáveis de controlo, conseguindo-se a redução de variância, qualquer que seja a relação entre a resposta e as variáveis de controlo. Como uma alternativa para assegurar que as regiões de importância estão devidamente representadas, a amostragem sistemática força amostragem uniforme ao longo do domínio de interesse através de o dividir em estratos congruentes, de tal forma que

um ponto amostrado aleatoriamente num estrato identifica automaticamente o ponto amostral correspondente em cada estrato. O método Monte Carlo condicional consiste em converter um problema de estimação, expresso como uma esperança condicionada, noutro problema, expresso como uma esperança não condicionada (ou vice-versa). A amostragem hipercúbica latina foi sugerida como uma técnica de redução de variância, mas também pode ser vista como uma técnica de varrimento ³. Nesta técnica, a selecção dos valores amostrais é fortemente controlada, embora ainda os deixando variar. A ideia base deste método consiste em fazer uma estratificação completa da distribuição amostral com uma selecção aleatória dentro de cada estrato.

3. Métodos de substituição Existem, referenciados na literatura, dois métodos de substituição: **esperanças condicionadas** e **estimação indirecta**. O método das esperanças condicionadas é frequentemente denominado de Monte Carlo condicional. Contudo, Monte Carlo condicional também se refere a uma técnica de amostragem desenvolvida por Trotter e Tukey em 1956 e que foi referida atrás como pertencente à classe dos métodos de importância. Aqui, o termo fica reservado para essa técnica e a expressão esperanças condicionadas é utilizada para o que Law e Kelton [73, Secção 11.6] designam por condicionamento. Esta técnica tira partido de alguma propriedade especial, existente na estrutura do modelo, com o objectivo de substituir um estimador, de uma determinada quantidade, pelo seu valor analítico exacto. Eliminando esta componente aleatória, espera-se que a variância da resposta seja menor (vários exemplos desta técnica podem ser encontrados, por exemplo, em [73]).

A estimação indirecta foi desenvolvida no âmbito da simulação de filas de espera para estudar medidas de desempenho tais como, por exemplo, as médias dos tempos de espera na fila e no sistema e o comprimento médio da fila. Alguns autores como, por exemplo, Carson e Law [18] e Glynn e Whitt [42] apresentam demonstrações que provam que se obtêm reduções de variância neste tipo de modelos, mas a ideia base do método pode ser aplicável a outros problemas. Suponha-se que pretendemos estimar

³*screening*

uma determinada média, μ , que pode ser expressa como uma função, conhecida, de uma outra quantidade, ρ : $\mu = f(\rho)$. Pode ser mais eficiente estimar ρ , em vez de μ , e depois aplicar a transformação f ao estimador de ρ .

2.3.1 Análise comparativa

Nesta secção, apresenta-se uma discussão sobre as diversas técnicas de redução de variância.

A grande desvantagem dos métodos de importância é que recorrem à distorção da(s) lei(s) probabilística(s) que rege(m) os dados. Por conseguinte, não devem ser aplicados se se pretender observar o comportamento natural do sistema. Outra grande desvantagem, talvez a maior, é que estes métodos são, em geral, extremamente sensíveis à escolha da distribuição de importância. Uma má escolha pode facilmente produzir um acréscimo na variância [75, Secção 6]. A distribuição de importância é escolhida por forma a imitar a relação entre os dados (aleatórios) e a resposta do sistema. Deste modo, é necessário ter um conhecimento detalhado dessa relação. Tal é quase impossível em simulações realistas de acontecimentos discretos, onde o comportamento dinâmico ao longo do tempo (simulado) pode ser extremamente complexo.

Os métodos de substituição têm a desvantagem de não se poder assegurar uma redução de variância. Estas técnicas são muito dependentes do modelo. Por exemplo, ao usar a estimação indirecta, no caso das filas de espera, existem relações teóricas que permitem descobrir f . Mas, no caso de outros modelos mais gerais, isso pode não ser fácil.

Os métodos de correlação induzida têm a desvantagem, tal como os métodos de importância, de exigir a distorção do processo aleatório de entrada de dados. Além disso, nem sempre se consegue garantir a correlação positiva ou negativa e, em certos casos, quando se consegue, pode ser extremamente trabalhoso calcular a covariância *a priori*.

Utilizando o método das variáveis de controlo, é possível conseguir reduções na variância apreciáveis através da escolha cuidadosa de variáveis de controlo apropriadas, no sentido em que se relacionam moderadamente com a resposta [90]. Uma desvantagem

desta técnica é que, se a correlação entre a resposta e as variáveis de controlo tiver um valor baixo, então pode ocorrer um aumento de variância.

A técnica das variáveis de controlo é uma das técnicas mais promissoras, dentro dos métodos de redução de variância, em termos do seu potencial uso alargado e foi ela a escolhida para esta investigação. Outras razões, para esta escolha, são a grande generalidade do método, a actual investigação na área, a adequação da técnica à simulação digital e o grande potencial que ela parece ter nas aplicações práticas [92, 89, 14]. Além disso, esta técnica não é intrusiva (não requer a distorção do processo estocástico) e é fácil de implementar.

A área das técnicas de redução de variância é de grande actualidade, uma vez que não param de surgir trabalhos de investigação nesta área. Existem duas formas de obter maior redução de variância em experiências de simulação: uma é generalizar métodos de redução de variância conhecidos e outra é utilizar **métodos combinados**, isto é, usar diversas técnicas de redução de variância simultaneamente. Por exemplo, foi referido anteriormente que uma das condições do método das variáveis de controlo é que a média das variáveis de controlo seja conhecida. Schmeiser, Taaffe e Wang [102] estudaram a técnica relaxando esta hipótese. Esta relaxação, ou método das variáveis de controlo enviesadas, obtém-se substituindo as médias desconhecidas por aproximações das mesmas. Os autores concluíram que, quando o coeficiente de controlo óptimo (coeficiente que minimiza a variância do estimador obtido através do método das variáveis de controlo enviesadas) é conhecido, o valor da eficiência é elevado se a correlação induzida também for, sendo a redução dependente do tempo de computação, duração da simulação e erro de análise (diferença entre o vector média desconhecido e a sua aproximação). Quando o coeficiente de controlo é desconhecido, eles analisaram dois estimadores e descobriram que só havia vantagem, na utilização da técnica proposta, em determinados casos particulares de experiências de simulação. Contudo, os autores crêem que é possível melhorar esta técnica, encontrando melhores estimadores para o coeficiente de controlo óptimo. Shih e Song [110] provaram que, sob determinadas condições, a aplicação do método das variáveis de controlo (mas em que as variáveis de

controlo, em vez de serem independentes, são correlacionadas) pode melhorar a precisão dos estimadores; contudo, os resultados empíricos que obtiveram não mostraram melhoria de precisão significativa. Avramidis e Wilson [7] apresentaram uma técnica de indução de correlação negativa que é uma generalização do método das variáveis antitéticas. Esta técnica tem como casos particulares o método das variáveis antitéticas e a amostragem hipercúbica latina e garante a redução de variância sob determinadas condições de regularidade.

Em relação aos métodos combinados, Kwon e Tew [68] apresentaram três métodos, para combinar as técnicas das variáveis antitéticas e variáveis de controlo, que podem ser facilmente realizadas numa experiência de simulação, sem trabalho adicional significativo de programação (ver, também, [124] e [93]). Sob determinadas condições gerais, estes métodos combinados têm desempenho superior (estimador com menor variância), comparativamente com o método das variáveis de controlo. Os resultados teóricos foram suportados por experimentação numérica que também mostrou que, se as hipóteses exigidas não forem satisfeitas, então pode não se obter qualquer melhoria em relação ao método das variáveis de controlo.

2.4 Método das variáveis de controlo

A situação mais simples, em que se aplicam variáveis de controlo como técnica de redução de variância, é quando se utiliza uma única variável de controlo (concomitante com média conhecida) e a resposta com interesse é univariada.

Na Figura 2.1, ilustra-se o caso mais simples, em que o objectivo é estimar a média da resposta Y ($\mu_Y = E[Y]$). Um exemplo desta situação poderia ser o estudo da média do tempo de espera para os primeiros $n = 500$ clientes numa fila $M/M/1$ (Y), utilizando como variável de controlo o tempo médio de serviço na fila (C). Vimos anteriormente

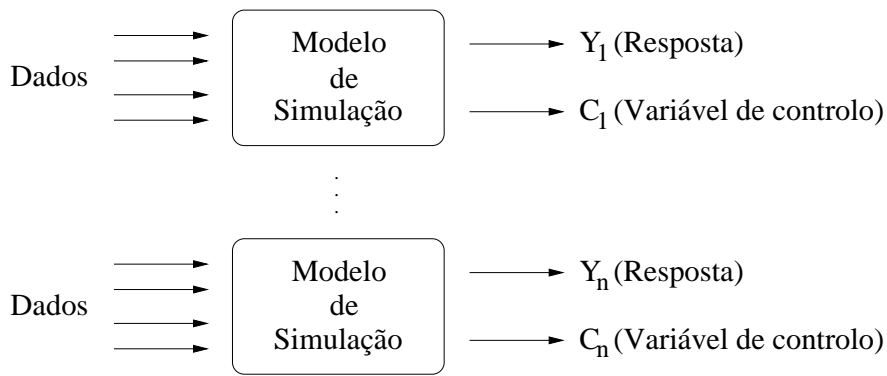


Figura 2.1: Conceito de variável de controlo.

que, a maneira mais simples de estimar μ_Y , a partir da sequência de observações recolhidas Y_1, Y_2, \dots, Y_n , é calcular a média aritmética das observações:

$$\hat{\mu}_Y = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}.$$

A variância de $\hat{\mu}_Y$ é dada por $\text{Var}[\hat{\mu}_Y] = \sigma_Y^2/n$. Se se tiver em conta a variável de controlo, então consideram-se as observações *controladas*

$$Y_i(\phi) = Y_i - \phi(C_i - \mu_C), \quad i = 1, 2, \dots, n,$$

o que dá origem ao estimador *controlado*

$$\hat{\mu}_Y(\phi) = \frac{1}{n} \sum_{i=1}^n Y_i(\phi) = \bar{Y}(\phi).$$

Para qualquer escalar fixo ϕ , o estimador controlado

$$Y(\phi) = Y - \phi(C - \mu_C)$$

é não enviesado

$$E[Y(\phi)] = E[Y]$$

com variância

$$\text{Var}[Y(\phi)] = \text{Var}[Y] + \phi^2 \text{Var}[C] - 2\phi \text{Cov}[Y, C]. \quad (2.37)$$

O coeficiente de controlo óptimo (no sentido em que torna a variância mínima), δ , pode ser obtido minimizando (2.37), ou seja, resolvendo a equação

$$\left. \frac{d}{d\phi} \text{Var}[Y(\phi)] \right|_{\phi=\delta} = 0,$$

cuja solução é:

$$\delta = \frac{\text{Cov}[Y, C]}{\text{Var}[C]}. \quad (2.38)$$

A variância mínima, que pode ser obtida através de uma resposta controlada, é dada por

$$\text{Var}[Y(\delta)] = \text{Var}[Y](1 - \rho_{YC}^2),$$

onde

$$\rho_{YC} = \frac{\text{Cov}[Y, C]}{\sqrt{\text{Var}[Y]\text{Var}[C]}}$$

é o *coeficiente de correlação* entre Y e C .

A quantidade $1 - \rho_{YC}^2$ é a chamada razão mínima entre variâncias. É o factor pelo qual a variância de Y , $\text{Var}[Y]$, poderá ser reduzida se o coeficiente δ for conhecido. Na maior parte das situações δ tem de ser estimado. Para isso, procede-se da seguinte forma: Se Y e C tiverem distribuição normal conjunta, a *distribuição condicional de Y dado C* é também normal:

$$Y|C = c \sim N(\mu_Y + \delta(c - \mu_C), \sigma_\epsilon^2),$$

com $\sigma_\epsilon^2 = \sigma_Y^2(1 - \rho_{YC}^2)$ e $\sigma_Y^2 = \text{Var}[Y]$. Por conseguinte, existe uma *regressão linear de Y em C* . Dado $C_i = c_i$ tem-se

$$Y_i = \mu_Y + \delta(c_i - \mu_C) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

onde os resíduos ϵ_i são variáveis aleatórias independentes e identicamente distribuídas, com distribuição $\epsilon_i \sim N(0, \sigma_\epsilon^2)$.

Para estimar os parâmetros desconhecidos, μ_Y e δ , pode-se utilizar os métodos da máxima verosimilhança ou mínimos quadrados (quer a distribuição dos erros seja normal ou não, o estimador de máxima verosimilhança coincide com o estimador obtido através do método dos mínimos quadrados). O método dos mínimos quadrados, para estimar parâmetros dos modelos, tem sido largamente utilizado em estatística. A sua maior vantagem é ser facilmente tratável analiticamente. Além disso, os estimadores dos mínimos quadrados possuem algumas propriedades convenientes, como por exemplo, o facto de, sob determinadas condições, serem os estimadores não enviesados com variância mínima.

Após aplicação do método dos mínimos quadrados, obtém-se:

$$\hat{\mu}_Y(\hat{\delta}) = \bar{Y}(\hat{\delta}) = \bar{Y} - \hat{\delta}(\bar{c} - \mu_C)$$

com

$$\hat{\delta} = \frac{\sum_{i=1}^n (c_i - \bar{c})(Y_i - \bar{Y})}{\sum_{i=1}^n (c_i - \bar{c})^2}.$$

Observe-se que $\hat{\delta}$ é simplesmente o estimador amostral do coeficiente óptimo de controlo (2.38).

Para obter uma estimativa do erro, calcula-se o *quadrado da média dos resíduos*

$$\hat{\sigma}_\epsilon^2 = \frac{\epsilon^T \epsilon}{n-2} = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (2.39)$$

onde $\hat{Y}_i = \hat{\mu}_Y(\hat{\delta}) + \hat{\delta}(c_i - \mu_C)$, com $i = 1, 2, \dots, n$. Pela teoria da regressão, a variância da intercepção, com o eixo dos YY' s, $\hat{\mu}_Y(\hat{\delta})$ é dada por

$$\text{Var} \left[\hat{\mu}_Y(\hat{\delta}) | C_i = c_i \right] = \sigma_\epsilon^2 \frac{\sum_{i=1}^n (c_i - \mu_C)^2}{n \sum_{i=1}^n (c_i - \bar{c})^2}, \quad i = 1, 2, \dots, n.$$

Uma vez que não é conhecido o valor exacto de σ_ϵ^2 , apenas se consegue calcular uma estimativa da variância anterior:

$$\text{Var} \left[\hat{\mu}_Y(\hat{\delta}) | C_i = c_i \right] = \hat{\sigma}_\epsilon^2 \frac{\sum_{i=1}^n (c_i - \mu_C)^2}{n \sum_{i=1}^n (c_i - \bar{c})^2}, \quad i = 1, 2, \dots, n,$$

onde σ_ϵ^2 foi substituído pelo seu estimador (quadrado da média dos resíduos ou erros) dado em (2.39).

Seguindo Lavenberg, Moeller e Welch [69], para se ter uma ideia da eficiência do método, mede-se a redução de variância conseguida utilizando variáveis de controlo através da *razão entre variâncias*, η . Se for conhecido o coeficiente de controlo óptimo, δ , a razão mínima entre variâncias, $\eta(\delta)$ é dada por

$$\eta(\delta) = \frac{\text{Var}[\bar{Y}(\delta)]}{\text{Var}[\bar{Y}]}. \quad (2.40)$$

O valor desta razão indica a máxima redução de variância que é possível obter com variáveis de controlo. Contudo, tal como foi referido anteriormente, δ necessita ser

estimado. Por conseguinte, em geral, é apenas possível calcular a razão entre variâncias associada ao estimador $\hat{\delta}$:

$$\eta(\hat{\delta}) = \frac{\text{Var}[\bar{Y}(\hat{\delta})]}{\text{Var}[\bar{Y}]} \quad (2.41)$$

A degradação da redução de variância, pelo facto de δ ter de ser estimado, pode ser medida pelo *factor de perda*:

$$\text{FP}(\hat{\delta}) = \frac{\eta(\hat{\delta})}{\eta(\delta)} \quad (2.42)$$

Neste exemplo, obtém-se (ver [69])

$$\eta(\delta) = 1 - \rho_{YC}^2,$$

$$\eta(\hat{\delta}) = \frac{n-2}{n-3}(1 - \rho_{YC}^2),$$

e

$$\text{FP}(\hat{\delta}) = \frac{n-2}{n-3},$$

onde ρ_{YC}^2 é o coeficiente de correlação entre a resposta e a variável de controlo.

Lavenberg, Moeller e Welch [69] estudaram a aplicação de diversas variáveis de controlo, como técnica de redução de variância, na obtenção de um estimador pontual e na construção de intervalos de confiança para a média de uma única resposta de um modelo de simulação. Este estudo foi realizado no contexto do método das réplicas independentes. Os autores definiram as seguintes medidas de eficiência para o estimador controlado da média: *i*) a razão mínima entre variâncias, que é o complemento da percentagem máxima de redução na variância, que pode ser obtida utilizando um determinado conjunto de variáveis de controlo – ver (2.40); e *ii*) o factor de perda, que é um menos a percentagem de aumento de variância, que ocorre quando o coeficiente de controlo óptimo, δ , é estimado pelo método dos mínimos quadrados – ver (2.42). Por conseguinte, a variância do estimador controlado é o produto da razão mínima entre variâncias, do factor de perda e da variância do estimador directo (estimador obtido sem o auxílio das variáveis de controlo). Na formulação básica sobre variáveis de controlo, apresentada na secção anterior, tem-se $\text{Var}[\bar{Y}(\hat{\delta})] = \text{FP}(\hat{\delta})\text{Var}[\bar{Y}(\delta)]\eta(\delta)\text{Var}[\bar{Y}]$; ver (2.40) e (2.41). Loh [78] estendeu o trabalho desenvolvido por Lavenberg, Moeller e

Welch a esquemas de variáveis de controlo não lineares. O autor concluiu que, sob determinadas condições de regularidade pouco exigentes, a melhor forma de combinar variáveis de controlo com a resposta, considerando todas as combinações lineares e não lineares possíveis, está dentro da classe das relações lineares. Para um número de pontos experimentais n , suficientemente grande, qualquer esquema de controlo não linear é equivalente a um esquema linear, do ponto de vista de redução de variância [42]. Como resultado, neste trabalho, a relação entre a resposta e as diversas variáveis de controlo é sempre linear.

O trabalho de Lavenberg, Moeller e Welch é sobre a estimação da média da resposta para um valor fixo de cada variável de decisão (relevante no sistema em estudo). No contexto da construção de um modelo linear de regressão da resposta expressa em termos de um conjunto apropriado de variáveis explicativas (regressores), isto limita os dados de simulação como sendo uma constante escalar fixa. Com o objectivo de ultrapassar esta limitação, Nozari, Arnold e Pegden [92] utilizam uma experiência de simulação para estimar um modelo linear geral, em que a resposta depende de diversas variáveis de decisão que descrevem o sistema em estudo. Os autores consideram, como exemplo de aplicação, um problema de tráfego de automóveis. As medidas de eficiência que utilizam não são as que foram definidas por Lavenberg, Moeller e Welch, sendo análogas às usadas em [99].

Em Nozari, Arnold e Pegden [92] considera-se $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ um vector n -dimensional de observações independentes e normalmente distribuídas, com variância comum σ_Y^2 , onde $Y_i, i = 1, \dots, n$ é obtido a partir da i -ésima execução do programa de simulação. O valor esperado de \mathbf{Y} , $\mathcal{E}[\mathbf{Y}]$, assume-se ser $\mathbf{X}\boldsymbol{\theta}$, onde \mathbf{X} é uma matriz conhecida $n \times m$ com *rank* m , e $\boldsymbol{\theta}$ é um vector m -dimensional de parâmetros desconhecidos. Mais especificamente,

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\theta}, \sigma_Y^2 \mathbf{I}_n), \quad (2.43)$$

onde \mathbf{I}_n denota a matriz identidade n dimensional.

Com a i -ésima observação Y_i , é observado um vector aleatório de dimensão q , $\mathbf{C}_i = (C_{i1}, \dots, C_{iq})^T$, cujo valor esperado é conhecido. Sem perda de generalidade,

assumimos que $\mathcal{E}[\mathbf{C}_i] = \mathbf{0}$. Além disso, assume-se que

$$\mathbf{Z}_i = \begin{bmatrix} Y_i \\ \mathbf{C}_i \end{bmatrix} \sim N_{q+1} \left(\begin{bmatrix} \mu_i \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_Y^2 & \boldsymbol{\sigma}_{YC} \\ \boldsymbol{\sigma}_{CY} & \boldsymbol{\Sigma}_C \end{bmatrix} \right) \quad (2.44)$$

onde μ_i é o valor esperado de Y_i , i.e., $(\mu_1, \dots, \mu_n)^T = \mathbf{X}\boldsymbol{\theta}$. Note-se que, por hipótese, todos os \mathbf{C}_i têm a mesma matriz de variância-covariância e que $\boldsymbol{\sigma}_{YC}$, o vector linha cujos elementos são as covariâncias entre Y_i e os elementos de \mathbf{C}_i , não depende de i . Se considerarmos $\tilde{\mathbf{C}} = (\mathbf{C}_1^T, \dots, \mathbf{C}_n^T)^T$, então iremos ter

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \tilde{\mathbf{C}} \end{bmatrix} \sim N_{n(q+1)}(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z), \quad (2.45)$$

onde

$$\boldsymbol{\Sigma}_Z = \begin{bmatrix} \sigma_Y^2 & & 0 & & \boldsymbol{\sigma}_{YC} & & 0 \\ & \ddots & & & & \ddots & \\ 0 & & \sigma_Y^2 & & 0 & & \boldsymbol{\sigma}_{YC} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \boldsymbol{\sigma}_{CY} & & 0 & & \boldsymbol{\Sigma}_C & & 0 \\ & \ddots & & & & \ddots & \\ 0 & & \boldsymbol{\sigma}_{CY} & & 0 & & \boldsymbol{\Sigma}_C \end{bmatrix}$$

ou equivalentemente

$$\boldsymbol{\Sigma}_Z = \begin{bmatrix} \sigma_Y^2 \mathbf{I}_n & \boldsymbol{\sigma}_{YC} \otimes \mathbf{I}_n \\ \boldsymbol{\sigma}_{CY} \otimes \mathbf{I}_n & \boldsymbol{\Sigma}_C \otimes \mathbf{I}_n \end{bmatrix}$$

Sendo ou não $\boldsymbol{\Sigma}_Z$ conhecida, são apresentados estimadores pontuais e intervalos de confiança para $\boldsymbol{\theta}$ e são dados procedimentos para testar a hipótese linear geral

$$H_0 : \mathbf{A}\boldsymbol{\theta} = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0},$$

onde \mathbf{A} é uma matriz conhecida $(m - k) \times m$ com $\text{rank } m - k$. São apresentadas várias medidas de eficiência, para averiguar a eficácia da utilização das variáveis de controlo. Além disso, são feitas considerações sobre a influência do factor de perda na selecção das variáveis de controlo, recorrendo a alguns exemplos ilustrativos.

• **Caso em que Σ_Z é conhecida** Nesta situação, o estimador não enviesado com variância mínima de θ é

$$\hat{\theta}(\delta) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \left[\mathbf{Y} - (\sigma_{YC} \Sigma_C^{-1} \otimes \mathbf{I}_n) \tilde{\mathbf{C}} \right]. \quad (2.46)$$

e tem a seguinte distribuição:

$$\hat{\theta}(\delta) \sim N_m(\theta, \tau^2 (\mathbf{X}^T \mathbf{X})^{-1}), \quad (2.47)$$

onde $\tau^2 = \sigma_Y^2 - \sigma_{YC} \Sigma_C^{-1} \sigma_{CY}$. Para estudar os efeitos das variáveis de controlo, observamos que, quando $\tilde{\mathbf{C}}$ é ignorado, o nosso modelo linear será a relação (2.43), que é um modelo linear vulgar com σ_Y^2 conhecida. O estimador não enviesado com variância mínima de $\hat{\theta}$ é, neste caso,

$$\hat{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (2.48)$$

e além disso

$$\hat{\theta} \sim N_m(\theta, \sigma_Y^2 (\mathbf{X}^T \mathbf{X})^{-1}). \quad (2.49)$$

A medida de eficiência, utilizada pelos autores, é a matriz (comparar as relações (2.47) e (2.49))

$$\mathcal{D}[\hat{\theta}] - \mathcal{D}[\hat{\theta}(\delta)] = (\sigma_Y^2 - \tau^2) (\mathbf{X}^T \mathbf{X})^{-1}$$

que é sempre uma matriz definida positiva. Por conseguinte, quando a matriz Σ_Z é conhecida, é sempre melhor utilizar todas as variáveis de controlo disponíveis. Para este cenário, também é dada uma fórmula para os intervalos de confiança simultâneos de Scheffé.

• **Caso em que Σ_Z é desconhecida** Neste caso, considera-se a distribuição condicional

$$Y_i | \mathbf{C}_i. \sim N(\mu_i + \sigma_{YC} \Sigma_C^{-1} \mathbf{C}_i., \tau^2).$$

Se se definir a matriz de controlo

$$\mathbf{C} = \begin{bmatrix} C_{11} & \dots & C_{1q} \\ \vdots & & \vdots \\ C_{n1} & \dots & C_{nq} \end{bmatrix},$$

então, mostra-se que, o estimador controlado para θ é dado por

$$\hat{\theta}(\hat{\delta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{I}_n - \mathbf{C}(\mathbf{C}^T \mathbf{P} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{P}] \mathbf{Y},$$

onde $\mathbf{P} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Para discutir os efeitos das variáveis de controlo prova-se que, sob as condições da relação (2.45) e se $n - m - q - 1 > 0$:

$$\mathcal{D}[\hat{\theta}(\hat{\delta})] = \frac{n - m - 1}{n - m - q - 1} \tau^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (2.50)$$

A partir deste resultado, observa-se que q não pode ser maior que $n - m - 1$, porque nesse caso as variâncias dos elementos de $\hat{\theta}(\hat{\delta})$ não são finitos. Por conseguinte, os autores limitam a sua discussão à situação em que $q < n - m - 1$. O resultado (2.50) é um caso particular do Teorema 1 de Lavenberg, Moeller e Welch [69]; observe-se que, se considerarmos $m = 1$ e \mathbf{X} um vector coluna n -dimensional, então todos os seus elementos serão iguais a 1 e obtém-se o resultado do teorema referido. Para avaliar o efeito das variáveis de controlo, a medida de eficiência utilizada é, neste caso:

$$\mathcal{D}[\hat{\theta}] - \mathcal{D}[\hat{\theta}(\hat{\delta})] = \left(\sigma_Y^2 - \tau^2 \frac{n - m - 1}{n - m - q - 1} \right) (\mathbf{X}^T \mathbf{X})^{-1}.$$

É interessante observar, na medida de eficiência anterior, a existência de um factor de perda igual a

$$\frac{n - m - 1}{n - m - q - 1}.$$

Para finalizar, também é dada uma fórmula para os intervalos de confiança simultâneos.

Na perspectiva de utilizar variáveis de controlo para construir metamodelos não lineares de simulação, generalizou-se o trabalho de Nozari, Arnold e Pegden para metamodelos em que a relação entre a resposta e as variáveis de decisão é não linear, mas em que as variáveis de controlo e a resposta se relacionam linearmente. Este tipo de estudo é importante, porque existem diversas situações em que um metamodelo não linear é o que consegue aproximar correctamente a relação entre a resposta e as variáveis explicativas (ver a Secção 2.1).

Capítulo 3

Metodologia para metamodelação

Ao longo das últimas três décadas, tem sido muito frequente a construção e utilização de metamodelos de regressão linear, no contexto da simulação. Contudo, existem dois inconvenientes que se têm vindo a verificar e que são discutidos em seguida.

(i) De uma maneira geral, os metamodelos utilizados em simulação têm sido metamodelos de regressão linear polinomial. Acontece que, eles são pouco realistas e não têm a capacidade de produzir ajustes satisfatórios a formas de funções regulares arbitrárias. Além disso, frequentemente, a previsão destes metamodelos, fora da região experimental, não é satisfatória. Para ultrapassar esta dificuldade utilizam-se, neste trabalho, metamodelos que não apresentam estes inconvenientes e que, além disso, têm as vantagens dos metamodelos de regressão polinomial: metamodelos não lineares de regressão (ver a página 30 e a Secção 1.3).

(ii) Os metamodelos de simulação são frequentemente construídos de uma maneira *ad hoc*. Quando surge a necessidade de construir um metamodelo, para resolver um problema, numa determinada área, essa construção é levada a cabo com base em características muito específicas desse problema. Não há, portanto, a preocupação de reutilização do método de construção, para que este possa ser utilizado para resolver problemas

futuros. Acontece porém que, em geral, quando surge um novo problema, não se aproveita o trabalho e esforço dispendidos na resolução de problemas resolvidos anteriormente e faz-se novamente uma construção adaptada a um problema específico. Sendo assim, existe uma necessidade urgente em criar um processo sistemático que possa ser utilizado para resolver um conjunto amplo de problemas (mesmo que sejam de áreas diferentes). Por este motivo, propõe-se uma metodologia para construir metamodelos não lineares de regressão, no contexto da simulação.

Na Secção 3.1, apresenta-se a referida metodologia, propondo um procedimento detalhado para obtenção de resultados para construção de um metamodelo, selecção do tipo de metamodelo e estimação e validação do metamodelo seleccionado. A Secção 3.2 é reservada à apresentação dos estimadores dos parâmetros desconhecidos de um metamodelo e das distribuições a eles associadas. Também são apresentadas regiões de confiança e testes de hipóteses relativos aos parâmetros desconhecidos do metamodelo. Devido à sua extensão e importância, o problema da validação de metamodelos é tratado no Capítulo 4, onde é dado um procedimento detalhado e se fornecem ferramentas estatísticas apropriadas para validar metamodelos de regressão não linear em simulação.

Seguindo a terminologia de Kleijnen e Sargent [65], distinguem-se os seguintes elementos básicos (ver a Figura 3.1, transcrita de [65], onde a notação c.r.a. significa 'com respeito ao'):

- Um *problema real* é uma entidade, situação ou sistema (real ou proposto) que se pretende modelar;
- Um *modelo de simulação* é um programa de computador que descreve o problema real;
- Um *metamodelo* é um modelo, da forma (2.2), que se propõe descrever, tanto o problema real, como o modelo de simulação – se a função f em (2.2) é não linear em, pelo menos, um dos parâmetros desconhecidos $\theta_1, \dots, \theta_m$, então o metamodelo diz-se ser não linear.

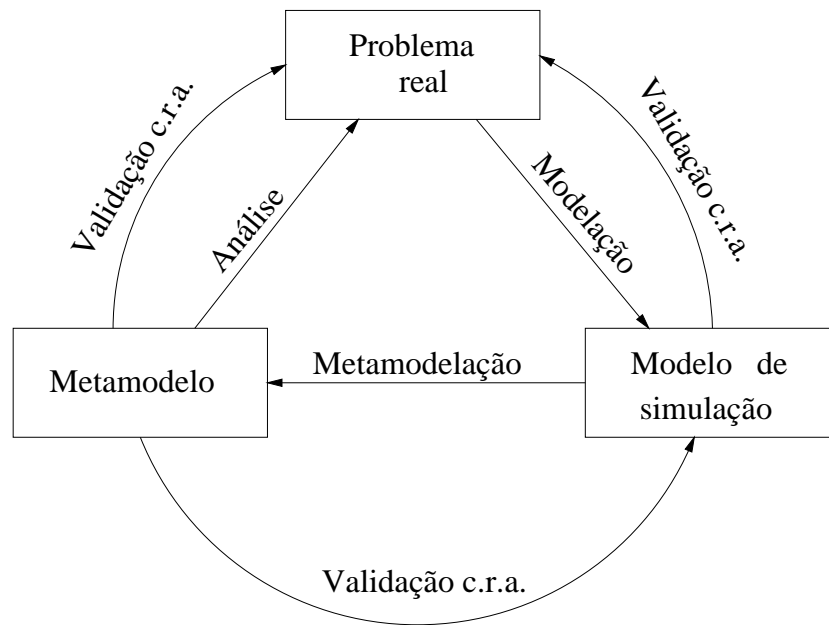


Figura 3.1: Metamodelo, modelo de simulação e problema real.

As setas interiores, na Figura 3.1, descrevem os processos que relacionam os elementos entre si e as outras referem-se a procedimentos que avaliam a credibilidade desses processos. Antes de descrever a terminologia utilizada nesta figura, para avaliação da credibilidade dos processos de análise, modelação e metamodelação, é necessário definir os dois conceitos que se seguem.

(i) *Validação de A, com respeito a B*: Comprovação que *A*, dentro do seu *domínio de aplicabilidade, com respeito a B*, está suficientemente próximo de *B*, tendo em conta o objectivo geral de aplicação de *A*.

(ii) *Domínio de aplicabilidade de A, com respeito a B*: Condições aconselhadas para as quais *A* foi testado, até onde foi possível, e é considerado apropriado para utilização como substituto de *B*. No exemplo tratado na Secção 6.2 (sistema de filas de espera $M/M/1$), foi testado um metamodelo hiperbólico no domínio de aplicabilidade $0 \leq \rho < 1$ e foi considerado apropriado para utilização como substituto do modelo de simulação e, também, do problema real.

Com base nestas definições, já se podem descrever as seguintes expressões:

- Quando A é o modelo de simulação e B é o problema real, então diz-se "Validação do modelo de simulação com respeito ao problema real".
- Quando A é o metamodelo e B é o problema real, então diz-se "Validação do metamodelo com respeito ao problema real".
- Quando A é o metamodelo e B é o modelo de simulação, então diz-se "Validação do metamodelo com respeito ao modelo de simulação".

O estudo do problema real distingue, em geral, diversas respostas do sistema, que poderemos ter interesse em analisar. Por conseguinte, o modelo de simulação correspondente também tem múltiplas variáveis de resposta. Contudo, na prática, em vez de utilizar um metamodelo com resposta multivariada, é mais frequente usar diversos metamodelos, cada um a descrever uma das variáveis de resposta que se pretendem analisar. Deste modo, a metodologia aqui apresentada considera apenas metamodelos com uma única variável de resposta.

Antes de se começar a descrever o processo de construção de metamodelos não lineares, convém salientar que, por hipótese, os modelos de simulação com que se vai trabalhar já foram todos devidamente verificados e validados. Considera-se que o modelo de simulação descreve convenientemente o problema real e que está disponível o programa de computador correspondente, para efeitos de experimentação. Neste contexto, o modelo de simulação é uma caixa preta válida, à qual fornecemos dados – variáveis e parâmetros de entrada – e observamos os resultados correspondentes – variáveis de resposta (ou simplesmente, respostas) – necessários à construção do metamodelo correspondente (ver a Figura 3.2).

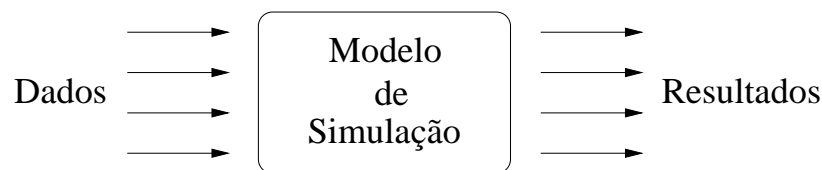


Figura 3.2: Modelo de Simulação.

Uma variável de entrada é uma quantidade que pode ser directamente observada, no problema real. Exemplos de variáveis são os instantes de chegada de clientes ao sistema, numa fila de espera $M/M/1$. Um parâmetro de entrada é uma quantidade que não pode ser directamente observada,

no problema real. Exemplo de um parâmetro é a taxa de chegadas ao sistema, na mesma fila de espera $M/M/1$.

3.1 Processo de metamodelação

A metodologia, proposta neste trabalho, é retratada na Figura 3.3 e consiste em determinados passos, que são discriminados em seguida.

1. Formulação do problema

Neste passo, é preciso formular o problema, determinando qual é o objectivo geral de aplicação do metamodelo e escolhendo as variáveis importantes para atingir esse objectivo.

Um objectivo do metamodelo pode ser, por exemplo, obter informação com respeito ao problema real. Neste caso, é comum fazer análise de sensibilidade e responder a questões do tipo 'O que acontece se ...?'. A análise de sensibilidade permite obter informação sobre a sensibilidade da variável de resposta a pequenas variações das variáveis de entrada. As questões do tipo 'O que acontece se ...?' têm interesse para saber qual é o valor da resposta para um determinado conjunto específico de variáveis de entrada. Outros objectivos possíveis podem ser: utilizar o metamodelo rotineiramente para prever valores da variável de resposta, ou fazer optimização.

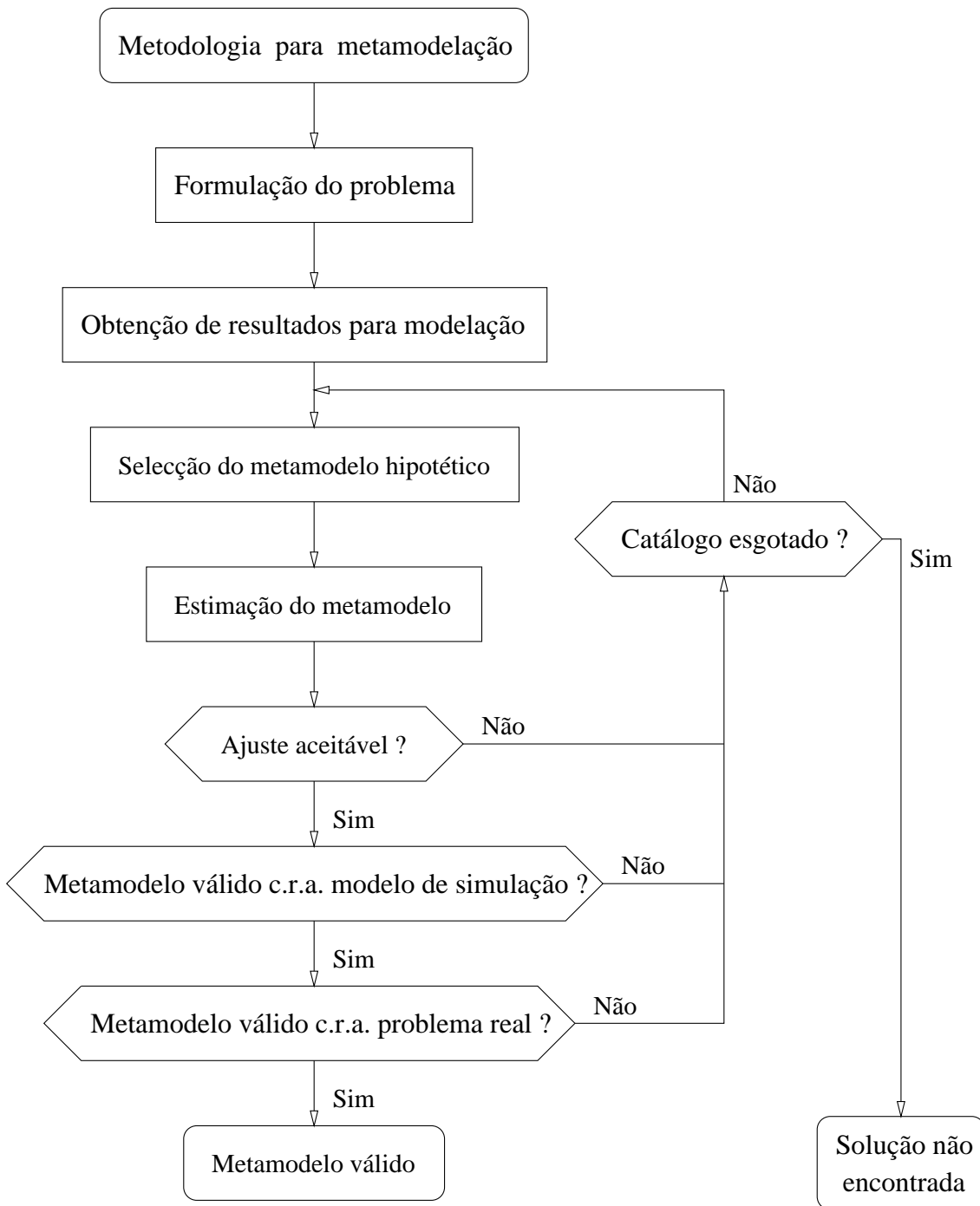


Figura 3.3: Processo de metamodelação.

Na escolha de variáveis, é preciso identificar a variável dependente e as variáveis independentes do metamodelo, bem como a região experimental. A variável dependente coincide com a variável de resposta, seleccionada, do modelo de simulação. Por exemplo, no modelo de filas de espera $M/M/1$, a variável dependente pode ser a média do tempo de espera na fila. As variáveis independentes do metamodelo podem ser idênticas às variáveis ou parâmetros de entrada do modelo de simulação, ou então transformações destes. Por exemplo, no caso do modelo de filas de espera $M/M/1$, uma variável independente do metamodelo pode ser o factor de utilização $\rho = \lambda/\mu$, em vez da taxa de chegadas λ , ou do tempo médio de serviço $1/\mu$ (ver o Capítulo 6). A região experimental é aquela onde as variáveis independentes do metamodelo podem assumir valores. Por exemplo, no caso do modelo de filas de espera $M/M/1$, a região experimental é $0 \leq \rho < 1$.

2. Obtenção de resultados para metamodelação

Em simulação estocástica, é preciso que, para cada execução do programa de simulação, sejam completamente especificadas as condições iniciais. Em geral, estas condições iniciais representam estados do sistema extremamente raros como, por exemplo, num sistema de filas de espera $M/M/1$, começar o programa de simulação sem nenhum cliente à espera de ser atendido. Como consequência, o programa de simulação começa com uma sequência de acontecimentos que podem ter uma probabilidade baixa de ocorrer. Isto faz com que os resultados obtidos, com base nessa experiência de simulação, possam estar contaminados com um viés inicial.

O viés inicial pode ser a maior fonte de erro na estimação dos parâmetros de um metamodelo; o método usual, para controlar este viés, é permitir que a execução do programa de simulação seja suficientemente longa, para que os resultados possam ser recolhidos fora do período em que o sistema ainda não está em equilíbrio estatístico. Isto é equivalente a deitar fora uma porção de resultados iniciais. Este procedimento é referido como truncatura de resultados e a observação a partir da qual os resultados são recolhidos é denominada ponto de truncatura. Existe um compromisso envolvido na selecção do ponto de truncatura. Se for deitada fora uma porção pequena de resultados, então o viés inicial, que resta, pode ainda afectar significativamente os resultados. Por outro lado, ignorar demasiados resultados aumenta a sua variância e não é económico (cada execução do programa de simulação pode ser bastante demorada, o que é especialmente importante no contexto do método das réplicas independentes).

O procedimento para obtenção de resultados para modelação, que em seguida é descrito em quatro passos, é um processo iterativo e recorre ao procedimento de Welch para eliminar o problema do viés inicial [118] (procedimento fortemente recomendado por Crawford e Gallwey [27]).

1º passo

Escolher n pontos experimentais (combinações de variáveis e parâmetros de entrada do modelo de simulação). Estes pontos experimentais têm de ser diferentes uns dos outros e têm de corresponder a valores das variáveis independentes do metamodelo pertencentes à região experimental. Para cada um destes pontos experimentais, executar r longas réplicas piloto, cada uma com k observações (k deverá ser bastante superior ao valor que se suspeita ser o ponto de truncatura). Na escolha do número de réplicas é preciso ter em conta que, por vezes, se utiliza o método dos mínimos quadrados ponderados estimados na estimação do metamodelo; neste caso, é conveniente considerar um valor de r que permita obter estimativas aceitáveis da variância da resposta, em cada ponto experimental. Com este objectivo, aconselha-se a escolher $r > 9$ [94], por exemplo $r = 10$.

2º passo

Em seguida, para cada ponto experimental, calcular a sequência de médias

$$\bar{Z}_{.l} = \frac{1}{r} \sum_{j=1}^r Z_{jl}, \quad l = 1, \dots, k,$$

ao longo das réplicas, onde Z_{jl} é a l -ésima observação da j -ésima réplica, $j = 1, \dots, r$, e determinar e representar graficamente as médias móveis

$$\bar{Z}_l(w) = \frac{1}{2w+1} \sum_{s=-w}^w \bar{Z}_{.l+s}, \quad \text{para } l = w+1, \dots, k-w,$$

$$\bar{Z}_l(w) = \frac{1}{2l-1} \sum_{s=-(l-1)}^{l-1} \bar{Z}_{.l+s}, \quad \text{para } l = 1, \dots, w,$$

para diferentes valores da janela de Welch, w , com $w \leq \lfloor k/2 \rfloor$ (ver a Figura 3.4). Seleccionar o valor mais pequeno de w que corresponde a uma representação gráfica suficientemente regular e escolher o valor de l , l_0 , a partir do momento em que a sequência $\{\bar{Z}_l(w) : l = 1, \dots, k-w\}$ parece ter convergido, ou seja, escolher o ponto de truncatura l_0 .

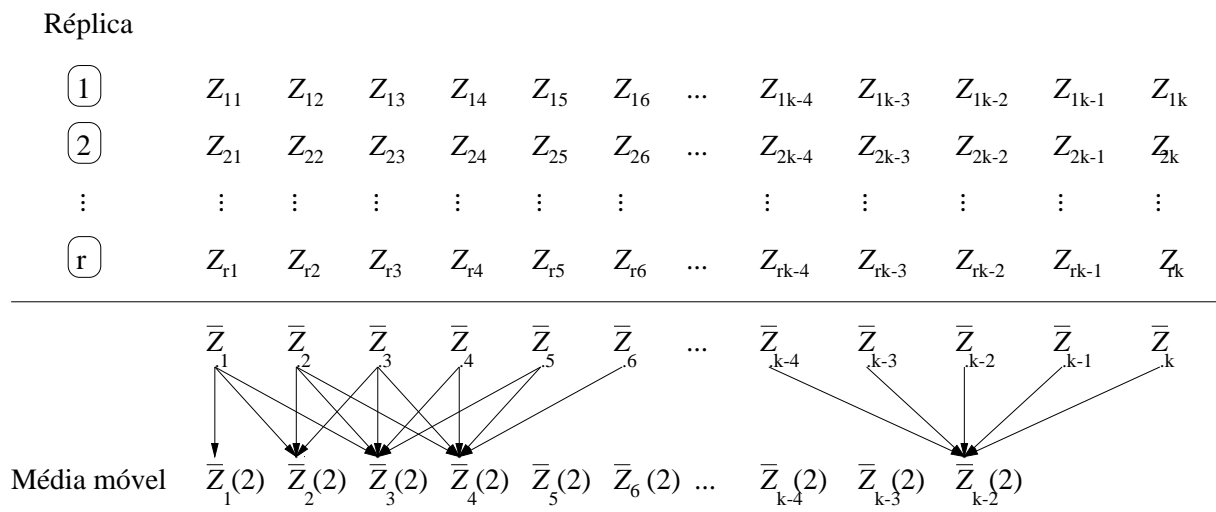


Figura 3.4: Cálculo da sequência de médias móveis com $w = 2$ e baseado em r réplicas.

Por exemplo, na situação descrita na Secção 6.1 (o problema do centro de inspeção e reparação de automóveis), para o ponto experimental 3, obtiveram-se os resultados presentes na Figura 3.5. Se o valor de w for pequeno, por exemplo $w = 10$, então o gráfico das médias móveis tem oscilações com frequência elevada, que é o que se pretende evitar ao definir o conceito de média móvel. Se o valor de w for menor que 10 (valor mais pequeno que os que estão na figura), então a frequência das oscilações ainda é mais elevada que para $w = 10$. Para $w = 50$, o gráfico já apresenta um comportamento suficientemente regular, no sentido em que a frequência das oscilações é pequena. A partir deste valor, o comportamento das oscilações não melhora significativamente, como pode, por exemplo, ser observado para $w = 100$. É preciso ter atenção para que w não seja demasiado elevado porque, neste caso, as observações \bar{Z}_l ficam demasiadamente agregadas e não se consegue ter uma boa ideia da curva transiente. Sendo assim, a janela escolhida é $w = 50$. Observando o gráfico correspondente a este valor de w , escolheu-se $l_0 = 100$, uma vez que é possível confinar a sequência $\{\bar{Z}_l(w) : l = l_0, \dots, k - w\}$ numa faixa horizontal bastante estreita.

3º passo

Se existir algum ponto experimental, para o qual não foi possível encontrar um valor para w satisfatório, então considerar $r = r + 10$ e repetir o 2º passo. Caso contrário, para cada réplica, do ponto experimental i , ignorar as observações até ao ponto de truncatura correspondente e

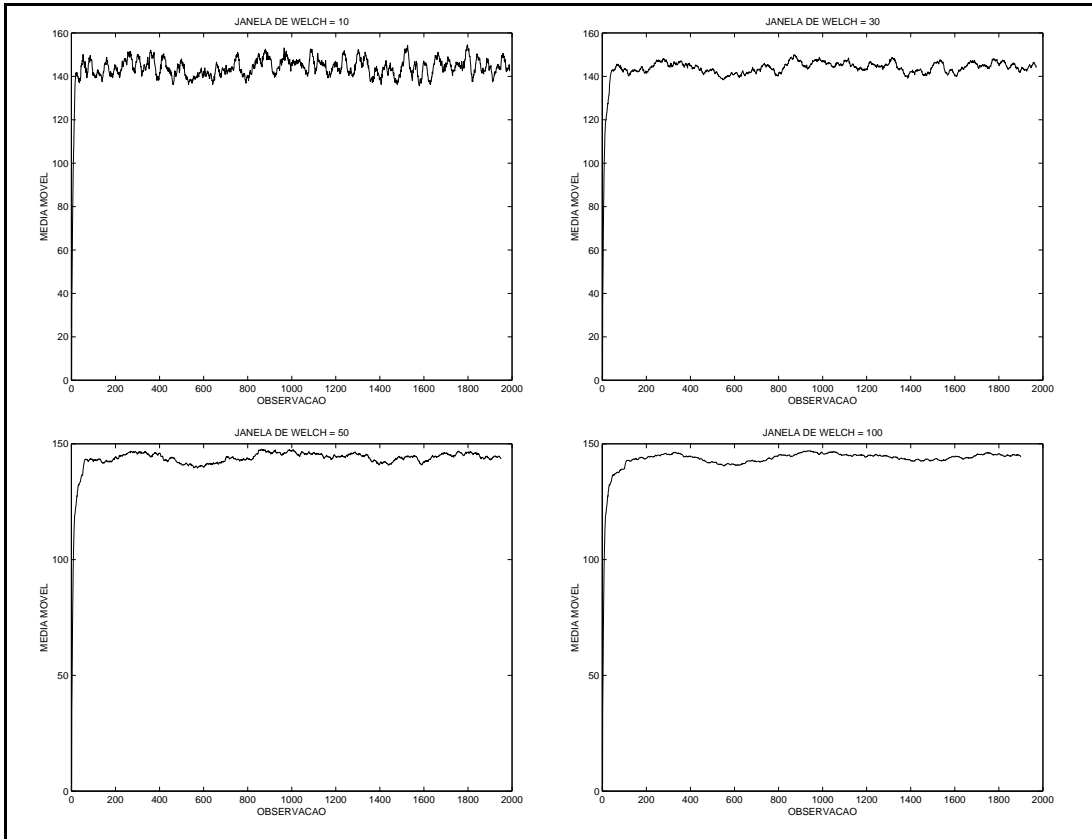


Figura 3.5: Médias móveis para vários valores da janela de Welch: $w = 10, 30, 50, 100$.

recolher os resultados de simulação

$$\{(\mathbf{X}_i, Y_{ij}, \hat{\sigma}_i) : i = 1, \dots, n, j = 1, \dots, r\}$$

(ver a Tabela 3.1). Y_{ij} é a resposta relevante do sistema correspondente ao ponto experimental i e

Tabela 3.1: Disposição dos resultados para estimação de metamodelos.

Ponto experimental	Variáveis independentes	Variável dependente	Desvio padrão estimado
1	$X_{11} \dots X_{1d}$	$Y_{11} \dots Y_{1r}$	$\hat{\sigma}_1$
\vdots	\vdots	\vdots	\vdots
i	$X_{i1} \dots X_{id}$	$Y_{i1} \dots Y_{ir}$	$\hat{\sigma}_i$
\vdots	\vdots	\vdots	\vdots
n	$X_{n1} \dots X_{nd}$	$Y_{n1} \dots Y_{nr}$	$\hat{\sigma}_n$

réplica j e

$$\hat{\sigma}_i = \left[\frac{1}{r-1} \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2 \right]^{1/2}, \quad (3.1)$$

com

$$\bar{Y}_i = \frac{1}{r} \sum_{j=1}^r Y_{ij}, \quad (3.2)$$

é a estimativa do desvio padrão, da resposta, no ponto experimental i (ver [64, Capítulo 8]).

4º passo

Calcular as distâncias entre os pontos consecutivos \bar{Y}_i , ou seja, $|\bar{Y}_{i+1} - \bar{Y}_i|$, onde $i = 1, \dots, n$. Verificar se $\epsilon_1 < |\bar{Y}_{i+1} - \bar{Y}_i| < \epsilon_2$, onde ϵ_1 e ϵ_2 são dois valores escolhidos. Se $|\bar{Y}_{i+1} - \bar{Y}_i| \leq \epsilon_1$, então não é preciso considerar os pontos $\{(\mathbf{X}_{i+1}, Y_{i+1,j}, \hat{\sigma}_{i+1}) : j = 1, \dots, r\}$ para efeitos de construção do metamodelo. Neste caso, a variável de resposta apresenta pouca variação entre os pontos \mathbf{X}_i e \mathbf{X}_{i+1} , e, por conseguinte, os pontos $\{(\mathbf{X}_i, Y_{ij}, \hat{\sigma}_i) : j = 1, \dots, r\}$ são geralmente suficientes para explicar o comportamento da resposta nesta região. Se $|\bar{Y}_{i+1} - \bar{Y}_i| \geq \epsilon_2$, então a variável de resposta varia significativamente entre os pontos \mathbf{X}_i e \mathbf{X}_{i+1} . Sendo assim, para

que seja possível perceber o verdadeiro comportamento da variável de resposta, é aconselhável escolher um ou mais pontos experimentais que correspondam a pontos entre \mathbf{X}_i e \mathbf{X}_{i+1} . Repetir os 2º, 3º, e 4º passos, considerando esses pontos experimentais e o número de réplicas igual ao utilizado na recolha dos resultados, correspondentes à primeira escolha de pontos experimentais.

3. Selecção do tipo de metamodelo

Idealmente, a selecção do metamodelo hipotético, para efeitos de estimação, é baseada em justificações físicas. Por exemplo, prova-se teoricamente que, numa amostra de gás mantida a uma temperatura constante, o volume v e a pressão p do gás satisfazem a relação $pv^\gamma = c$. Considerando a variável dependente $y = p$ e a variável independente $x = 1/v$, tem-se a relação não linear $y = cx^\gamma$, onde os parâmetros desconhecidos do metamodelo são c e γ . Contudo, em geral, não é conhecida uma equação teórica, como neste exemplo. Nesse caso, aconselha-se a escolher o metamodelo hipotético visualmente, seguindo uma perspectiva semelhante à utilizada na estatística, quando se comparam histogramas empíricos com funções de densidade de probabilidade conhecidas, para seleccionar uma função de distribuição.

Seguindo esta perspectiva, é conveniente em primeiro lugar construir o diagrama de dispersão da variável dependente versus cada variável independente. Ou seja, representar graficamente, para cada valor fixo de l ($l = 1, \dots, d$), os pares (X_{il}, Y_{ij}) onde $i = 1, \dots, n$ e $j = 1, \dots, r$. Em seguida, compara-se cada representação gráfica com diversas curvas analíticas, de um catálogo de curvas, e selecciona-se a curva que parece aproximar melhor os resultados contidos nessa representação. No Apêndice A, apresenta-se um catálogo de curvas, que inclui as curvas não lineares mais comuns presentes na literatura. A representação gráfica dessas curvas não é difícil de obter, uma vez que os computadores recentes fazem diferentes tipos de gráficos e todos os gráficos são relativamente simples de desenhar. Uma pequena parte desse catálogo é apresentada na Figura 3.6.

4. Estimação do metamodelo

A estimação de um metamodelo consiste em utilizar técnicas estatísticas para estimar os parâmetros desconhecidos do metamodelo hipotético (seleccionado no passo 3). Esta estimação é realizada utilizando um método muito conhecido, e largamente utilizado em estatística, para estimação de

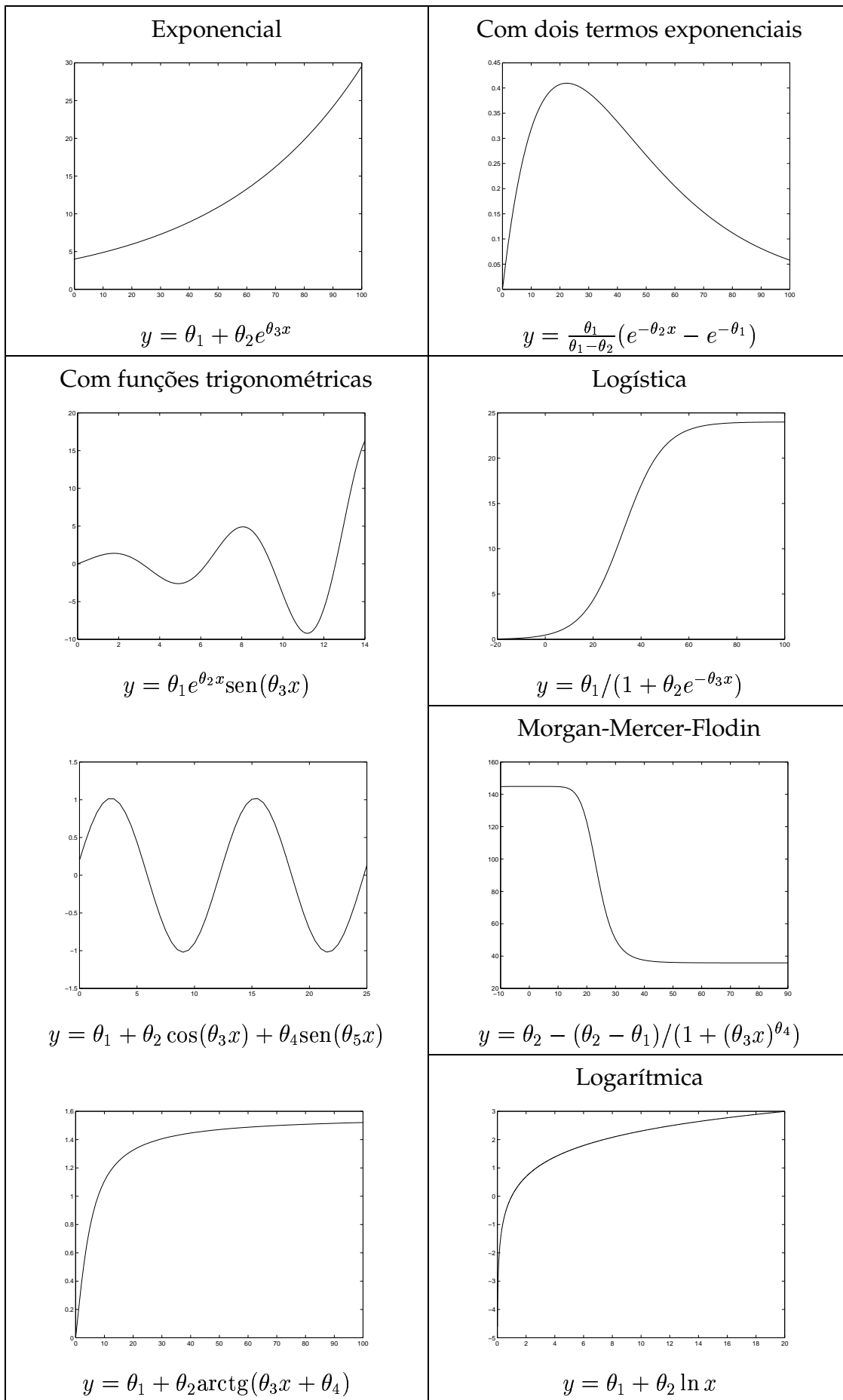


Figura 3.6: Algumas relações funcionais não lineares.

parâmetros desconhecidos em modelos: o método dos mínimos quadrados. Este método consiste em determinar um mínimo global para a soma dos quadrados dos erros (ou resíduos)

$$\text{SSE}(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^r [Y_{ij} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})]^2,$$

ou seja, obter um ponto $\hat{\boldsymbol{\theta}}$ tal que $\text{SSE}(\hat{\boldsymbol{\theta}}) < \text{SSE}(\boldsymbol{\theta})$, para todo o $\boldsymbol{\theta}$ pertencente a \mathbb{R}^m ($\boldsymbol{\theta} \neq \hat{\boldsymbol{\theta}}$). Para a maior parte dos metamodelos não lineares, a função $\text{SSE}(\boldsymbol{\theta})$ não pode ser minimizada analiticamente e é preciso utilizar um método numérico iterativo para encontrar $\hat{\boldsymbol{\theta}}$. Neste trabalho, usa-se o método de Levenberg-Marquardt porque é unânime a opinião que, na maior parte dos casos, este método é robusto e funciona bem na resolução de problemas de mínimos quadrados [107].

É de salientar que, ao contrário do caso dos mínimos quadrados lineares, $\text{SSE}(\boldsymbol{\theta})$ pode ter diversos mínimos locais para além do mínimo absoluto $\hat{\boldsymbol{\theta}}$. Sendo assim, na maior parte das situações, o melhor que se pode esperar é que o método numérico convirja para um mínimo local. Para que $\hat{\boldsymbol{\theta}}$ seja um mínimo local é suficiente que:

(i) as derivadas parciais de $\text{SSE}(\boldsymbol{\theta})$ com respeito a $\theta_1, \dots, \theta_m$, calculadas no ponto $\hat{\boldsymbol{\theta}}$, sejam nulas, isto é,

$$\frac{\partial \text{SSE}(\hat{\boldsymbol{\theta}})}{\partial \theta_k} = 0, \quad k = 1, \dots, m;$$

(ii) a matriz Hessiana de $\text{SSE}(\boldsymbol{\theta})$, calculada no ponto $\hat{\boldsymbol{\theta}}$, seja definida positiva, isto é,

$$\sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \text{SSE}}{\partial \theta_i \partial \theta_j} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \theta_i \theta_j > 0$$

para todo o $\boldsymbol{\theta}$ diferente de zero e pertencente a \mathbb{R}^m .

Na obtenção de resultados descrita no passo 2, considerou-se o mesmo número de réplicas em cada ponto experimental (num estudo de simulação isto não põe qualquer tipo de problema). Por conseguinte, é possível substituir as observações individuais Y_{ij} pelas suas médias \bar{Y}_i . e, depois, aplicar o critério dos mínimos quadrados ao vector das n médias \bar{Y}_i . Desta forma simplifica-se bastante o procedimento computacional de estimação.

Após a obtenção do estimador dos mínimos quadrados, podem-se determinar as regiões de confiança, para o verdadeiro vector de parâmetros do metamodelo, e realizar testes de hipóteses.

Procedimentos para construir estas regiões e efectuar estes testes de hipóteses são apresentados na Secção 3.2. Na mesma secção, sob determinadas condições de regularidade, apresenta-se o estimador dos mínimos quadrados não lineares e a sua distribuição estatística. Todo este desenvolvimento é levado a cabo considerando as duas situações seguintes: (i) variância dos erros constante; e (ii) variância dos erros heterogénea.

5. Validação do metamodelo estimado

Para garantir que o metamodelo estimado, no passo 4, aproxima suficientemente bem, tanto os resultados obtidos no passo 2 (adequação do metamodelo), como o modelo de simulação (validade do metamodelo com respeito ao modelo de simulação) e, também, o problema real (validade do metamodelo com respeito ao problema real), têm de ser seguidos determinados procedimentos de validação. Estes procedimentos de validação são devidamente apresentados no Capítulo 4. Será preciso retomar o passo 3, se o metamodelo não for considerado: (i) adequado; (ii) válido com respeito ao modelo de simulação; e (iii) válido com respeito ao problema real. Caso contrário, o metamodelo é considerado válido e pode ser utilizado para o fim desejado.

Pode acontecer que seja impossível encontrar um metamodelo válido para determinado problema. Por conseguinte, o passo 3 só é retomado se existir pelo menos uma curva, no catálogo de curvas, que ainda não tenha sido seleccionada. Caso contrário, termina-se o processo de modelação e conclui-se que, neste caso, a metodologia não é aplicável.

3.2 Estimação de metamodelos

Suponha-se um desenho experimental, que consiste em n pontos experimentais $\{X_{il} : i = 1, \dots, n; l = 1, \dots, d\}$. Para cada ponto experimental, realizam-se r réplicas independentes do modelo de simulação e a experiência produz $\{(\mathbf{X}_{i.}, Y_{ij}, \hat{\sigma}_i) : i = 1, \dots, n; j = 1, \dots, r\}$. Considere-se também o seguinte metamodelo não linear com réplicas

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \epsilon_{ij}, \quad (3.3)$$

para $i = 1, \dots, n$ e $j = 1, \dots, r$, onde as variáveis aleatórias ϵ_{ij} são independentes e identicamente distribuídas com distribuição $N(0, \sigma_i^2)$, com $\sigma_i > 0$.

O modelo de simulação (2.1) define uma população estatística de observações, Y . As características destas observações dependem das variáveis de decisão (parâmetros e variáveis de entrada). Os membros da população correspondem a todas as possíveis sementes de números pseudo-aleatórios, isto é, em teoria temos uma população de dimensão infinitamente grande. A hipótese clássica é que esta população tem uma distribuição gaussiana. Se o metamodelo (3.3) for válido, então a esperança e a variância condicionais da população são dadas por

$$E[Y_{ij}|\mathbf{X}_{i.}] = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) = \mu_i,$$

$$\text{Var}[Y_{ij}|\mathbf{X}_{i.}] = \sigma_i^2.$$

Por conseguinte, o desenho experimental implica que os resultados da experiência de simulação Y_{ij} sejam interpretados como r observações independentes da distribuição normal $N(\mu_i, \sigma_i^2)$. Daqui resulta que faz sentido considerar o metamodelo (3.3).

A estimação é levada a cabo utilizando o método dos mínimos quadrados não lineares e consideram-se duas situações distintas:

(i) variância dos erros constante, isto é, $\sigma_i^2 = \sigma^2 > 0$ ($i = 1 \dots, n$); e

(ii) variância dos erros heterogénea (situação mais realista), isto é, σ_i^2 varia com i , $i = 1 \dots, n$.

Neste trabalho, a heterogeneidade da variância é medida através da quantidade

$$het = \frac{\max_{i=1,n} \hat{\sigma}_i}{\min_{i=1,n} \hat{\sigma}_i} \quad (3.4)$$

onde $\hat{\sigma}_i$ é dado por (3.1) [64].

3.2.1 Estimador dos mínimos quadrados

(i) Variância dos erros constante.

Proposição 3.1 Dados $\epsilon_{ij} \sim N(0, \sigma^2)$, com $\sigma^2 > 0$ em (3.3), e condições de regularidade apropriadas [107, páginas 563-572]. Então, para $N = nr$ suficientemente grande, o estimador dos mínimos quadrados dos coeficientes do metamodelo de simulação $\boldsymbol{\theta}$ em (3.3) satisfaz aproximadamente:

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{F}^T(\boldsymbol{\theta}^*)\mathbf{F}(\boldsymbol{\theta}^*)]^{-1}\mathbf{F}^T(\boldsymbol{\theta}^*)[\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*)], \quad (3.5)$$

$$\hat{\boldsymbol{\theta}} \sim N_m \left(\boldsymbol{\theta}^*, \frac{\sigma^2}{r} [\mathbf{F}^T(\boldsymbol{\theta}^*)\mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \right), \quad (3.6)$$

$$(N - m) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{N-m}^2, \quad (3.7)$$

$$\hat{\sigma}^2 = \frac{1}{N - m} \sum_{i=1}^n \sum_{j=1}^r [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})]^2, \quad (3.8)$$

onde $\boldsymbol{\theta}^*$ é o valor exacto de $\boldsymbol{\theta}$, $\mathbf{f}(\boldsymbol{\theta}^*) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}^*), \dots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}^*))^T$, $\mathbf{F}(\boldsymbol{\theta}^*)$ é a matriz Jacobiana de \mathbf{f} , calculada em $\boldsymbol{\theta}^*$ e $\bar{\mathbf{Y}} = (\bar{Y}_{1.}, \dots, \bar{Y}_{n.})^T$.

Verificação Em [107, Secção 2.1.5] observa-se que, encontrar o estimador dos mínimos quadrados $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$, no problema (3.3), é equivalente a minimizar o somatório

$$r \sum_{i=1}^n [\bar{Y}_{i.} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})]^2,$$

com respeito a $\boldsymbol{\theta}$, ou seja,

$$\sum_{i=1}^n [\bar{Y}_{i.} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})]^2.$$

Esta é a soma dos quadrados dos erros correspondente ao problema

$$\bar{Y}_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \bar{\epsilon}_i, \quad i = 1, 2, \dots, n. \quad (3.9)$$

Uma vez que os erros são variáveis aleatórias independentes e identicamente distribuídas com distribuição $\epsilon_{ij} \sim N(0, \sigma^2)$, então

$$\text{Var}[\bar{\epsilon}_i] = \frac{1}{r^2} \sum_{j=1}^r \text{Var}[\epsilon_{ij}] = \frac{\sigma^2}{r}, \quad (3.10)$$

$$\text{E}[\bar{\epsilon}_i] = \frac{1}{r} \sum_{j=1}^r \text{E}[\epsilon_{ij}] = 0,$$

e, conseqüentemente, $\bar{\epsilon}_i$ ($i = 1, \dots, n$) são variáveis aleatórias independentes e identicamente distribuídas com distribuição

$$\bar{\epsilon}_i \sim N(0, \sigma^2/r). \quad (3.11)$$

Uma vez que σ^2 em (3.6) é desconhecida, estima-se esta quantidade através de (3.8). Note-se que o estimador (3.8) pode ser escrito na forma

$$\hat{\sigma}^2 = \frac{\text{SSE}(\hat{\boldsymbol{\theta}})}{\mathbf{N} - \mathbf{m}} \quad (3.12)$$

com $\text{SSE}(\hat{\boldsymbol{\theta}}) = \|\mathbf{Y} - \tilde{\mathbf{f}}(\hat{\boldsymbol{\theta}})\|_2^2$, onde $\mathbf{Y} = (Y_{11}, \dots, Y_{1r}, \dots, Y_{n1}, \dots, Y_{nr})^T$ e

$$\tilde{\mathbf{f}}(\hat{\boldsymbol{\theta}}) = (f(\mathbf{X}_{1,1}, \hat{\boldsymbol{\theta}}), \dots, f(\mathbf{X}_{1,r}, \hat{\boldsymbol{\theta}}), \dots, f(\mathbf{X}_{n,1}, \hat{\boldsymbol{\theta}}), \dots, f(\mathbf{X}_{n,r}, \hat{\boldsymbol{\theta}}))^T.$$

Utilizando o resultado (2.18) de [107, página 24] obtém-se

$$\text{SSE}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\epsilon}^T (\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}}) \boldsymbol{\epsilon} \quad (3.13)$$

onde $\boldsymbol{\epsilon} = (\epsilon_{11}, \dots, \epsilon_{1r}, \dots, \epsilon_{n1}, \dots, \epsilon_{nr})^T$ e $\mathbf{P}_{\tilde{\mathbf{F}}}$ é uma matriz projecção $\mathbf{P}_{\tilde{\mathbf{F}}} = \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T$, sendo $\tilde{\mathbf{F}}$ a matriz Jacobiana de $\tilde{\mathbf{f}}$ calculada no ponto $\hat{\boldsymbol{\theta}}$.

Uma vez que a matriz $\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}}$ é simétrica (isto é, $\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}} = (\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}})^T$), idempotente (isto é, $(\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}})^2 = \mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}}$) e $\epsilon_{ij} \sim N(0, \sigma^2)$, $i = 1, \dots, n$, $j = 1, \dots, r$, então aplica-se o Teorema B.4 de [109, página 291] a (3.13) e pode-se afirmar que $\sigma^{-2} \text{SSE}(\hat{\boldsymbol{\theta}})$ tem distribuição qui-quadrado com $\text{tr}(\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}})$ graus de liberdade (a notação tr representa o traço da matriz [109, página 271]).

Mas

$$\begin{aligned}
tr(\mathbf{I}_N - \mathbf{P}_{\tilde{\mathbf{F}}}) &= tr(\mathbf{I}_N) - tr(\mathbf{P}_{\tilde{\mathbf{F}}}) = \\
&= N - tr[\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T] = \\
&= N - tr[\tilde{\mathbf{F}}^T \tilde{\mathbf{F}} (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1}] = \\
&= N - tr(\mathbf{I}_m) = \\
&= N - m,
\end{aligned}$$

ou seja, $\sigma^{-2}SSE(\hat{\boldsymbol{\theta}}) \sim \chi_{N-m}^2$. Além disso $SSE(\hat{\boldsymbol{\theta}}) = (N - m)\hat{\sigma}^2$ (ver (3.12)), logo verifica-se o resultado (3.7).

Finalmente, aplicando o resultado (12.21) de [107, página 570] ao problema (3.9), obtém-se

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{F}^T(\boldsymbol{\theta}^*)\mathbf{F}(\boldsymbol{\theta}^*)]^{-1}\mathbf{F}^T(\boldsymbol{\theta}^*)[\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*)].$$

(3.6) obtém-se aplicando (12.23) de [107, página 571] ao problema (3.9); ver também [107, Teorema 2.1(i)]. Também [107, páginas 564-572] enumera algumas hipóteses de regularidade que asseguram a eficiência assintótica de $\hat{\boldsymbol{\theta}}$ e consistência forte de $\hat{\boldsymbol{\theta}}$ e $\hat{\sigma}$. \diamond

(ii) Variância dos erros heterogénea.

Na maior parte das situações reais, a hipótese de homogeneidade das variâncias dos erros é irrealista; um caso muito comum é, por exemplo, quando as variâncias dos resultados crescem com a grandeza dos mesmos (ver os exemplos do Capítulo 6). Quando isto acontece, é preferível utilizar o método dos mínimos quadrados ponderados. Observe-se que, neste caso, o problema a estudar é (3.9):

$$\bar{Y}_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \bar{\epsilon}_i, \quad i = 1, 2, \dots, n,$$

com

$$\text{Var}[\bar{Y}_i] = \frac{\sigma_i^2}{r}, \quad (3.14)$$

e

$$\bar{\epsilon}_i \sim N\left(0, \frac{\sigma_i^2}{r}\right).$$

Portanto,

$$\bar{\boldsymbol{\epsilon}} \sim N_n\left(\mathbf{0}, \frac{1}{r}\boldsymbol{\Sigma}\right), \quad (3.15)$$

onde

$$\Sigma = \text{diag}[\sigma_1^2, \dots, \sigma_n^2]. \quad (3.16)$$

Proposição 3.2 *Suponha-se que $\epsilon_{ij} \sim N(0, \sigma_i^2)$, com $\sigma_i > 0$ em (3.3), e condições de regularidade apropriadas [119]. Então, para $N = nr$ suficientemente grande, tem-se aproximadamente:*

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{F}^T(\boldsymbol{\theta}^*)\Sigma^{-1}\mathbf{F}(\boldsymbol{\theta}^*)]^{-1}\mathbf{F}^T(\boldsymbol{\theta}^*)\Sigma^{-1}[\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*)], \quad (3.17)$$

$$\hat{\boldsymbol{\theta}} \sim N_m \left(\boldsymbol{\theta}^*, \frac{1}{r}[\mathbf{F}^T(\boldsymbol{\theta}^*)\Sigma^{-1}\mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \right). \quad (3.18)$$

Verificação *De acordo com [107, Secção 2.1.4], o problema (3.9), com uma distribuição de erros (3.15), é um problema de mínimos quadrados ponderados, onde a função a ser minimizada é*

$$[\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]^T \left(\frac{1}{r}\Sigma \right)^{-1} [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})].$$

Minimizar esta função, com respeito a $\boldsymbol{\theta}$, é equivalente a minimizar

$$[\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})]^T \Sigma^{-1} [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})].$$

Σ é uma matriz simétrica e definida positiva que admite a decomposição de Cholesky:

$$\Sigma = \mathbf{U}^T \mathbf{U}, \quad (3.19)$$

onde \mathbf{U} é uma matriz triangular superior. Multiplicando ambos os membros do modelo não linear (3.9) por $\mathbf{R} = (\mathbf{U}^T)^{-1}$, obtém-se

$$\mathbf{W} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\eta}, \quad (3.20)$$

onde $\mathbf{W} = \mathbf{R}\bar{\mathbf{Y}}$, $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$ e $\boldsymbol{\eta} = \mathbf{R}\bar{\boldsymbol{\epsilon}}$. Então, observa-se que

$$\begin{aligned} \mathcal{E}[\boldsymbol{\eta}] &= \mathbf{0}, \\ \mathcal{D}[\boldsymbol{\eta}] &= \mathbf{R}\mathcal{D}[\bar{\boldsymbol{\epsilon}}]\mathbf{R}^T = \frac{1}{r}\mathbf{R}\Sigma\mathbf{R}^T. \end{aligned}$$

Mas Σ admite a decomposição (3.19) e $\mathbf{R} = (\mathbf{U}^T)^{-1}$, logo

$$\mathcal{D}[\boldsymbol{\eta}] = \frac{1}{r}(\mathbf{U}^T)^{-1}\mathbf{U}^T\mathbf{U}[(\mathbf{U}^T)^{-1}]^T = \frac{1}{r}\mathbf{I}_n.$$

Sendo assim, o problema (3.9) foi transformado no problema (3.20), problema este que possui variância dos erros constante e esperança nula. Por conseguinte, pode-se aplicar o método dos mínimos quadrados usuais ao problema (3.20). O estimador dos mínimos quadrados ponderados de $\boldsymbol{\theta}$ é dado por [107, Teorema 2.1]

$$\dot{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{G}^T(\boldsymbol{\theta}^*)\mathbf{G}(\boldsymbol{\theta}^*)]^{-1}\mathbf{G}^T(\boldsymbol{\theta}^*)[\mathbf{W} - \mathbf{g}(\boldsymbol{\theta}^*)], \quad (3.21)$$

onde

$$\mathbf{G} = \frac{\partial \mathbf{g}(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$$

é a matriz Jacobiana de \mathbf{g} .

Mas $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$, então

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{R} \frac{\partial \mathbf{f}(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \mathbf{R}\mathbf{F}(\boldsymbol{\theta}).$$

Além disso, $\mathbf{W} = \mathbf{R}\bar{\mathbf{Y}}$ e $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, logo (3.21) é equivalente a:

$$\begin{aligned} \dot{\boldsymbol{\theta}} &\approx \boldsymbol{\theta}^* + [\mathbf{F}^T\mathbf{R}^T\mathbf{R}\mathbf{F}]^{-1}(\mathbf{R}\mathbf{F})^T[\mathbf{R}\bar{\mathbf{Y}} - \mathbf{R}\mathbf{f}(\boldsymbol{\theta}^*)] \\ &= \boldsymbol{\theta}^* + [\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}]^{-1}\mathbf{F}^T\mathbf{R}^T\mathbf{R}[\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*)] \\ &= \boldsymbol{\theta}^* + [\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}]^{-1}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}[\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*)], \end{aligned}$$

onde, para simplificar a notação, se utiliza \mathbf{f} e \mathbf{F} para representar $\mathbf{f}(\boldsymbol{\theta}^*)$ e $\mathbf{F}(\boldsymbol{\theta}^*)$ respectivamente. Fica assim estabelecido (3.17).

O resultado (3.18) obtém-se aplicando [107, Teorema 2.1(i)] ao problema (3.21):

$$\dot{\boldsymbol{\theta}} \sim N_m \left[\boldsymbol{\theta}, \frac{1}{r}(\mathbf{G}^T\mathbf{G})^{-1} \right].$$

Uma vez que $\mathbf{G} = \mathbf{R}\mathbf{F}$ e $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, obtém-se

$$\dot{\boldsymbol{\theta}} \sim N_m \left[\boldsymbol{\theta}, \frac{1}{r}(\mathbf{F}^T\mathbf{R}^T\mathbf{R}\mathbf{F})^{-1} \right],$$

e depois

$$\dot{\boldsymbol{\theta}} \sim N_m \left[\boldsymbol{\theta}, \frac{1}{r}(\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1} \right] \diamond.$$

Sempre que seja necessário, deve-se substituir $\mathbf{F}(\boldsymbol{\theta}^*)$ por $\hat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$. Se $\boldsymbol{\Sigma}$ for conhecida, então o estimador $\hat{\boldsymbol{\theta}}$ tem propriedades conhecidas. Contudo, nos estudos de simulação, $\boldsymbol{\Sigma}$ é estimada por (2.7), o que conduz ao estimador não linear dos mínimos quadrados ponderados estimados

$$\tilde{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + [\mathbf{F}^T(\boldsymbol{\theta}^*)\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}(\boldsymbol{\theta}^*)]^{-1}\mathbf{F}^T(\boldsymbol{\theta}^*)\hat{\boldsymbol{\Sigma}}^{-1}[\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*)].$$

Tanto no caso linear como no não linear, o estimador dos mínimos quadrados ponderados estimados não é fácil de analisar. No contexto da regressão linear, existem condições técnicas sob as quais os estimadores, dos mínimos quadrados ponderados e dos mínimos quadrados ponderados estimados, têm o mesmo comportamento assintótico [61]. Se existir uma parametrização, para a qual o metamodelo não linear possa ser aproximado por um metamodelo linear, então no caso não linear também se pode considerar que os dois referidos estimadores têm o mesmo comportamento assintótico, se determinadas condições técnicas forem satisfeitas.

3.2.2 Regiões de confiança

Em seguida, são propostos dois tipos de intervalos de confiança aproximados: intervalos de Scheffé e intervalos de Bonferroni. Os intervalos de Scheffé tendem a ser muito compridos e portanto a sua utilidade é limitada. Enquanto que os intervalos de Bonferroni são utilizados para combinações lineares pré-escolhidas, os de Scheffé usam-se para fazer julgamentos sobre combinações lineares sugeridas pelo conjunto de resultados. Contudo, os intervalos de Bonferroni são habitualmente os apropriados para metamodelos não lineares [107, Secção 5.2.1].

A. Intervalos de Bonferroni

(i) Variância dos erros constante.

Com o objectivo de obter um rectângulo de confiança para $\boldsymbol{\theta}^*$, formado por intervalos de confiança, um para cada θ_i^* ($i = 1, \dots, m$), procede-se como se segue. Começa-se por construir um intervalo de confiança, para uma combinação linear $\mathbf{a}^T \boldsymbol{\theta}$ ($\mathbf{a} \in \mathbb{R}^m$), aplicando a Proposição 3.1. Em particular, temos o resultado assintótico (3.6)

$$\hat{\boldsymbol{\theta}} \sim N_m \left(\boldsymbol{\theta}^*, \frac{\sigma^2}{r} [\mathbf{F}^T(\boldsymbol{\theta}^*) \mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \right). \quad (3.22)$$

A partir deste resultado, é possível afirmar que assintoticamente

$$\mathbf{a}^T \hat{\boldsymbol{\theta}} \sim N \left(\mathbf{a}^T \boldsymbol{\theta}^*, \frac{\sigma^2}{r} \mathbf{a}^T [\mathbf{F}^T(\boldsymbol{\theta}^*) \mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \mathbf{a} \right). \quad (3.23)$$

Além disso, (3.8) é um estimador não enviesado da variância da população e estatisticamente independente de $\hat{\boldsymbol{\theta}}$. Por conseguinte, $\hat{\sigma}^2$ é não enviesado e independente de $\mathbf{a}^T \hat{\boldsymbol{\theta}}$. Como resultado, para N relativamente grande, tem-se aproximadamente,

$$\frac{\mathbf{a}^T \hat{\boldsymbol{\theta}} - \mathbf{a}^T \boldsymbol{\theta}^*}{\hat{\sigma} \{r^{-1} \mathbf{a}^T [\mathbf{F}^T(\boldsymbol{\theta}^*) \mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \mathbf{a}\}^{1/2}} \sim t_{N-m},$$

onde t_{N-m} é a distribuição t com $N - m$ graus de liberdade. Um intervalo de confiança aproximado, com probabilidade de cobertura igual a $1 - \alpha$, para $\mathbf{a}^T \boldsymbol{\theta}^*$ é então

$$\mathbf{a}^T \hat{\boldsymbol{\theta}} \pm t_{N-m; \alpha/2} \hat{\sigma} \{r^{-1} \mathbf{a}^T [\mathbf{F}^T(\boldsymbol{\theta}^*) \mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \mathbf{a}\}^{1/2}. \quad (3.24)$$

No intervalo de confiança anterior, pode-se estimar $\mathbf{F}(\boldsymbol{\theta}^*)$ através de $\hat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$. Considerando $\mathbf{a}^T = (0, \dots, 0, 1, 0, \dots, 0)$, onde o k -ésimo elemento de \mathbf{a} é igual a um e os restantes elementos iguais a zero, então um intervalo de confiança para o k -ésimo elemento de $\boldsymbol{\theta}^*$, θ_k^* , é

$$\hat{\theta}_k \pm t_{N-m; \alpha/2} \frac{\hat{\sigma}}{r^{1/2}} \left[\hat{\mathbf{F}}^T \hat{\mathbf{F}} \right]_{kk}^{-1/2}.$$

O rectângulo de confiança pretendido, com nível de confiança aproximado de pelo menos $100(1 - \alpha)\%$, obtém-se utilizando o método de Bonferroni:

$$\hat{\theta}_k \pm t_{N-m; \alpha/(2m)} \frac{\hat{\sigma}}{r^{1/2}} \left[\hat{\mathbf{F}}^T \hat{\mathbf{F}} \right]_{kk}^{-1/2}, \quad k = 1, 2, \dots, m. \quad (3.25)$$

(ii) Variância dos erros heterogénea

Quando a variância dos erros não é constante ao longo dos pontos experimentais, propõe-se o seguinte rectângulo de confiança de Bonferroni para $\boldsymbol{\theta}^*$:

$$\tilde{\theta}_k \pm ct_{r-n+m-1; \alpha/(2m)} \left[r \hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{F}} \right]_{kk}^{-1/2}, \quad k = 1, 2, \dots, m, \quad (3.26)$$

onde $\hat{\mathbf{F}} = \mathbf{F}(\tilde{\boldsymbol{\theta}})$, $\hat{\boldsymbol{\Sigma}}$ é da forma (2.7) e a constante c é calculada através de

$$c = \left[\frac{1 + F_{Rao}(n-m)/(r-n+m)}{1 - (n-m)/(r-1)} \right]^{1/2},$$

onde F_{Rao} é a estatística associada ao teste de Rao (ver a página 90).

Este rectângulo obteve-se adaptando o resultado (3.13) de Kleijnen [61] a metamodelos não lineares (neste caso, $\widehat{\text{Var}}[\hat{\theta}_k] = [r \hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{F}}]_{kk}^{-1}$) e aplicando, em seguida, o método de Bonferroni.

B. Intervalos de Scheffé

(i) Variância dos erros constante.

A região de confiança assintótica, para $\boldsymbol{\theta}^*$, com probabilidade de cobertura igual a $1 - \alpha$, obtém-se aplicando [107, (2.26), página 25] ao problema (3.9) e tem a forma:

$$\left\{ \boldsymbol{\theta}^* : \frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \hat{\mathbf{F}}^T \hat{\mathbf{F}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)}{m \hat{\sigma}^2} \leq F_{m, N-m; \alpha} \right\}. \quad (3.27)$$

Com o objectivo de apresentar os intervalos simultâneos de Scheffé para θ_k^* , com $k = 1, \dots, m$, começa-se por aplicar [107, (5.11), página 194] ao problema (3.9) e obtém-se o seguinte intervalo de confiança válido para qualquer combinação linear $\mathbf{a}^T \theta^*$, onde \mathbf{a} é um vector real tal que $\mathbf{a} \neq \mathbf{0}$:

$$\mathbf{a}^T \hat{\theta} \pm (m F_{m, N-m; \alpha})^{1/2} \hat{\sigma} \left(\mathbf{a}^T \left[\hat{\mathbf{F}}^T \hat{\mathbf{F}} \right]^{-1} \mathbf{a} \right)^{1/2}. \quad (3.28)$$

(ii) Variância dos erros heterogénea.

Sob determinadas condições técnicas, o estimador dos mínimos quadrados ponderados estimados satisfaz assintoticamente

$$\tilde{\theta} \sim N_m \left(\theta^*, \frac{1}{r} [\hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}}]^{-1} \right),$$

onde $\hat{\mathbf{F}} = \mathbf{F}(\hat{\theta})$ e $\hat{\Sigma}$ é dado por (2.7). Consequentemente, aplicando [3, Teorema 3.3.3, página 54], vem

$$(\tilde{\theta} - \theta^*)^T r \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} (\tilde{\theta} - \theta^*) \sim \chi_m^2.$$

Sendo assim, a região de confiança aproximada, com probabilidade de cobertura igual a $1 - \alpha$, para θ^* , é a seguinte:

$$\left\{ \theta^* : (\tilde{\theta} - \theta^*)^T r \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} (\tilde{\theta} - \theta^*) \leq \chi_{m; \alpha}^2 \right\}. \quad (3.29)$$

Para exibir os intervalos de confiança simultâneos, para qualquer combinação linear $\mathbf{a}^T \theta^*$, onde $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}$, aplica-se a teoria linear usual para o método-S de Scheffé [106, Capítulo 5]. Para começar, a desigualdade de Cauchy-Schwartz permite escrever [107, Apêndice A7.2]

$$(\tilde{\theta} - \theta^*)^T r \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} (\tilde{\theta} - \theta^*) = \sup_{\mathbf{a}: \mathbf{a} \neq \mathbf{0}} \frac{\left[(\tilde{\theta} - \theta^*)^T \mathbf{a} \right]^2}{\mathbf{a}^T \left(r \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} \right)^{-1} \mathbf{a}}$$

Utilizando a igualdade anterior e a região de confiança (3.29), vem aproximadamente

$$\left\{ \mathbf{a}^T \theta^* : \left| \mathbf{a}^T (\tilde{\theta} - \theta^*) \right| \leq \left(\chi_{m; \alpha}^2 \mathbf{a}^T \left[r \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} \right]^{-1} \mathbf{a} \right)^{1/2} \right\}.$$

Donde se obtém, para todo o \mathbf{a} real tal que $\mathbf{a} \neq \mathbf{0}$,

$$\mathbf{a}^T \tilde{\theta} \pm \left(\frac{1}{r} \chi_{m; \alpha}^2 \right)^{1/2} \left(\mathbf{a}^T \left[\hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} \right]^{-1} \mathbf{a} \right)^{1/2}.$$

3.2.3 Testes de hipóteses

(i) *Variância dos erros constante.*

Com o objectivo de testar a hipótese $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, pode-se utilizar (3.27) para considerar

$$F_1 = \frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \hat{\mathbf{F}}^T \hat{\mathbf{F}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}{m \hat{\sigma}^2},$$

onde $\hat{\sigma}^2$ é dado por (3.8). F_1 tem a distribuição aproximada $F_{m, N-m}$ quando H_0 é verdadeira. Portanto, rejeita-se H_0 , com nível de significância α , se $F_1 > F_{m, N-m; \alpha}$ (para testar outro tipo de hipóteses, lineares ou não lineares, consultar por exemplo [107]).

(ii) *Variância dos erros heterogénea.*

Para testar a hipótese $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, é possível utilizar (3.29) e calcular a estatística

$$\chi = r(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{F}} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \quad (3.30)$$

que tem distribuição aproximada χ_m^2 quando H_0 é verdadeira. Neste caso, rejeita-se H_0 , com nível de significância α , se $\chi > \chi_{m, \alpha}^2$. Para uma consulta mais aprofundada sobre vários tipos de testes, aconselha-se a consultar o trabalho desenvolvido por Halbert White [119].

Capítulo 4

Validação de metamodelos

É de extrema importância o papel da validação, no contexto da construção de metamodelos de simulação. Após a estimação do metamodelo, nada garante que ele é credível para ser aplicado como substituto, ou do problema real, ou do modelo de simulação. Por conseguinte, antes de um metamodelo ser utilizado em qualquer tipo de aplicação, é preciso garantir que esse metamodelo é válido. A validação de um metamodelo consiste em três passos, que são seguidamente discriminados.

(i) Verificação da adequação do metamodelo. Neste passo, em primeiro lugar, verifica-se se as hipóteses distribucionais, para a porção estocástica do metamodelo, são satisfeitas. Em segundo lugar, averigua-se até que ponto a porção determinística do metamodelo é adequada, no sentido estatístico. Finalmente, verifica-se se a capacidade previsiva do metamodelo é satisfatória.

(ii) Validação do metamodelo com respeito ao modelo de simulação. Neste passo, verifica-se se o metamodelo, dentro do seu domínio de aplicabilidade (com respeito ao modelo de simulação) está suficientemente próximo do modelo de simulação, tendo em conta o objectivo geral de aplicação do metamodelo. Esta verificação é feita utilizando um procedimento de partição de resultados.

(iii) Validação do metamodelo com respeito ao problema real. Neste passo, verifica-se se o metamodelo, dentro do seu domínio de aplicabilidade (com respeito ao problema real) está suficientemente próximo do problema real, tendo em conta o objectivo geral de aplicação do metamodelo.

Esta verificação pode ser levada a cabo recorrendo a procedimentos utilizados no contexto da validação de modelos de simulação. A proximidade do metamodelo, em relação ao problema real, depende de duas aproximações que podem somar-se ou cancelar-se entre elas: construção do modelo de simulação com base no problema real e construção do metamodelo com base no modelo de simulação (ver a Figura 3.1).

Se o objectivo geral, de aplicação de um metamodelo, for compreender a fundo o problema real, então é conveniente obter respostas do problema real para um conjunto numeroso de dados. Infelizmente, em geral, isto é impossível e o passo (iii) da validação do metamodelo é levado a cabo fazendo apenas algumas comparações entre as respostas do metamodelo e as respostas do problema real. Pode mesmo acontecer que o problema real não seja observável e então é impossível determinar se o metamodelo satisfaz algum critério de validação com respeito ao problema real. Nesse caso, apenas se considera a verificação da adequação do metamodelo e a validação do metamodelo com respeito ao modelo de simulação. Existe outra situação em que, geralmente, a validação também só é realizada em relação à adequação do metamodelo e à validação com respeito ao modelo de simulação: quando o objectivo geral de aplicação do metamodelo é ajudar na verificação e validação de um modelo de simulação.

Neste capítulo, apresentam-se procedimentos numéricos e estatísticos para verificar a adequação do metamodelo (Secção 4.1), avaliar a validade do metamodelo com respeito ao modelo de simulação (Secção 4.2) e avaliar a validade do metamodelo com respeito ao problema real (Secção 4.3). Alguns dos procedimentos referidos são conhecidos, o que é novo é a sua aplicação na construção de metamodelos de regressão não linear em simulação. Outros, embora conhecidos da teoria da regressão linear, tiveram de ser devidamente adaptados ao contexto dos metamodelos não lineares de simulação.

4.1 Adequação do metamodelo

Para obter informação sobre a potencial violação das hipóteses distribucionais, para a porção estocástica do metamodelo, são propostos, na Secção 4.1.1, métodos gráficos de análise de resíduos.

Na Secção 4.1.2, apresentam-se testes estatísticos que ajudam o experimentador a averiguar a adequação, no sentido estatístico, da porção determinística de um metamodelo de regressão não linear. Fora do contexto da simulação, Rao [97] propôs um teste F de falta de ajustamento. Um dos testes de falta de ajustamento, sugeridos nesta secção, é uma adaptação desse teste de Rao a metamodelos não lineares de simulação. Além deste teste, também são sugeridos um teste clássico de falta de ajustamento e o conhecido coeficiente de determinação, R^2 . Na Secção 4.1.3, com o objectivo de testar a capacidade previsiva do metamodelo, apresentam-se a estatística PRESS e um procedimento de validação cruzada. Também são apresentadas estatísticas de diagnóstico, relacionadas com a estatística utilizada no referido procedimento. Este método de validação cruzada é uma generalização, para metamodelos não lineares de regressão, de um método proposto por Kleijnen em 1983 [59]. Assim, como no caso linear [94], demonstra-se que, tanto no caso dos mínimos quadrados usuais, como no caso dos mínimos quadrados ponderados estimados, é possível realizar este procedimento de validação utilizando apenas uma única regressão.

4.1.1 Hipóteses distribucionais

As hipóteses distribucionais que se pretendem analisar, num metamodelo de regressão não linear, são as da variância constante e da normalidade dos erros. Com o objectivo de averiguar se estas hipóteses são satisfeitas, habitualmente, realiza-se uma análise gráfica de resíduos [107, Secção 4.6]. Embora este tipo de análise seja inerentemente subjectiva, consegue frequentemente dar mais informação sobre a potencial violação das hipóteses distribucionais que qualquer teste formal (ver [84] e [87]). Na interpretação das representações gráficas de resíduos, é conveniente lembrar que a soma dos resíduos, no caso não linear, pode não ser zero. Se a curvatura intrínseca for insignificante [107, página 131], como em geral é o caso, então é apropriado construir as representações gráficas de resíduos utilizadas, usualmente, para metamodelos lineares. Contudo, se a curvatura intrínseca for substancial, então podem ser utilizadas representações gráficas com resíduos projectados [107, Secção 4.6.4].

As representações gráficas de resíduos (relacionadas entre si), que habitualmente se utilizam, são as seguintes:

(i) gráficos de resíduos versus variáveis independentes;

(ii) gráfico de resíduos versus valores previstos; e

(iii) gráfico de probabilidade normal dos resíduos.

Nas duas primeiras representações gráficas, (i) e (ii), os resíduos deverão encontrar-se dentro de uma banda horizontal centrada em zero, não apresentando tendência sistemática para serem positivas ou negativas (ver, por exemplo, a Figura 6.6). Neste caso, pode-se assumir que a variância dos erros é constante. Na terceira representação gráfica, (iii), cada resíduo é representado graficamente versus o seu valor esperado quando a distribuição é normal. Uma representação gráfica aproximadamente linear sugere que a distribuição é normal; caso contrário, contradiz a hipótese que a distribuição dos erros é normal (ver, por exemplo, a Figura 6.7). Pode ser útil fazer o gráfico de probabilidade normal para cada ponto experimental. Por exemplo, no caso da distribuição ser normal, mas a variância não constante, cada gráfico é aproximadamente linear, mas a inclinação varia de ponto para ponto (ver, por exemplo, a Figura 6.5).

4.1.2 Falta de ajustamento

Após o metamodelo hipotético ter sido escolhido e os parâmetros desconhecidos desse metamodelo terem sido estimados, surge a necessidade de averiguar a bondade do ajustamento, do metamodelo estimado, aos resultados obtidos no passo 2 da metodologia.

Para ajudar a concluir se um determinado metamodelo, de regressão não linear, ajusta adequadamente os resultados obtidos no passo 2 da metodologia, apresentam-se em seguida três procedimentos formais.

A. Teste F Considere-se o metamodelo com réplicas (3.3), onde se assume que os erros são variáveis aleatórias independentes e identicamente distribuídas, tal que $\epsilon_{ij} \sim N(0, \sigma^2)$, com $i = 1, \dots, n$ e $j = 1, \dots, r$. Representando o metamodelo (3.3) na forma

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

tem-se a seguinte decomposição usual

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \mu_i)^2 &= \sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_i + \bar{Y}_i - \mu_i)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^n r (\bar{Y}_i - \mu_i)^2,\end{aligned}$$

onde

$$\sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2$$

é a soma dos quadrados dos erros puros. Se μ_i for substituído por $f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) = \hat{\mu}_i$, na decomposição anterior, então

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^r [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})]^2 &= \sum_{i=1}^n \sum_{j=1}^r [Y_{ij} - \bar{Y}_i + \bar{Y}_i - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})]^2 \\ &= \sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^n r [\bar{Y}_i - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})]^2,\end{aligned}$$

ou seja

$$\text{SSE} = \text{SSPE} + (\text{SSE} - \text{SSPE}).$$

O primeiro membro é a soma dos quadrados dos resíduos e é decomposto na soma dos quadrados dos erros puros, SSPE, e na soma dos quadrados de falta de ajustamento, SSE – SSPE. No caso linear, em que $f(\mathbf{X}_{i.}, \boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$, um **teste clássico** para a falta de ajustamento é

$$F = \frac{(\text{SSE} - \text{SSPE})/(\mathbf{n} - \mathbf{m})}{\text{SSPE}/(\mathbf{N} - \mathbf{n})}, \quad (4.1)$$

onde $F \sim F_{\mathbf{n}-\mathbf{m}, \mathbf{N}-\mathbf{n}}$, se o metamodelo for válido. Para metamodelos não lineares, uma vez que se verifica a linearidade assintótica, o teste F anterior é aproximadamente válido para \mathbf{N} grande [107, página 32].

Se os erros, em (3.3), forem variáveis aleatórias independentes tal que $\epsilon_{ij} \sim N(0, \sigma_i^2)$, $i = 1, 2, \dots, \mathbf{n}$ e $j = 1, 2, \dots, \mathbf{r}$, então o teste de falta de ajustamento é

$$F = \frac{(\text{SSE} - \text{SSPE})/(\mathbf{n} - \mathbf{m})}{\text{SSPE}/(\mathbf{N} - \mathbf{n})},$$

onde

$$\text{SSE} = \sum_{i=1}^n \sum_{j=1}^r \frac{1}{\sigma_i^2} [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})]^2$$

é a soma dos quadrados dos erros usual (ou, soma dos quadrados dos resíduos) e

$$SSPE = \sum_{i=1}^n \sum_{j=1}^r \frac{1}{\sigma_i^2} (Y_{ij} - \bar{Y}_{i.})^2 \quad (4.2)$$

[107, página 82]. Se existir uma parametrização, para a qual o metamodelo possa ser aproximado adequadamente por um metamodelo linear, então F terá aproximadamente uma distribuição $F_{n-m, N-n}$, quando o metamodelo é válido. O teste consiste em rejeitar o metamodelo se, para um dado nível de confiança α , $F > F_{n-m, N-n}^{1-\alpha}$. Quando σ_i ($i = 1, 2, \dots, n$) tem de ser estimado, como em geral é o caso, o número usual de graus de liberdade $N - n$ tem de ser reduzido em n ficando apenas $N - 2n$.

No caso em que a variância dos erros não é homogênea e no contexto dos metamodelos lineares de simulação, Kleijnen sugeriu utilizar o **teste F de Rao**. Neste trabalho, generalizou-se este teste para o caso não linear. Ele consiste em obter um valor positivo, F_{Rao} , tal que quanto mais pequeno ele for, melhor será a aproximação utilizando o metamodelo. Consequentemente, um ajuste ideal corresponderá a $F_{Rao} = 0$. Sendo assim, o teste de Rao permite ordenar diversos metamodelos por "grau de aproximação".

O teste de Rao tem as seguintes condições de aplicabilidade:

(i) $n > m = \text{rank}(\mathbf{F})$ e $r > n (> n - m)$ (para que $\hat{\Sigma}$, dada por (2.7), seja não singular); e

(ii) as respostas do modelo de simulação têm uma distribuição simétrica, como por exemplo, uma distribuição normal ou uniforme (deste modo, obtêm-se melhores resultados).

Dadas estas condições de aplicabilidade, o teste de Rao tem a forma

$$F_{Rao} = \frac{r(r - n + m)}{(n - m)(r - 1)} \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} \left[\bar{Y}_{i.} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) \right]^2. \quad (4.3)$$

Se existir uma parametrização para a qual o metamodelo possa ser aproximado adequadamente por um metamodelo linear então, quando o metamodelo é válido, F_{Rao} é aproximadamente distribuído como uma distribuição $F_{n-m, r-n+m}$.

B. Coeficiente de determinação Quando se tem um bom ajuste, no contexto da regressão linear ou não linear [109, Apêndice C.4], os resíduos obtidos são pequenos. Por conseguinte, a qualidade do ajuste pode ser medida através da soma dos quadrados dos resíduos SSE. Contudo, esta quantidade é dependente das unidades em que é expressa a resposta. Assim, uma medida mais adequada é o chamado coeficiente de determinação,

$$R^2 = 1 - \frac{SSE}{SST}.$$

onde SST é a soma dos quadrados total. No caso em que a variância dos erros é constante,

$$SST = \sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2$$

onde $\bar{Y}_{..}$ é a média geral de todas as observações

$$\bar{Y}_{..} = \frac{1}{nr} \sum_{i=1}^n \sum_{j=1}^r Y_{ij}.$$

Quando a variância dos erros é heterogênea, tem-se

$$SST = \sum_{i=1}^n \sum_{j=1}^r \left(\frac{Y_{ij}}{\sigma_i} - \bar{Y}_{..} \right)^2 \quad (4.4)$$

e

$$\bar{Y}_{..} = \frac{1}{nr} \sum_{i=1}^n \sum_{j=1}^r \frac{Y_{ij}}{\sigma_i} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i} \bar{Y}_i. \quad (4.5)$$

Este coeficiente mede a proporção de variação total da resposta, em torno da média da resposta, que a regressão consegue explicar. Quando se utiliza esta medida de falta de ajustamento, não se rejeita a hipótese de o metamodelo ser adequado sempre que R^2 esteja próximo de 1. O ajuste perfeito é quando todas as respostas são iguais aos seus valores previstos pelo metamodelo ($R^2 = 1$). Devido à sensibilidade de R^2 em relação à dimensão do vector θ , é preferível utilizar o coeficiente de determinação ajustado (para o número de parâmetros m)

$$R_{adj}^2 = 1 - (1 - R^2) \frac{N - 1}{N - m};$$

quando aumenta o número de parâmetros desconhecidos do metamodelo, o valor de SSE diminui sempre, enquanto que o valor de SST não é alterado e, por conseguinte, o valor de R^2 aumenta sempre (ver [87, página 241]).

4.1.3 Capacidade previsiva

Quando se ajusta um metamodelo, os resíduos usuais, $e_{ij} = Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})$, não são, em geral, indicativos de como é que o metamodelo irá prever novos resultados. Uma vez que $f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})$ não é independente de Y_{ij} e é com efeito desenhado para isso, estes resíduos são frequentemente mais pequenos que os verdadeiros erros de previsão. Por conseguinte, estes resíduos, assim como a soma $\sum_{i=1}^n \sum_{j=1}^r e_{ij}$, medem a qualidade do ajuste em vez de avaliar a qualidade do metamodelo como predictor. Sendo assim, os resíduos previstos $e_{ij(-i)} = Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{(-i)})$ são uma medida mais eficaz para averiguar a capacidade previsiva de um metamodelo ($\hat{\boldsymbol{\theta}}_{(-i)}$ é o estimador dos mínimos quadrados que se obtém utilizando todas as observações, excepto as correspondentes ao ponto experimental i). Esta motivação é a base dos dois métodos de validação que são propostos em seguida.

A. Estatística PRESS Um procedimento usual, para averiguar a capacidade previsiva de um modelo linear, é a estatística PRESS [87, página 450]. Esta estatística, adaptada ao problema (3.3), é dada por

$$PRESS = \sum_{i=1}^n \sum_{j=1}^r \frac{1}{\sigma_i^2} [Y_{ij} - f(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}}_{(-i)})]^2, \quad (4.6)$$

se a variância dos erros é heterogénea. Uma vez que $Y_{ij} - f(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}}_{(-i)})$ é o erro de previsão, quando o metamodelo é ajustado sem o i -ésimo ponto experimental e $f(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}}_{(-i)})$ é usado como predictor, um metamodelo é considerado bom predictor se o valor de PRESS for pequeno. Se os valores de PRESS e SSE forem bastante próximos, então MSE poderá ser um indicador válido da capacidade previsiva do metamodelo seleccionado.

Se a variância dos erros for constante, então

$$PRESS = \sum_{i=1}^n \sum_{j=1}^r [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{(-i)})]^2.$$

B. Um método de validação cruzada Comece-se por utilizar o metamodelo (3.9) para prever o resultado de uma nova combinação de dados, \mathbf{X}_0 , ou seja,

$$\hat{Y}_0 = f(\mathbf{X}_0, \hat{\boldsymbol{\theta}}). \quad (4.7)$$

Para averiguar se (4.7) é ou não um bom previsor (ou seja, se o metamodelo pode ser usado para prever valores), executa-se o programa de simulação com os novos dados \mathbf{X}_0 , fazendo r réplicas independentes; ou seja, utilizando para cada execução uma sequência de entrada de números pseudo-aleatórios diferente das restantes. O resultado para comparação com o valor previsto \hat{Y}_0 é

$$\bar{Y}_0 = \frac{1}{r} \sum_{j=1}^r Y_{0j}. \quad (4.8)$$

Aplicando a estatística t de Kleijnen (2.10), para comparar \hat{Y}_0 e \bar{Y}_0 , obtém-se

$$t_0 = \frac{\bar{Y}_0 - \hat{Y}_0}{\widehat{\text{Var}}[\bar{Y}_0 - \hat{Y}_0]}.$$

\hat{Y}_0 e \bar{Y}_0 são independentes porque $\hat{\boldsymbol{\theta}}$ não depende dos valores Y_{0j} , $j = 1, \dots, r$. Por conseguinte, pode-se escrever a estatística anterior na forma

$$t_0 = \frac{\bar{Y}_0 - \hat{Y}_0}{\widehat{\text{Var}}[\bar{Y}_0] + \widehat{\text{Var}}[\hat{Y}_0]}. \quad (4.9)$$

Em seguida, calcula-se o valor desta estatística t_0 . Se ele for significativo, ou seja $|t_0| > t_{r-1; \alpha/2}$ onde $t_{r-1; \alpha/2}$ é o valor crítico fixado, então não se utiliza o metamodelo para previsão.

Para se obter o valor t_0 observe-se como é que, na prática, se pode determinar o denominador de (4.9). Uma vez que \bar{Y}_0 é dado por (4.8), pode-se escrever:

$$\widehat{\text{Var}}[\bar{Y}_0] = \frac{1}{r^2} \sum_{k=1}^r \widehat{\text{Var}}[Y_{0k}]. \quad (4.10)$$

Utilizando (2.8) sabe-se que

$$\widehat{\text{Var}}[Y_{0k}] = \frac{1}{r-1} \sum_{j=1}^r (Y_{0j} - \bar{Y}_0)^2, \quad k = 1, \dots, r. \quad (4.11)$$

Substituindo (4.11) em (4.10) obtém-se

$$\widehat{\text{Var}}[\bar{Y}_0] = \frac{1}{r(r-1)} \sum_{j=1}^r (Y_{0j} - \bar{Y}_0)^2. \quad (4.12)$$

Para calcular $\widehat{\text{Var}}[\hat{Y}_0]$, procede-se do seguinte modo: uma vez que, para \mathbf{N} suficientemente grande, $\hat{\boldsymbol{\theta}}$ está próximo do verdadeiro valor $\boldsymbol{\theta}^*$, pode-se utilizar o desenvolvimento em série de Taylor de $f(\mathbf{X}_0, \hat{\boldsymbol{\theta}})$, em torno de $(\mathbf{X}_0, \boldsymbol{\theta}^*)$, e escrever [107, página 193]:

$$\hat{Y}_0 \approx f(\mathbf{X}_0, \boldsymbol{\theta}^*) + \mathbf{f}_0^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*),$$

onde $\mathbf{f}_0^T = \mathbf{f}_0^T(\boldsymbol{\theta}^*)$ é o vector gradiente de $f(\mathbf{X}_0, \boldsymbol{\theta}^*)$. Sendo assim, a variância da previsão \hat{Y}_0 é

$$\text{Var}[\hat{Y}_0] \approx \mathbf{f}_0^T \mathcal{D}[\hat{\boldsymbol{\theta}}] \mathbf{f}_0. \quad (4.13)$$

Convém referir que, afirmar que $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\theta}^*$ estão próximos significa que a distância entre eles, medida através da norma euclidiana, é bastante pequena, ou seja

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = \sqrt{\sum_{i=1}^m (\hat{\theta}_i - \theta_i^*)^2} < \epsilon,$$

sendo ϵ um valor pequeno.

Contudo, existem duas situações distintas:

(i) Variância dos erros constante. Neste caso, o resultado (4.13) resulta em (ver a proposição 3.1)

$$\text{Var}[\hat{Y}_0] \approx \frac{\sigma^2}{r} \mathbf{f}_0^T [\mathbf{F}^T \mathbf{F}]^{-1} \mathbf{f}_0.$$

Por conseguinte,

$$\widehat{\text{Var}}[\hat{Y}_0] \approx \frac{\hat{\sigma}^2}{r} \hat{\mathbf{f}}_0^T [\hat{\mathbf{F}}^T \hat{\mathbf{F}}]^{-1} \hat{\mathbf{f}}_0 \quad (4.14)$$

onde $\hat{\sigma}^2$ é dada na proposição 3.1, equação (3.8), $\hat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$ e $\hat{\mathbf{f}}_0^T = \mathbf{f}_0^T(\hat{\boldsymbol{\theta}})$. Observe-se que, uma vez que $\boldsymbol{\theta}^*$ é desconhecido, houve a necessidade de o aproximar para ser possível calcular uma estimativa aproximada da variância. O vector escolhido foi a melhor aproximação disponível. Ao longo deste trabalho, este tipo de aproximação vai ser feita sempre que haja necessidade.

(ii) Variância dos erros heterogénea. Nesta situação, aplica-se a proposição 3.2 e obtém-se a seguinte variância estimada:

$$\widehat{\text{Var}}[\hat{Y}_0] \approx \frac{1}{r} \hat{\mathbf{f}}_0^T [\hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{F}}]^{-1} \hat{\mathbf{f}}_0, \quad (4.15)$$

com $\hat{\mathbf{F}} = \mathbf{F}(\tilde{\boldsymbol{\theta}})$, $\hat{\mathbf{f}}_0^T = \mathbf{f}_0^T(\tilde{\boldsymbol{\theta}})$ e $\hat{\boldsymbol{\Sigma}}$ dada por (2.7).

Com base na exposição anterior, propõe-se um método de validação cruzada para avaliar a capacidade previsiva de metamodelos não lineares de simulação. Este método consiste em realizar: (i) os passos primeiro a quarto (descritos em seguida), para $i = 1, 2, \dots, n$; e depois o quinto passo (também descrito em seguida).

Primeiro passo

Considerar o conjunto que se obtém eliminando, do conjunto de resultados obtido no passo 2 da metodologia, a informação relativa ao i -ésimo ponto experimental

$$\{(\mathbf{X}_{i.}, Y_{ij}, \hat{\sigma}_i) : j = 1, \dots, r\}.$$

Segundo passo

Estimar o metamodelo de regressão, a partir do conjunto de resultados obtido no passo anterior. Se a variância dos erros for constante, então o estimador dos mínimos quadrados é dado por (ver a proposição 3.1)

$$\hat{\boldsymbol{\theta}}_{(-i)} \approx \boldsymbol{\theta}^* + [\mathbf{F}_{(-i)}^T(\boldsymbol{\theta}^*) \mathbf{F}_{(-i)}(\boldsymbol{\theta}^*)]^{-1} \mathbf{F}_{(-i)}^T(\boldsymbol{\theta}^*) [\bar{\mathbf{Y}}_{(-i)} - \mathbf{f}_{(-i)}(\boldsymbol{\theta}^*)].$$

Caso contrário, obtém-se o estimador dos mínimos quadrados ponderados estimados (ver a posição 3.2):

$$\tilde{\boldsymbol{\theta}}_{(-i)} \approx \boldsymbol{\theta}^* + [\mathbf{F}_{(-i)}^T(\boldsymbol{\theta}^*) \hat{\boldsymbol{\Sigma}}_{(-i)}^{-1} \mathbf{F}_{(-i)}(\boldsymbol{\theta}^*)]^{-1} \mathbf{F}_{(-i)}^T(\boldsymbol{\theta}^*) \hat{\boldsymbol{\Sigma}}_{(-i)}^{-1} [\bar{\mathbf{Y}}_{(-i)} - \mathbf{f}_{(-i)}(\boldsymbol{\theta}^*)].$$

O índice $(-i)$ significa que toda a informação, correspondente ao ponto experimental i , foi eliminada. Por exemplo, $\hat{\boldsymbol{\Sigma}}_{(-i)}$ é a matriz que se obtém ao retirar, à matriz de covariâncias $\hat{\boldsymbol{\Sigma}}$ (ver (2.7)), as i -ésimas linha e coluna. Considera-se que $n > m$, ou seja, o número de pontos experimentais é superior ao número de parâmetros regressores (caso contrário, $\mathbf{F}_{(-i)}$ seria certamente singular).

Terceiro passo

Prever a resposta para o ponto experimental i (eliminado) através de

$$\hat{Y}_{i(-i)} = f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{(-i)}), \quad (4.16)$$

ou

$$\hat{Y}_{i(-i)} = f(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}}_{(-i)}),$$

consoante a variância seja, ou não, constante respectivamente (observa-se que a previsão $\hat{Y}_{i(-i)}$ não depende do ponto experimental i).

Quarto passo

Comparar o resultado do terceiro passo, $\hat{Y}_{i(-i)}$, com o valor da média das respostas eliminadas no primeiro passo, \bar{Y}_i . Esta comparação é feita calculando a estatística (4.9)

$$t_i = \frac{\bar{Y}_i - \hat{Y}_{i(-i)}}{\left(\widehat{\text{Var}}[\bar{Y}_i] + \widehat{\text{Var}}[\hat{Y}_{i(-i)}]\right)^{1/2}}, \quad (4.17)$$

onde \bar{Y}_i é dado por (3.2) e

$$\widehat{\text{Var}}[\bar{Y}_i] = \frac{1}{r(r-1)} \sum_{j=1}^r (Y_{ij} - \bar{Y}_i)^2, \quad (4.18)$$

(ver (4.9) e (4.12) respectivamente). O cálculo da variância estimada, $\widehat{\text{Var}}[\hat{Y}_{i(-i)}]$, também aqui depende do caso: variância dos erros constante ou não.

No primeiro caso, tem-se

$$\widehat{\text{Var}}[\hat{Y}_{i(-i)}] \approx \frac{\hat{\sigma}_{(-i)}^2}{r} \hat{\mathbf{f}}_i^T [\hat{\mathbf{F}}_{(-i)}^T \hat{\mathbf{F}}_{(-i)}]^{-1} \hat{\mathbf{f}}_i, \quad (4.19)$$

onde $\hat{\mathbf{F}}_{(-i)} = \mathbf{F}_{(-i)}(\hat{\boldsymbol{\theta}}_{(-i)})$, $\hat{\mathbf{f}}_i^T = \mathbf{f}_i^T(\hat{\boldsymbol{\theta}}_{(-i)})$ e $\hat{\sigma}_{(-i)}^2$ é dada por

$$\hat{\sigma}_{(-i)}^2 = \frac{r}{n-1-m} \sum_{k=1, k \neq i}^n [\bar{Y}_k - f(\mathbf{X}_k, \hat{\boldsymbol{\theta}}_{(-i)})]^2 \quad (4.20)$$

(ver (3.8) considerando $n-1$ pontos experimentais).

No segundo caso, em vez de (4.19), tem-se

$$\widehat{\text{Var}}[\hat{Y}_{i(-i)}] \approx \frac{1}{r} \hat{\mathbf{f}}_i^T [\hat{\mathbf{F}}_{(-i)}^T \hat{\boldsymbol{\Sigma}}_{(-i)}^{-1} \hat{\mathbf{F}}_{(-i)}]^{-1} \hat{\mathbf{f}}_i,$$

onde $\hat{\mathbf{F}}_{(-i)} = \mathbf{F}_{(-i)}(\tilde{\boldsymbol{\theta}}_{(-i)})$ e $\hat{\mathbf{f}}_i^T = \mathbf{f}_i^T(\tilde{\boldsymbol{\theta}}_{(-i)})$.

Quinto passo

Aplicar a desigualdade de Bonferroni da seguinte forma: testam-se os n valores individuais t_i , a um nível α_E/n , e rejeita-se o metamodelo de regressão se

$$\max_{i=1, \dots, n} |t_i| > t_{r-1; \alpha_E/(2n)} \quad (4.21)$$

onde α_E é a probabilidade de rejeitar o metamodelo e o factor 2 é necessário visto tratar-se de um teste bilateral.

Observe-se que, uma vantagem deste método é que estão disponíveis n valores da estatística de teste, t , sem ser preciso realizar mais experiências de simulação que aquelas que já foram realizadas. Observa-se que, os erros de previsão normalizados são dependentes mesmo que as sementes, dos números pseudo-aleatórios utilizadas nas simulações, tenham sido diferentes. Por conseguinte, no quarto passo, existe a necessidade de recorrer à desigualdade de Bonferroni (ver [64, página 155] para uma explicação detalhada da desigualdade de Bonferroni, no contexto da validação de metamodelos).

B.1. "Otimização" computacional do método de validação cruzada. Observa-se que, o método de validação cruzada proposto é relativamente dispendioso do ponto de vista computacional, principalmente quando o número de pontos experimentais é grande. Isto acontece porque é preciso fazer n regressões não lineares e, por cada regressão, é necessário recorrer a um método iterativo. Por conseguinte, fazer n regressões pode ser um processo ineficiente, que consome algum tempo computacional, principalmente para grandes conjuntos de dados. Uma solução para aumentar a rapidez do processo iterativo é utilizar técnicas de aceleração de convergência, conjuntamente com o método iterativo. Pode-se também seguir a sugestão de Kleijnen [59] para metamodelos lineares e utilizar o procedimento num subconjunto do conjunto total de pontos experimentais. Mas neste caso, como o próprio Kleijnen afirma, existe a desvantagem de ignorar informação potencial. Contudo, é possível utilizar o método de validação cruzada proposto fazendo uma única regressão. Nas duas proposições seguintes observa-se que, para determinar $\hat{Y}_{i(-i)} - \bar{Y}_i$, não é necessário fazer uma nova regressão. Este valor é calculado à custa de somar a $\hat{Y}_i - \bar{Y}_i$ (onde $\hat{Y}_i = f(\mathbf{X}_i, \hat{\boldsymbol{\theta}})$, sendo $\hat{\boldsymbol{\theta}}$ o estimador dos mínimos quadrados, obtido a partir de todo o conjunto de resultados) uma quantidade que é calculada uma única vez e que depende apenas de \mathbf{F} , $\bar{\mathbf{Y}}$, \mathbf{f} e $\boldsymbol{\theta}^*$ (valor exacto). Sempre que exista necessidade, pode-se substituir $\boldsymbol{\theta}^*$ por $\hat{\boldsymbol{\theta}}$; não se conhecendo o valor exacto, o vector disponível mais próximo de $\boldsymbol{\theta}^*$ é $\hat{\boldsymbol{\theta}}$.

Proposição 4.1 Considere-se o modelo de regressão não linear (3.9):

$$\bar{Y}_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \bar{\epsilon}_i, \quad i = 1, 2, \dots, n,$$

sendo a variância dos erros constante e dada por (3.11), $\bar{\epsilon}_i \sim N(0, \sigma^2/r)$. Sejam $\hat{Y}_{i(-i)} - \bar{Y}_i$ e $\hat{Y}_i - \bar{Y}_i$ os resíduos (ou erros) correspondentes à regressão onde é eliminado o ponto experimental

$i, i = 1, 2, \dots, n$, e à regressão com todos os resultados, respectivamente. Então

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \hat{Y}_i - \bar{Y}_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}}\right) \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \hat{\boldsymbol{\theta}})], \quad (4.22)$$

onde os \hat{h}_{ii} são os elementos da diagonal principal da matriz "chapéu" estimada

$$\hat{\mathbf{H}} = \hat{\mathbf{F}} (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T, \quad (4.23)$$

correspondente à regressão com todos os dados, $\hat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$ e $\hat{\mathbf{f}}_i = \mathbf{f}_i(\hat{\boldsymbol{\theta}})$. Além disso,

$$\widehat{\text{Var}}[\hat{Y}_{i(-i)}] \approx \frac{\hat{\sigma}_{(-i)}^2}{r} \hat{\mathbf{f}}_i^T [\hat{\mathbf{F}}^T \hat{\mathbf{F}} - \hat{\mathbf{f}}_i \hat{\mathbf{f}}_i^T]^{-1} \hat{\mathbf{f}}_i, \quad (4.24)$$

onde $\hat{\sigma}_{(-i)}^2$ é dado por (4.20) e

$$\hat{\boldsymbol{\theta}}_{(-i)} \approx \hat{\boldsymbol{\theta}} + \frac{1}{1 - \hat{h}_{ii}} (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i [\bar{Y}_i - \mathbf{f}(\mathbf{X}_i, \hat{\boldsymbol{\theta}})]. \quad (4.25)$$

Verificação Comece-se por verificar o resultado (4.22). Para isso considere-se o problema (3.9), com distribuição de erros (3.11). Sabe-se que, numa vizinhança suficientemente próxima de $\boldsymbol{\theta}^*$, tem-se o desenvolvimento linear de Taylor [107, página 23]:

$$f(\mathbf{X}_i, \boldsymbol{\theta}) \approx f(\mathbf{X}_i, \boldsymbol{\theta}^*) + \sum_{j=1}^m \frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \theta_j} (\theta_j - \theta_j^*), \quad (4.26)$$

Substituindo (4.26) em (3.9), obtém-se

$$\bar{Y}_i \approx f(\mathbf{X}_i, \boldsymbol{\theta}^*) + \sum_{j=1}^m \frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \theta_j} (\theta_j - \theta_j^*) + \bar{\epsilon}_i.$$

ou seja,

$$\bar{Y}_i - f(\mathbf{X}_i, \boldsymbol{\theta}^*) \approx \sum_{j=1}^m \frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \theta_j} (\theta_j - \theta_j^*) + \bar{\epsilon}_i.$$

É possível afirmar que o problema anterior é aproximadamente

$$Z_i = \mathbf{f}_i^T(\boldsymbol{\theta}^*) \boldsymbol{\beta} + \bar{\epsilon}_i, \quad i = 1, \dots, n, \quad (4.27)$$

ou, de outro modo,

$$\mathbf{Z} = \mathbf{F} \boldsymbol{\beta} + \bar{\boldsymbol{\epsilon}},$$

onde $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$. Utilizando a conhecida teoria da regressão linear, pode-se afirmar que

$$d_i = \frac{e_i}{1 - h_{ii}},$$

onde e_i são os resíduos usuais e d_i são os resíduos previstos definidos por [87, página 399]:

$$d_i = Z_i - \hat{Z}_{i(-i)}$$

e

$$e_i = Z_i - \hat{Z}_i.$$

Mas

$$\begin{aligned} Z_i &= \bar{Y}_i - f(\mathbf{X}_i, \boldsymbol{\theta}^*), \\ \hat{Z}_{i(-i)} &= \hat{Y}_{i(-i)} - f(\mathbf{X}_i, \hat{\boldsymbol{\theta}}_{(-i)}), \end{aligned}$$

logo

$$\hat{Y}_{i(-i)} - \bar{Y}_i = \hat{Z}_{i(-i)} - Z_i + f(\mathbf{X}_i, \hat{\boldsymbol{\theta}}_{(-i)}) - f(\mathbf{X}_i, \boldsymbol{\theta}^*)$$

Numa vizinhança de $\boldsymbol{\theta}^*$, como $\hat{\boldsymbol{\theta}}_{(-i)}$ estará próximo de $\boldsymbol{\theta}^*$, tem-se

$$f(\mathbf{X}_i, \hat{\boldsymbol{\theta}}_{(-i)}) \approx f(\mathbf{X}_i, \boldsymbol{\theta}^*) + \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^*). \quad (4.28)$$

Por conseguinte,

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \hat{Z}_{i(-i)} - Z_i + \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^*). \quad (4.29)$$

Aplicando a teoria da regressão linear, pode-se escrever

$$\hat{Z}_{i(-i)} - Z_i = \frac{\hat{Z}_i - Z_i}{1 - h_{ii}}, \quad (4.30)$$

onde

$$h_{ii} = \mathbf{f}_i^T(\boldsymbol{\theta}^*)[\mathbf{F}^T(\boldsymbol{\theta}^*)\mathbf{F}(\boldsymbol{\theta}^*)]^{-1}\mathbf{f}_i(\boldsymbol{\theta}^*)$$

é o i -ésimo elemento da diagonal principal da matriz "chapéu"

$$\mathbf{H} = \mathbf{F}(\boldsymbol{\theta}^*)[\mathbf{F}^T(\boldsymbol{\theta}^*)\mathbf{F}(\boldsymbol{\theta}^*)]^{-1}\mathbf{F}^T(\boldsymbol{\theta}^*)$$

(ver [107, página 211]). Substituindo (4.30) em (4.29) vem,

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \frac{\hat{Z}_i - Z_i}{1 - h_{ii}} + \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^*). \quad (4.31)$$

Mas $\hat{Z}_i = \hat{Y}_i - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})$ e $Z_i = \bar{Y}_i - f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*)$, logo

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \frac{\hat{Y}_i - \bar{Y}_i + f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}})}{1 - h_{ii}} + \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^*). \quad (4.32)$$

Tal como $\hat{\boldsymbol{\theta}}_{(-i)}$, também $\hat{\boldsymbol{\theta}}$ está próximo do valor exacto $\boldsymbol{\theta}^*$. Por conseguinte, numa vizinhança de $\boldsymbol{\theta}^*$

$$f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) \approx f(\mathbf{X}_{i.}, \boldsymbol{\theta}^*) + \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*),$$

logo (4.32) surge na forma

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \frac{\hat{Y}_i - \bar{Y}_i}{1 - h_{ii}} - \frac{1}{1 - h_{ii}} \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + \mathbf{f}_i^T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^*). \quad (4.33)$$

Mas é verdade que,

$$\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*. \quad (4.34)$$

Além disso, a aplicação do resultado (5.42) de [107, página 211] ao problema (3.9) resulta na seguinte aproximação:

$$\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}} \approx \frac{(\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i}{1 - \hat{h}_{ii}} (\bar{Y}_i - \hat{Y}_i). \quad (4.35)$$

Substituindo (4.35) em (4.34) obtém-se

$$\hat{\boldsymbol{\theta}}_{(-i)} - \boldsymbol{\theta}^* \approx \frac{(\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i}{1 - \hat{h}_{ii}} (\bar{Y}_i - \hat{Y}_i) + \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*.$$

É possível aplicar este resultado em (4.33) e escrever

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \frac{\hat{Y}_i - \bar{Y}_i}{1 - \hat{h}_{ii}} - \frac{1}{1 - \hat{h}_{ii}} \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + \frac{\hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i}{1 - \hat{h}_{ii}} (\bar{Y}_i - \hat{Y}_i) + \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \quad (4.36)$$

onde foram feitas as aproximações $\mathbf{f}_i(\boldsymbol{\theta}^*) \approx \mathbf{f}_i(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{f}}_i$, $h_{ii} \approx \hat{h}_{ii} = \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i$ e $\mathbf{F}(\boldsymbol{\theta}^*) \approx \mathbf{F}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{F}}$. Uma vez que $\hat{h}_{ii} = \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i$, então

$$\begin{aligned} \hat{Y}_{i(-i)} - \bar{Y}_i &\approx \frac{\hat{Y}_i - \bar{Y}_i}{1 - \hat{h}_{ii}} - \frac{1}{1 - \hat{h}_{ii}} \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) - \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} (\hat{Y}_i - \bar{Y}_i) + \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \\ &= \frac{1}{1 - \hat{h}_{ii}} (1 - \hat{h}_{ii}) (\hat{Y}_i - \bar{Y}_i) - \frac{1}{1 - \hat{h}_{ii}} \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \\ &= \hat{Y}_i - \bar{Y}_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}}\right) \hat{\mathbf{f}}_i^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \end{aligned} \quad (4.37)$$

Aplicando (2.15) de [107, página 23] ao problema (3.9), obtém-se a aproximação seguinte:

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \approx [\mathbf{F}^T(\boldsymbol{\theta}^*) \mathbf{F}(\boldsymbol{\theta}^*)]^{-1} \mathbf{F}^T(\boldsymbol{\theta}^*) [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)].$$

Utilizando $\mathbf{F}(\boldsymbol{\theta}^*) \approx \hat{\mathbf{F}}$, é possível escrever a aproximação anterior na forma

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \approx (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)]. \quad (4.38)$$

Substituindo (4.38) em (4.37) obtém-se

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \hat{Y}_i - \bar{Y}_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}}\right) \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)]. \quad (4.39)$$

A verificação da aproximação (4.22) fica concluída substituindo $\boldsymbol{\theta}^*$ por $\hat{\boldsymbol{\theta}}$ na aproximação anterior.

O resultado (4.24) obtém-se da forma descrita em seguida. Aplicando (3.6), obtém-se

$$\text{Var}[\hat{Y}_{i(-i)}] \approx \frac{\sigma_{(-i)}^2}{r} [\mathbf{F}_{(-i)}^T \mathbf{F}_{(-i)}]^{-1},$$

donde

$$\widehat{\text{Var}}[\hat{Y}_{i(-i)}] \approx \frac{\hat{\sigma}_{(-i)}^2}{r} [\hat{\mathbf{F}}_{(-i)}^T \hat{\mathbf{F}}_{(-i)}]^{-1}.$$

Mas $\hat{\mathbf{F}}_{(-i)}^T \hat{\mathbf{F}}_{(-i)} = \hat{\mathbf{F}}^T \hat{\mathbf{F}} - \hat{\mathbf{f}}_i \hat{\mathbf{f}}_i^T$ (ver (5.40) de [107, página 211]), logo verifica-se (4.24).

O resultado (4.20) é útil para calcular $\hat{\sigma}_{(-i)}^2$, sem ser preciso fazer regressões sucessivas. Uma vez que $\hat{Y}_i = f(\mathbf{X}_i, \hat{\boldsymbol{\theta}})$, (4.25) resulta directamente da aproximação (4.35). Fica aqui concluída a verificação da proposição 4.1. \diamond

Proposição 4.2 Considere-se o modelo de regressão não linear (3.9)

$$\bar{Y}_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \bar{\epsilon}_i, \quad i = 1, 2, \dots, n,$$

quando a variância dos erros satisfaz (3.14)

$$\text{Var}[\bar{Y}_i] = \frac{\sigma_i^2}{r}.$$

Sejam $\hat{Y}_{i(-i)} - \bar{Y}_i$ e $\hat{Y}_i - \bar{Y}_i$ os resíduos (ou erros) correspondentes à regressão onde é eliminado o ponto experimental i , $i = 1, 2, \dots, n$, e a regressão com todos os dados respectivamente. Então

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \hat{Y}_i - \bar{Y}_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}^*}\right) \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \tilde{\boldsymbol{\theta}})], \quad (4.40)$$

onde os \hat{h}_{ii}^* são os elementos da diagonal principal da matriz "chapéu" modificada estimada

$$\hat{\mathbf{H}}^* = \hat{\mathbf{R}}\hat{\mathbf{F}}(\hat{\mathbf{F}}^T\hat{\Sigma}^{-1}\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}^T\hat{\mathbf{R}}^T, \quad (4.41)$$

correspondente à regressão com todos os dados. $\hat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$, $\hat{\mathbf{f}}_i = \mathbf{f}_i(\hat{\boldsymbol{\theta}})$ e as matrizes $\hat{\Sigma}$ e $\hat{\mathbf{R}}$ são dadas respectivamente por (3.16) e

$$\hat{\mathbf{R}} = \begin{bmatrix} 1/\hat{\sigma}_1 & & 0 \\ & \ddots & \\ 0 & & 1/\hat{\sigma}_n \end{bmatrix}, \quad (4.42)$$

onde os valores $\hat{\sigma}_i$, $i = 1, \dots, n$, são calculados utilizando (3.1). Além disso,

$$\widehat{\text{Var}}[\hat{Y}_{i(-i)}] \approx \frac{1}{r-1}\hat{\mathbf{f}}_i^T[\hat{\mathbf{F}}^T\hat{\Sigma}^{-1}\hat{\mathbf{F}} - \frac{1}{\hat{\sigma}_i^2}\hat{\mathbf{f}}_i\hat{\mathbf{f}}_i^T]^{-1}\hat{\mathbf{f}}_i. \quad (4.43)$$

Verificação Foi observado na Secção 3.2 que o problema (3.9), com variância heterogénea, é equivalente ao problema (3.20). Aplicando a aproximação (4.39) ao problema (3.20) obtém-se

$$\hat{W}_{i(-i)} - W_i \approx \hat{W}_i - W_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}^*}\right) \hat{\mathbf{g}}_i^T (\hat{\mathbf{G}}^T \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T [\mathbf{W} - \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}^*)],$$

onde \hat{h}_{ii}^* são os elementos da diagonal principal da matriz "chapéu" estimada

$$\hat{\mathbf{H}} = \hat{\mathbf{G}}(\hat{\mathbf{G}}^T \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T. \quad (4.44)$$

Note-se que

$$\hat{\mathbf{g}}_i = \frac{1}{\sigma_i} \hat{\mathbf{f}}_i.$$

Além disso, $\hat{\mathbf{G}} = \hat{\mathbf{R}}\hat{\mathbf{F}}$, $\mathbf{W} = \hat{\mathbf{R}}\bar{\mathbf{Y}}$, $\hat{\mathbf{g}} = \hat{\mathbf{R}}\hat{\mathbf{f}}$ e $\hat{\mathbf{R}}^T\hat{\mathbf{R}} = \hat{\Sigma}^{-1}$, com $\hat{\mathbf{R}}$ dado por (4.42). Por conseguinte,

$$\hat{W}_{i(-i)} - W_i \approx \hat{W}_i - W_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}^*}\right) \frac{1}{\sigma_i} \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)].$$

Mas

$$\begin{aligned} \hat{W}_{i(-i)} &= \frac{1}{\sigma_i} \hat{Y}_{i(-i)}, \\ \hat{W}_i &= \frac{1}{\sigma_i} \hat{Y}_i, \\ W_i &= \frac{1}{\sigma_i} \bar{Y}_i. \end{aligned}$$

logo

$$\hat{Y}_{i(-i)} - \bar{Y}_i \approx \hat{Y}_i - \bar{Y}_i + \left(1 - \frac{1}{1 - \hat{h}_{ii}^*}\right) \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T \hat{\Sigma}^{-1} [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}^*)].$$

A verificação de (4.40) termina substituindo $\boldsymbol{\theta}^*$ por $\tilde{\boldsymbol{\theta}}$ na aproximação anterior. Para calcular a matriz "chapéu" modificada estimada, substitui-se $\hat{\mathbf{G}} = \hat{\mathbf{R}}\hat{\mathbf{F}}$ em (4.44) para obter

$$\hat{\mathbf{H}}^* = \hat{\mathbf{R}}\hat{\mathbf{F}}(\hat{\mathbf{F}}^T \hat{\mathbf{R}}^T \hat{\mathbf{R}}\hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}^T \hat{\mathbf{R}}^T.$$

Aplicando $\hat{\mathbf{R}}^T \hat{\mathbf{R}} = \hat{\Sigma}^{-1}$ na igualdade anterior obtém-se (4.41).

Aplicando a Proposição 3.2 e considerando todos os pontos experimentais, excepto o ponto experimental i , podemos escrever:

$$\mathcal{D}[\tilde{\boldsymbol{\theta}}_{(-i)}] \approx \frac{1}{r-1} [\mathbf{F}_{(-i)}^T(\boldsymbol{\theta}^*) \boldsymbol{\Sigma}_{(-i)}^{-1} \mathbf{F}_{(-i)}(\boldsymbol{\theta}^*)]^{-1}.$$

Por conseguinte,

$$\hat{\mathcal{D}}[\tilde{\boldsymbol{\theta}}_{(-i)}] \approx \frac{1}{r-1} [\hat{\mathbf{F}}_{(-i)}^T \hat{\Sigma}_{(-i)}^{-1} \hat{\mathbf{F}}_{(-i)}]^{-1}$$

onde $\hat{\mathbf{F}}_{(-i)} = \mathbf{F}_{(-i)}(\tilde{\boldsymbol{\theta}})$. Mas $\hat{\Sigma}_{(-i)}^{-1} = \hat{\mathbf{R}}_{(-i)}^T \hat{\mathbf{R}}_{(-i)}$ logo

$$\hat{\mathcal{D}}[\tilde{\boldsymbol{\theta}}_{(-i)}] \approx \frac{1}{r-1} [\hat{\mathbf{F}}_{(-i)}^T \hat{\mathbf{R}}_{(-i)}^T \hat{\mathbf{R}}_{(-i)} \hat{\mathbf{F}}_{(-i)}]^{-1},$$

onde $\hat{\mathbf{R}}_{(-i)}$ é a matriz que se obtém ao eliminar as i -ésimas linha e coluna da matriz (4.42).

Porém,

$$\hat{\mathbf{F}}_{(-i)}^T \hat{\mathbf{R}}_{(-i)}^T \hat{\mathbf{R}}_{(-i)} \hat{\mathbf{F}}_{(-i)} = (\hat{\mathbf{R}}_{(-i)} \hat{\mathbf{F}}_{(-i)})^T \hat{\mathbf{R}}_{(-i)} \hat{\mathbf{F}}_{(-i)} = (\hat{\mathbf{R}}\hat{\mathbf{F}})^T \hat{\mathbf{R}}\hat{\mathbf{F}} - \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^T$$

onde $\hat{\mathbf{g}}_i^T$ é a i -ésima linha de $\mathbf{R}\hat{\mathbf{F}}$,

$$\hat{\mathbf{g}}_i^T = \frac{1}{\sigma_i} \hat{\mathbf{f}}_i^T. \quad (4.45)$$

Sendo assim, é possível escrever

$$\begin{aligned} \hat{\mathcal{D}}[\tilde{\boldsymbol{\theta}}_{(-i)}] &\approx \frac{1}{r-1} [\hat{\mathbf{F}}^T \hat{\mathbf{R}}^T \hat{\mathbf{R}}\hat{\mathbf{F}} - \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^T]^{-1} = \\ &= \frac{1}{r-1} [\hat{\mathbf{F}}^T \hat{\Sigma}^{-1} \hat{\mathbf{F}} - \frac{1}{\sigma_i^2} \hat{\mathbf{f}}_i \hat{\mathbf{f}}_i^T]^{-1} \diamond \end{aligned}$$

Com base nos resultados das proposições anteriores, apresenta-se em seguida o método optimizado.

Primeiro passo

Calcular o estimador dos mínimos quadrados, $\hat{\theta}$ ou $\tilde{\theta}$ consoante a variância dos erros seja ou não constante, respectivamente.

Segundo passo

Construir a matriz "chapéu" aproximada através de (4.23), se a variância dos erros for constante; caso contrário, utiliza-se (4.41).

Para $i = 1, 2, \dots, n$, repetir os passos terceiro a sexto.

Terceiro passo

Calcular os resíduos $\hat{Y}_{i(-i)} - \bar{Y}_i$, utilizando (4.22) ou (4.40), consoante a variância dos erros seja ou não constante.

Quarto passo

Calcular $\widehat{\text{Var}}[\hat{Y}_{i(-i)}]$ com base em (4.24), (4.20) e (4.25), se a variância dos erros for constante; caso contrário, utiliza-se (4.43) e (3.1).

Quinto passo

Calcular $\widehat{\text{Var}}[\bar{Y}_i]$ utilizando (4.18).

Sexto passo

Calcular t_i através de (4.17), $i = 1, \dots, n$.

Sétimo passo

Rejeitar o metamodelo de regressão, se se verificar o critério (4.21).

B.2. Diagnósticos relacionados com a estatística t de Kleijnen Uma característica importante da estatística (4.17) é que ela fornece uma medida, que não depende de uma escala, para avaliar a dimensão dos resíduos; mais concretamente, não depende da localização do i -ésimo ponto experimental que, como se sabe, é medida pelo i -ésimo elemento da diagonal principal da matriz "chapéu". Pannis, Myers e Houck [94] dão ênfase à semelhança entre esta estatística e a estatística R de Student, utilizada em diagnósticos de regressão. Ou seja, a estatística proposta por Kleijnen é uma medida que permite aceder quase directamente a um grande número de técnicas para a análise de resíduos. Uma das suas mais importantes aplicações é a detecção de

observações muito afastadas ou suspeitas ¹. Para identificar essas observações, \bar{Y}_i , $i = 1, \dots, n$, examinam-se os valores absolutos da estatística (4.17) e utiliza-se a distribuição t apropriada, para averiguar até que ponto essas observações estão afastadas. A razão para utilizar resíduos previstos, em vez dos resíduos usuais, na detecção de observações suspeitas é que se \bar{Y}_i for suspeita (muito afastada), então o metamodelo estimado com base em todos os pontos experimentais, pode ser influenciado para produzir um valor próximo de \bar{Y}_i , prevendo um valor de \hat{Y}_i perto de \bar{Y}_i . O resíduo usual, correspondente a este ponto experimental, será pequeno e não irá revelar que \bar{Y}_i está afastado. Por outro lado, se o i -ésimo ponto experimental for excluído, antes do metamodelo ser ajustado, a valor ajustado $\hat{Y}_{i(-i)}$ não é influenciado pela observação afastada \bar{Y}_i , e o resíduo terá tendência a ser maior; por conseguinte, mais provavelmente irá revelar o afastamento da média \bar{Y}_i . Observa-se que, se \hat{h}_{ii} estiver próximo de zero, ou seja se \bar{Y}_i não estiver afastado, então o valor do resíduo previsto é aproximadamente o do resíduo usual; ver o resultado (4.22). Para os mínimos quadrados ponderados estimados a conclusão é a mesma, mas com \hat{h}_{ii} substituído por \hat{h}_{ii}^* ; ver o resultado (4.40). Sendo assim, por vezes, os resíduos previstos permitem identificar observações afastadas, o que não acontece com os resíduos usuais, de um modo geral. Outras vezes, tanto os resíduos previstos, como os usuais levam à mesma identificação.

Após a identificação de valores da variável de resposta afastados, o próximo passo é averiguar até que ponto estes valores afastados são influentes. Um valor influente é um cuja exclusão provoca mudanças significativas no metamodelo ajustado. Neste trabalho, consideram-se três medidas de influência muito utilizadas em estatística, todas elas baseadas na omissão da informação correspondente a um ponto experimental para medir a sua influência.

- *Influência nos valores ajustados.* Uma estratégia, para examinar a influência da i -ésima média \bar{Y}_i , no valor ajustado \hat{Y}_i , é calcular o número aproximado de erros padrão estimados pelo qual o valor ajustado \hat{Y}_i se modifica, quando o i -ésimo ponto experimental é removido do conjunto de resultados. A estatística apropriada é dada por

$$(\text{DFFITs})_i = \frac{\hat{Y}_i - \hat{Y}_{i(-i)}}{(\widehat{\text{Var}}[\hat{Y}_i])^{1/2}}. \quad (4.46)$$

¹outliers [87, páginas 121-122]

O prefixo 'DF' denota a diferença entre o resultado com e sem a informação correspondente ao i -ésimo ponto experimental (ver [87, página 401]). Como guia para identificar um valor da variável de resposta influente, Neter, Wasserman e Kutner [87, página 401] sugerem um valor absoluto de $(\text{DFFITs})_i$ que exceda 1 para amostras de dimensão média e pequena e $2\sqrt{m/r}$ para grandes conjuntos de resultados. Os valores de $(\text{DFFITs})_i$ podem ser calculados utilizando apenas os resultados do ajuste com base em todo o conjunto de resultados. Consoante o método dos mínimos quadrados utilizado, assim é a forma de cálculo deste diagnóstico:

(i) no caso dos mínimos quadrados usuais, tem-se

$$\hat{Y}_i = f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}),$$

$$\hat{Y}_{i(-i)} = f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}_{(-i)}),$$

onde

$$\hat{\boldsymbol{\theta}}_{(-i)} \approx \hat{\boldsymbol{\theta}} + \frac{(\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i}{1 - \hat{h}_{ii}} (\bar{Y}_{i.} - \hat{Y}_i) \quad (4.47)$$

e

$$\widehat{\text{Var}}[\hat{Y}_i] \approx \frac{\hat{\sigma}^2}{r} \hat{\mathbf{f}}_i^T (\hat{\mathbf{F}}^T \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i, \quad (4.48)$$

com $\hat{\sigma}^2$ dada por (3.8). (4.47) e (4.48) resultam de (4.35) e (4.14) respectivamente; e

(ii) no caso dos mínimos quadrados ponderados estimados, tem-se $\hat{Y}_i = f(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}})$, $\hat{Y}_{i(-i)} = f(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}}_{(-i)})$, onde $\tilde{\boldsymbol{\theta}}_{(-i)}$ é dado por

$$\tilde{\boldsymbol{\theta}}_{(-i)} \approx \tilde{\boldsymbol{\theta}} + \frac{1}{\hat{\sigma}_i (1 - \hat{h}_{ii}^*)} (\hat{\mathbf{F}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathbf{F}})^{-1} \hat{\mathbf{f}}_i \hat{\mathbf{R}} [\bar{\mathbf{Y}} - \mathbf{f}(\mathbf{X}_{i.}, \tilde{\boldsymbol{\theta}})]. \quad (4.49)$$

O resultado (4.49) obtém-se aplicando (4.25) ao problema (3.20) e fazendo as devidas substituições.

• *Influência nos coeficientes de regressão.* Uma medida para avaliar a influência do i -ésimo ponto experimental em cada coeficiente de regressão estimado $\hat{\theta}_j$, $j = 1, \dots, m$, é o número de erros padrão estimados, pelo qual o coeficiente de regressão $\hat{\theta}_j$ se modifica, quando se exclui o i -ésimo ponto experimental. Devidamente normalizada, esta medida é definida como

$$(\text{DFTETAS})_{i,j} = \frac{\hat{\theta}_j - \hat{\theta}_{j(-i)}}{(\widehat{\text{Var}}[\hat{\theta}_j])^{1/2}}.$$

Um valor elevado de $(DFTETAS)_{i,j}$ indica que o i -ésimo ponto experimental tem um impacto grande no j -ésimo coeficiente de regressão (ver [87, página 402]).

Uma medida geral, do impacto combinado do i -ésimo ponto experimental em todos os coeficientes de regressão, é a medida de distância de Cook (ver [87, página 403]). Esta medida obtém-se a partir do conceito de região de confiança para todos os parâmetros de regressão θ_i , $i = 1, \dots, m$, simultaneamente.

- *Influência nas inferências.* Para fornecer mais informação, sobre a determinação de pontos experimentais influentes, é habitualmente uma boa ideia examinar de uma forma directa as inferências a partir do metamodelo ajustado que será obtido com e sem o ponto experimental em questão (ver [87, página 405]).

Para medir o efeito do ponto experimental j nas inferências a serem levadas a cabo a partir do metamodelo ajustado calcula-se, para os n pontos experimentais, a diferença entre os valores previstos \hat{Y}_i e $\hat{Y}_{i(-j)}$. Em seguida, obtém-se a média das diferenças absolutas percentuais

$$\frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{Y}_{i(-j)} - \hat{Y}_i}{\hat{Y}_i} \right| \times 100.$$

Finaliza-se esta secção salientando que, excepto as citações ao livro de Seber e Wild [107], todas outras referências citadas discutem o caso linear. Por este motivo, houve necessidade de fazer as devidas adaptações ao caso não linear.

4.2 Validação com respeito ao modelo de simulação

O método preferencial, para validar um metamodelo em relação ao modelo de simulação, é através da recolha de novos resultados de simulação. Uma alternativa razoável, quando o conjunto de resultados é suficientemente grande, é dividir os resultados em dois conjuntos. O primeiro, denominado conjunto de construção do modelo, é utilizado para desenvolver o metamodelo. O segundo conjunto de resultados, denominado conjunto de validação, é utilizado para avaliar a

razoabilidade e a capacidade previsiva do metamodelo. Este procedimento de validação é frequentemente denominado por validação cruzada. A partição dos resultados simula a replicação parcial, ou completa, da experiência de simulação. O conjunto de resultados de validação é utilizado para validação da mesma forma que quando novos resultados são recolhidos. Por conseguinte, é possível reestimar os coeficientes de regressão com base no conjunto de validação e compará-los para consistência, com os coeficientes obtidos com o conjunto de construção do modelo. Também podem ser feitas previsões para a variável resposta no conjunto de validação a partir do metamodelo desenvolvido com o conjunto de construção. Deste modo calibra-se o metamodelo para novos valores da variável resposta.

O conjunto de resultados é frequentemente dividido em dois conjuntos com igual número de elementos: resultados de construção e de validação. É importante, contudo, que o conjunto de resultados de construção do modelo seja suficientemente grande, de tal forma que possa ser desenvolvido um metamodelo seguro. Neter, Wasserman e Kutner [87, página 467] aconselham que o número de casos seja, pelo menos, 6 a 10 vezes o número de variáveis independentes. No caso não linear, como o número de variáveis independentes não está directamente relacionado com o número de parâmetros do modelo, aconselhamos a que o número de casos seja, pelo menos, 6 a 10 vezes o número de parâmetros desconhecidos do metamodelo de simulação. Por conseguinte, se o número de parâmetros desconhecidos for 2, então o conjunto de resultados de construção do metamodelo deverá conter pelo menos 12 a 20 elementos. Se o conjunto total de resultados não for suficientemente grande, para ser possível uma divisão em dois conjuntos aproximadamente com o mesmo número de elementos e levando em conta a exigência atrás mencionada, então deverá escolher-se um conjunto de resultados de validação mais pequeno que o conjunto de resultados de construção. A divisão dos dados pode ser feita aleatoriamente ou intuitivamente. Outra possibilidade é juntar os casos em pares e colocar um de cada par em cada conjunto de resultados.

- **Validação cruzada dupla** Com base na técnica descrita anteriormente, sugere-se em seguida um método de partição de resultados, para averiguar a validade do metamodelo em relação ao modelo de simulação.

Primeiro passo

Dividir o conjunto de resultados em dois conjuntos de aproximadamente igual dimensão.

Segundo passo

Desenvolver um metamodelo independentemente para cada um dos conjuntos de resultados, obtidos no primeiro passo, e comparar os coeficientes dos dois metamodelos. Se os coeficientes forem muito diferentes, então o metamodelo pode ter falta de consistência interna. Neste caso, rejeita-se o metamodelo.

Terceiro passo

Utilizar cada metamodelo de regressão, construído com base num conjunto de resultados, para previsão no outro conjunto. Para isso, calcular dois valores do coeficiente de determinação: um para os resultados utilizados na construção do metamodelo, R_{const}^2 , e o outro para os resultados do conjunto de validação, R_{val}^2 . Se estes valores forem muito diferentes, então não se considera o metamodelo válido. Para cada metamodelo, calcular também o erro quadrático médio de previsão MSPR, com base nos resultados não utilizados na sua construção [87, página 466]. Para o problema (3.3), calcula-se MSPR através de

$$MSPR = \frac{1}{nr} \sum_{i \in I_v} \sum_{j=1}^r \frac{1}{\sigma_i^2} (Y_{ij} - \hat{Y}_{ij})^2,$$

onde \hat{Y}_{ij} é o valor previsto para Y_{ij} , baseado no conjunto de dados de construção do metamodelo. I_v é o conjunto dos índices dos pontos experimentais que não pertencem ao conjunto de construção do metamodelo. Se a variância dos erros for constante, então

$$MSPR = \frac{1}{nr} \sum_{i \in I_v} \sum_{j=1}^r (Y_{ij} - \hat{Y}_{ij})^2.$$

Se MSPR for bastante próximo do valor de MSE, correspondente ao ajuste baseado nos resultados utilizados na construção do modelo, então o erro quadrático médio MSE dá uma indicação apropriada da capacidade previsiva do modelo. Se MSE for muito menor que MSPR, então MSPR é um indicador da capacidade previsiva do modelo.

Uma possível desvantagem da separação de resultados é que as variâncias dos coeficientes de regressão estimados, desenvolvidos a partir de um subconjunto de resultados, irão frequentemente ser maiores que os obtidos a partir do ajuste com todo o conjunto de resultados. Contudo, se esse

subconjunto for suficientemente grande, estas variâncias geralmente não serão muito maiores que as obtidas a partir de todo o conjunto de resultados. Em qualquer dos casos, após o modelo ter sido validado, é prática usual utilizar o metamodelo estimado com base em todo o conjunto de resultados, embora talvez seja teoricamente mais apropriado utilizar o modelo ajustado a partir apenas do conjunto de construção do modelo.

4.3 Validação com respeito ao problema real

Nesta secção, o objectivo é averiguar se o metamodelo aproxima suficientemente bem o comportamento do problema real. Para isso, considera-se um determinado conjunto de n_w pontos experimentais. Consoante o conhecimento do problema real, este conjunto pode ser igual ou diferente do utilizado até ao presente momento; pode acontecer que só seja conhecida a resposta do problema real para um conjunto reduzido de pontos experimentais, que nem sequer coincida com o conjunto utilizado no desenho experimental escolhido para estimação do metamodelo. Em seguida, calcula-se a percentagem de erro relativo médio

$$\frac{1}{n_w} \sum_{i=1}^{n_w} \frac{|W_i - \hat{W}_i|}{|W_i|} \times 100\%,$$

onde W_i é o valor exacto (resposta do problema objecto) e \hat{W}_i é a resposta prevista pelo metamodelo, para o i -ésimo ponto experimental. Esta percentagem de erro foi utilizada, no contexto da simulação, por Friedman e Friedman [37]. Contudo, propomos utilizá-la também no caso não linear. No caso da resposta poder ser nula, propomos utilizar o erro absoluto em vez do erro relativo, ou seja

$$\frac{1}{n_w} \sum_{i=1}^{n_w} |W_i - \hat{W}_i|.$$

Capítulo 5

Variáveis de controlo

Uma característica interessante, das experiências de simulação em computador, é a possibilidade que o utilizador tem de observar diferentes variáveis aleatórias e parâmetros conhecidos. Esta propriedade deu origem ao método, de redução de variância, das variáveis de controlo. O método das variáveis de controlo tem sido largamente utilizado em simulação, nomeadamente na estimação de metamodelos de regressão. Neste capítulo, apresentam-se procedimentos para usar diversas variáveis de controlo concomitantes numa experiência de simulação, com acontecimentos discretos, em que o objectivo é estimar um modelo não linear geral com uma única resposta, expressa em termos de variáveis de decisão, seleccionadas no sistema em estudo. O objectivo de utilizar variáveis de controlo é reduzir o erro (variância) dos parâmetros estimados no metamodelo, em comparação com o que seria obtido através da estimação directa (sem variáveis de controlo). Assumindo que a resposta e as variáveis de controlo têm distribuição normal conjunta, com estrutura de variância-covariância homogénea ao longo de todos os pontos experimentais, desenvolvem-se procedimentos com variáveis de controlo para estimar os parâmetros desconhecidos do metamodelo, construir regiões de confiança e realizar testes de hipóteses. Com o objectivo de quantificar a eficiência máxima que é possível atingir, utilizando um determinado conjunto de variáveis de controlo, obtém-se a razão mínima entre as variâncias generalizadas. Para medir a degradação na eficiência que ocorre quando os coeficientes de controlo óptimos têm de ser estimados (através do método dos mínimos quadrados), apresenta-se o factor de perda generalizado.

5.1 Metamodelos com variáveis de controlo

Suponha-se que, o modelo de simulação pode ser representado pelo metamodelo (3.3), onde os erros ϵ_{ij} são variáveis aleatórias independentes, com distribuição

$$\epsilon_{ij} \sim N(0, \sigma^2), \quad \sigma > 0.$$

Assuma-se que, o erro $\epsilon_{ij} = Y_{ij} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})$, no problema (3.3), tem uma regressão linear no vector de controlo $\mathbf{C}_{ij} = (C_{ij1}, \dots, C_{ijq})$ com um vector desconhecido de coeficientes de controlo, $\boldsymbol{\delta}$, e com um resíduo ε_{ij} , tal que o modelo de simulação também pode ser descrito através do metamodelo de simulação com réplicas

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \mathbf{C}_{ij}\boldsymbol{\delta} + \varepsilon_{ij}, \quad (5.1)$$

onde $i = 1, \dots, n$ e $j = 1, \dots, r$.

Nesta formulação, utiliza-se o método das variáveis de controlo na sua forma linear, porque qualquer esquema não linear de variáveis de controlo se comporta assintoticamente como um método linear de variáveis de controlo [42]. Na verdade, não se consegue melhorar a eficiência assintótica do método, generalizando a forma de inclusão das variáveis de controlo, do caso linear para o não linear. No entanto, convém salientar que, não se põe de parte a possibilidade de obtenção de melhor desempenho utilizando a forma não linear do método das variáveis de controlo, no contexto de amostras de tamanho reduzido.

Nesta situação, para além dos resultados presentes na Tabela 3.1, é preciso também recolher os resultados correspondentes às variáveis de controlo (ver a Tabela 5.1).

Tabela 5.1: Disposição dos resultados para estimação de metamodelos, com variáveis de controlo.

Ponto experimental	Variáveis independentes	Variável dependente	Desvio padrão estimado
1	$X_{11} \dots X_{1d} \quad \mathbf{C}_{11} \dots \mathbf{C}_{1r}$	$Y_{11} \dots Y_{1r}$	$\hat{\sigma}_1$
\vdots	$\vdots \quad \quad \quad \vdots$	\vdots	\vdots
i	$X_{i1} \dots X_{id} \quad \mathbf{C}_{i1} \dots \mathbf{C}_{ir}$	$Y_{i1} \dots Y_{ir}$	$\hat{\sigma}_i$
\vdots	$\vdots \quad \quad \quad \vdots$	\vdots	\vdots
n	$X_{n1} \dots X_{nd} \quad \mathbf{C}_{n1} \dots \mathbf{C}_{nr}$	$Y_{n1} \dots Y_{nr}$	$\hat{\sigma}_n$

Seja \mathbf{Z} a seguinte matriz aleatória:

$$\mathbf{Z} = \begin{bmatrix} Y_{11} & C_{111} & \dots & C_{11q} \\ \vdots & \vdots & & \vdots \\ Y_{n1} & C_{n11} & \dots & C_{n1q} \\ \vdots & \vdots & & \vdots \\ Y_{1r} & C_{1r1} & \dots & C_{1rq} \\ \vdots & \vdots & & \vdots \\ Y_{nr} & C_{nr1} & \dots & C_{nrq} \end{bmatrix} = [\mathbf{Y}\mathbf{C}]. \quad (5.2)$$

Considere-se que, o vector linha aleatório \mathbf{Z}_l é contínuo e tem a mesma função de densidade de probabilidade para todos os valores de $l = 1, \dots, N$, tal que a sua matriz de variância-covariância existe e é a seguinte

$$\Sigma_Z = \mathcal{D}[\mathbf{Z}_l] = \begin{bmatrix} \sigma^2 & \boldsymbol{\sigma}_{YC} \\ \boldsymbol{\sigma}_{CY} & \Sigma_C \end{bmatrix}, \quad (5.3)$$

onde $\sigma^2 = \text{Var}[Y_{ij}]$. Como consequência $\Sigma_C = \mathcal{D}[\mathbf{C}_{ij}]$ é não singular e, por conseguinte, também é definida positiva, com probabilidade um, ao longo de todas as réplicas de todos os pontos experimentais [89]. O vector de covariâncias entre Y_{ij} e \mathbf{C}_{ij} , denotado por $\boldsymbol{\sigma}_{YC} = \mathcal{C}[Y_{ij}, \mathbf{C}_{ij}]$, assume-se ser constante para todo o $i = 1, \dots, n$ e todo o $j = 1, \dots, r$, com $\boldsymbol{\sigma}_{CY} = \boldsymbol{\sigma}_{YC}^T$.

No caso em que o modelo de simulação pode ser representado pelo metamodelo (3.3), onde os erros ϵ_{ij} são variáveis aleatórias independentes com distribuição normal, com médias nulas

e valores de variância dependentes do ponto experimental (isto é, $\epsilon_{ij} \sim N(0, \sigma_i^2)$ com $\sigma_i > 0$), considera-se o metamodelo

$$W_{ij} = g(\mathbf{X}_{i.}, \boldsymbol{\theta}) + e_{ij}$$

com $i = 1, \dots, n, j = 1, \dots, r, W_{ij} = Y_{ij}/\sigma_i, g(\mathbf{X}_{i.}, \boldsymbol{\theta}) = f(\mathbf{X}_{i.}, \boldsymbol{\theta})/\sigma_i$ e $e_{ij} = \epsilon_{ij}/\sigma_i$. Obtém-se assim, um metamodelo em que os erros já possuem variância constante e, por conseguinte, pode-se aplicar a metodologia desenvolvida neste capítulo. Outra hipótese é aplicar uma transformação de ambos os lados de (3.3), para estabilizar a variância, e depois aplicar também a metodologia aqui apresentada (ver a Secção 6.2.1).

Neste capítulo, apresentam-se dois conjuntos de resultados importantes:

(i) Na Secção 5.2, considera-se que o metamodelo (5.1) é válido e obtém-se a razão mínima aproximada entre variâncias generalizadas, o estimador dos mínimos quadrados $\hat{\delta}$ (para estimar $\boldsymbol{\delta}$), e o correspondente estimador dos mínimos quadrados controlado $\hat{\boldsymbol{\theta}}(\hat{\delta})$ (para estimar o vector de coeficientes de controlo $\boldsymbol{\theta}$).

(ii) Na Secção 5.3, considera-se a hipótese de que a resposta e as variáveis de controlo têm distribuição conjunta normal multivariada e obtém-se o factor de perda aproximado, constroem-se regiões de confiança assintóticas para $\boldsymbol{\theta}$, centradas em $\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})$ e em $\hat{\boldsymbol{\theta}}(\hat{\delta})$, e propõem-se procedimentos para realizar testes de hipóteses.

A secção 5.4 é reservada à descrição de uma meta-experiência de Monte-Carlo, em que se obtêm, experimentalmente, estimativas fiáveis para as razões entre variâncias e para o factor de perda.

5.2 Resultados gerais sobre estimação

Nesta secção, desenvolvem-se os resultados sobre a estimação de metamodelos com variáveis de controlo, que não dependem da normalidade conjunta entre a resposta e as variáveis de controlo.

5.2.1 Razão mínima entre variâncias generalizadas

O estimador dos mínimos quadrados de $\boldsymbol{\theta}$, no problema (3.3), é dado por

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T [\bar{\mathbf{Y}} - \mathbf{f}], \quad (5.4)$$

onde $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$ e $\mathbf{f} = \mathbf{f}(\boldsymbol{\theta}^*)$ (ver (3.5)). A média e a variância deste estimador satisfazem (ver (3.6)):

$$\mathcal{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}^*, \quad (5.5)$$

$$\mathcal{D}[\hat{\boldsymbol{\theta}}] = \frac{\sigma^2}{r} (\mathbf{F}^T \mathbf{F})^{-1}. \quad (5.6)$$

Quando se incluem as variáveis de controlo e para um determinado vector fixo de coeficientes de controlo $\boldsymbol{\phi}$, o estimador controlado dos mínimos quadrados é dado aproximadamente por

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T [\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi} - \mathbf{f}], \quad (5.7)$$

com

$$\bar{\mathbf{C}} = \begin{bmatrix} \bar{C}_{1.1} & \dots & \bar{C}_{1.q} \\ \vdots & & \vdots \\ \bar{C}_{n.1} & \dots & \bar{C}_{n.q} \end{bmatrix}$$

e onde, para $i = 1, 2, \dots, n$ e $k = 1, 2, \dots, q$,

$$\bar{C}_{i.k} = \frac{1}{r} \sum_{j=1}^r C_{ijk}.$$

Este estimador obtém-se do seguinte modo: considera-se o problema (5.1) escrito na forma

$$Y_{ij} - \mathbf{C}_{ij}\boldsymbol{\phi} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \varepsilon_{ij},$$

e, em seguida, calcula-se o estimador dos mínimos quadrados não lineares para $\boldsymbol{\theta}$ (considerando $\boldsymbol{\phi}$ fixo).

A aproximação (5.7) é equivalente a

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T [\bar{\mathbf{Y}} - \mathbf{f}] - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \bar{\mathbf{C}} \boldsymbol{\phi},$$

e sabe-se que, quando não se utilizam variáveis de controlo, o estimador dos mínimos quadrados usuais é dado por (3.5), logo

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \hat{\boldsymbol{\theta}} - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \bar{\mathbf{C}} \boldsymbol{\phi},$$

isto é, $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \neq \hat{\boldsymbol{\theta}}$, ou por outras palavras, a observação de variáveis de controlo com médias conhecidas resulta numa estimativa diferente para $\boldsymbol{\theta}$. Por conseguinte, se a matriz aleatória \mathbf{C} tem uma função densidade de probabilidade, então utilizando o facto de $\mathcal{E}[\bar{\mathbf{C}}] = \mathbf{0}$ e $\mathcal{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}^*$, obtém-se

$$\mathcal{E}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})] = \boldsymbol{\theta}^*. \quad (5.8)$$

Considere-se o vector aleatório $\bar{\mathbf{Z}}_{i.} = (\bar{Y}_{i.}, \bar{\mathbf{C}}_{i.})$ onde a matriz de variância-covariância a ele associada é

$$\boldsymbol{\Sigma}_{\bar{\mathbf{Z}}} = \mathcal{D}[\bar{\mathbf{Z}}_{i.}] = \begin{bmatrix} \sigma_{\bar{Y}}^2 & \sigma_{\bar{Y}\bar{\mathbf{C}}} \\ \sigma_{\bar{\mathbf{C}}\bar{Y}} & \boldsymbol{\Sigma}_{\bar{\mathbf{C}}} \end{bmatrix} = \frac{1}{r} \boldsymbol{\Sigma}_{\mathbf{Z}}, \quad (5.9)$$

com $\boldsymbol{\Sigma}_{\mathbf{Z}}$ é dada por (5.3).

Com o objectivo de obter a matriz de variância-covariância de $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})$, $\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})]$, é útil escrever o estimador (5.7) na forma

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \mathbf{A}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi}] + \mathbf{b},$$

onde $\mathbf{A} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ e $\mathbf{b} = \boldsymbol{\theta}^* - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{f}$. Deste modo podemos escrever

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})] = \mathbf{A} \mathcal{D}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi}] \mathbf{A}^T, \quad (5.10)$$

onde $\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi} = (\bar{Y}_{1.} - \bar{\mathbf{C}}_{1.}\boldsymbol{\phi}, \dots, \bar{Y}_{n.} - \bar{\mathbf{C}}_{n.}\boldsymbol{\phi})^T$. Para dois pontos experimentais diferentes, i e i' tem-se

$$\text{Cov}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi}, \bar{Y}_{i'.} - \bar{\mathbf{C}}_{i'.}\boldsymbol{\phi}] = 0, \quad i \neq i', \quad (5.11)$$

uma vez que estamos no contexto do método das réplicas independentes. Além disso, para $i = 1, \dots, n$,

$$\text{Var}[\bar{Y}_i - \bar{C}_i \phi] = \text{Var}[\bar{Y}_i] + \phi^T \mathcal{D}[\bar{C}_i] \phi - \phi^T \mathcal{C}[\bar{C}_i, \bar{Y}_i] - \mathcal{C}[\bar{Y}_i, \bar{C}_i] \phi,$$

ou seja

$$\text{Var}[\bar{Y}_i - \bar{C}_i \phi] = \sigma_Y^2 + \phi^T \Sigma_{\bar{C}} \phi - 2\phi^T \sigma_{\bar{C}\bar{Y}}. \quad (5.12)$$

O vector de coeficientes de controlo que minimiza esta variância é dado por

$$\delta = \Sigma_{\bar{C}}^{-1} \sigma_{\bar{C}\bar{Y}}$$

(ver (8) e (9) de [70]). Mas $\Sigma_{\bar{C}} = \Sigma_C/r$ e $\sigma_{\bar{C}\bar{Y}} = \sigma_{CY}/r$, logo

$$\delta = \Sigma_C^{-1} \sigma_{CY}. \quad (5.13)$$

Substituindo (5.13) em (5.12), e atendendo a que $\Sigma_{\bar{Z}} = \frac{1}{r} \Sigma_Z$, vem

$$\text{Var}[\bar{Y}_i - \bar{C}_i \delta] = \frac{1}{r} \tau^2, \quad (5.14)$$

onde

$$\tau^2 = \sigma^2 - \sigma_{YC} \Sigma_C^{-1} \sigma_{CY}. \quad (5.15)$$

(5.11), (5.14) e (5.15) implicam que

$$\mathcal{D}[\mathbf{Y} - \bar{\mathbf{C}}\delta] = \frac{1}{r} (\sigma^2 - \sigma_{YC} \Sigma_C^{-1} \sigma_{CY}) \mathbf{I}_n.$$

Substituindo esta matriz de variância-covariância em (5.10), e atendendo a que $\mathbf{A} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$, obtém-se

$$\begin{aligned} \mathcal{D}[\hat{\boldsymbol{\theta}}(\delta)] &= (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \frac{1}{r} (\sigma^2 - \sigma_{YC} \Sigma_C^{-1} \sigma_{CY}) \mathbf{I}_n \mathbf{F} (\mathbf{F}^T \mathbf{F})^{-1} \\ &= \frac{1}{r} (\sigma^2 - \sigma_{YC} \Sigma_C^{-1} \sigma_{CY}) (\mathbf{F}^T \mathbf{F})^{-1}. \end{aligned} \quad (5.16)$$

Utilizando (5.6) e (5.16), conclui-se que a redução máxima de variância que é possível atingir com variáveis de controlo, é dada aproximadamente pela razão mínima entre variâncias generalizadas

$$\eta(\delta) = \frac{|\mathcal{D}[\hat{\boldsymbol{\theta}}(\delta)]|}{|\mathcal{D}[\hat{\boldsymbol{\theta}}]} \approx 1 - \rho_{YC}^2, \quad (5.17)$$

onde

$$\rho_{YC}^2 = \frac{\boldsymbol{\sigma}_{YC} \boldsymbol{\Sigma}_C^{-1} \boldsymbol{\sigma}_{CY}}{\sigma^2}$$

é o coeficiente de correlação múltipla entre Y_{ij} e \mathbf{C}_{ij} , $i = 1, \dots, n$ e $j = 1, \dots, r$ (neste resultado utilizou-se o facto de $\rho_{\bar{Y}\bar{C}}^2 = \rho_{YC}^2$). Em geral não se conhece $\boldsymbol{\delta}$ ($\boldsymbol{\Sigma}_Z$ é desconhecida), logo tem de ser estimado e, como consequência, a variância irá aumentar. Ir-se-á utilizar o factor de perda para quantificar a percentagem de aumento da variância que ocorre quando $\boldsymbol{\delta}$ tem de ser estimado.

5.2.2 Estimador controlado dos mínimos quadrados

Com o objectivo de obter um estimador para o vector óptimo de coeficientes de controlo, $\hat{\boldsymbol{\delta}}$, juntamente com o estimador controlado correspondente $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, considera-se que (5.1) é um metamodelo válido e recorre-se ao método dos mínimos quadrados não lineares.

Proposição 5.1 Considere-se N suficientemente grande. Então, no metamodelo não linear (5.1), os estimadores dos mínimos quadrados para os parâmetros $\boldsymbol{\theta}$ e $\boldsymbol{\delta}$ são:

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T [\bar{\mathbf{Y}} - \mathbf{f} - \bar{\mathbf{C}} \hat{\boldsymbol{\delta}}] \quad (5.18)$$

e

$$\hat{\boldsymbol{\delta}} \approx (\bar{\mathbf{C}}^T \mathbf{P} \bar{\mathbf{C}})^{-1} \bar{\mathbf{C}}^T \mathbf{P} [\bar{\mathbf{Y}} - \mathbf{f}], \quad (5.19)$$

onde

$$\mathbf{P} = \mathbf{I}_n - \mathbf{F}(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T. \quad (5.20)$$

Verificação Para simplificar o problema de estimação, em vez de (5.1), considera-se o problema equivalente, no qual as observações individuais, em cada ponto experimental, são substituídas pelas suas médias:

$$\bar{Y}_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \bar{\mathbf{C}}_i \boldsymbol{\delta} + \bar{\varepsilon}_i, \quad i = 1, 2, \dots, n. \quad (5.21)$$

Considere-se $\boldsymbol{\theta}$ numa vizinhança de $\boldsymbol{\theta}^*$ (verdadeiro valor de $\boldsymbol{\theta}$). Então pode-se desenvolver $f_i = f(\mathbf{X}_i, \boldsymbol{\theta})$ em série de Taylor, em torno deste ponto, até à primeira ordem e vem

$$f_i(\boldsymbol{\theta}) \approx f_i(\boldsymbol{\theta}^*) + \sum_{j=1}^m \left. \frac{\partial f_i}{\partial \theta_j} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\theta_j - \theta_j^*),$$

ou seja,

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*),$$

com $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$. Utilizando este resultado, pode-se reescrever (5.21) na forma

$$\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*) \approx \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \bar{\mathbf{C}}\boldsymbol{\delta} + \bar{\boldsymbol{\varepsilon}}. \quad (5.22)$$

Aplicando [105, página 342] ao problema linearizado anterior, obtém-se

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^* \approx (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T [\bar{\mathbf{Y}} - \mathbf{f}] - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \bar{\mathbf{C}} \hat{\boldsymbol{\delta}} \quad (5.23)$$

onde $\hat{\boldsymbol{\delta}}$ tem a forma (5.19). A aproximação (5.18) resulta directamente de (5.23). \diamond

5.3 Resultados para metamodelos normais

Com o objectivo de simplificar a apresentação dos resultados, quando a resposta e as variáveis de controlo têm distribuição normal conjunta, introduz-se em seguida alguma notação e hipóteses adicionais. Considerem-se os metamodelos (3.3) e (5.1). Suponha-se que ao longo do l -ésimo ponto experimental

$$\mathbf{Z}_l \sim N_{q+1} \left((\boldsymbol{\mu}_l, \mathbf{0}^T), \boldsymbol{\Sigma}_Z \right), \quad (5.24)$$

onde $\boldsymbol{\Sigma}_Z$ é dada por (5.3). Logo, a matriz aleatória $\mathbf{N} \times (q+1)$, definida por (5.2), tem distribuição normal multivariada

$$\mathbf{Z} \sim N_{\mathbf{N}, q+1} (\boldsymbol{\mu}_Z, \boldsymbol{\Xi}, \boldsymbol{\Sigma}), \quad (5.25)$$

(onde cada elemento de \mathbf{Z} tem distribuição normal), com $\boldsymbol{\Sigma}_Z$ desconhecida e

$$\mathcal{E}[\mathbf{Z}] = \boldsymbol{\mu}_Z = (\boldsymbol{\mu}_Y, \mathbf{0}),$$

onde

$$\boldsymbol{\mu}_Y = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{n.}, \boldsymbol{\theta}))^T.$$

A matriz de variância-covariância, para cada linha, é

$$\mathcal{D}[\mathbf{Z}_{l.}] = \boldsymbol{\Sigma}_Z, \quad l = 1, \dots, N,$$

e a matriz de variância-covariância, para cada coluna, é

$$\mathcal{D}[\mathbf{Z}_{.s}] = \boldsymbol{\Xi}, \quad s = 1, \dots, q + 1.$$

Suponha-se também que $\boldsymbol{\Xi}$ e $\boldsymbol{\Sigma}_Z$ são definidas positivas. Além disso, as linhas de \mathbf{Z} são mutuamente independentes, uma vez que correspondem a execuções independentes do programa de simulação. Como resultado, considera-se $\boldsymbol{\Xi} = \mathbf{I}_N$ no desenvolvimento que se segue.

Se a matriz $q \times q$, $\boldsymbol{\Sigma}_C$, for definida positiva, então a distribuição condicional de \mathbf{Y} dado \mathbf{C} é [5, Teorema 17.2 g), página 312]

$$\mathbf{Y}|\mathbf{C} \sim N_N(\boldsymbol{\mu}_{Y.C}, \tau^2 \mathbf{I}_N) \quad (5.26)$$

com τ^2 dado por (5.15) e

$$\boldsymbol{\mu}_{Y.C} = \boldsymbol{\mu}_Y + \mathbf{C}\boldsymbol{\Sigma}_C^{-1}\boldsymbol{\sigma}_{CY}. \quad (5.27)$$

Por conseguinte, condicionando em \mathbf{C} , observa-se que o metamodelo correcto é (5.1). Ir-se-á ver que, os estimadores assintóticos dos mínimos quadrados $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ e $\hat{\boldsymbol{\delta}}$ são não enviesados, tanto condicionalmente, como incondicionalmente. Além disso, pode-se construir uma região de confiança aproximada, para o verdadeiro vector de parâmetros, centrada em $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$.

5.3.1 Distribuição do estimador controlado

- Caso em que $\boldsymbol{\Sigma}_Z$ é conhecida

Se $\boldsymbol{\Sigma}_Z$ é conhecida, então o estimador dos mínimos quadrados para $\boldsymbol{\theta}$ é dado aproximadamente por

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T [\bar{\mathbf{Y}} - \mathbf{f} - \bar{\mathbf{C}}\boldsymbol{\delta}], \quad (5.28)$$

onde $\boldsymbol{\delta}$ é o vector óptimo dos coeficientes de controlo (5.13). Observa-se que é possível escrever $\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})$ da seguinte maneira:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \approx \mathbf{A}\bar{\mathbf{Y}} + \mathbf{b},$$

onde $\mathbf{A} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ é uma matriz $m \times n$ e $\mathbf{b} = \boldsymbol{\theta}^* - \mathbf{A}[\mathbf{f} + \bar{\mathbf{C}}\boldsymbol{\delta}]$. Utilizando [5, Teorema 17.2 d), página 312] pode-se afirmar que assintoticamente,

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \sim N_{m,1}(\mathbf{A}\boldsymbol{\mu}_{Y.C} + \mathbf{b}, \mathbf{A}\mathbf{I}_n\mathbf{A}^T, \bar{\tau}^2),$$

onde $\bar{\tau}^2 = \tau^2/r$, com τ^2 dado por (5.15). Mas

$$\mathbf{A}\mathbf{I}_n\mathbf{A}^T = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{F} (\mathbf{F}^T \mathbf{F})^{-1} = (\mathbf{F}^T \mathbf{F})^{-1}$$

e, por conseguinte, tem-se aproximadamente

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \sim N_{m,1}\left(\boldsymbol{\theta}^*, (\mathbf{F}^T \mathbf{F})^{-1}, \frac{\tau^2}{r}\right). \quad (5.29)$$

Como consequência, conclui-se que se a hipótese de normalidade multivariada (5.25) se verificar, então o metamodelo não linear com inclusão de variáveis de controlo é verdadeiro. Observa-se que, uma vez que τ^2/r é um escalar [5, Teorema 17.2 a), página 312], (5.29) é equivalente a

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \sim N_m\left(\boldsymbol{\theta}^*, \frac{1}{r}\tau^2(\mathbf{F}^T \mathbf{F})^{-1}\right), \quad (5.30)$$

ou seja, a distribuição do estimador $\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})$ é aproximadamente normal.

• Caso em que Σ_Z é desconhecida

Se $\mathbf{Z} \sim N_{N,q+1}(\boldsymbol{\mu}_Z, \mathbf{I}_N, \Sigma_Z)$ com $\boldsymbol{\theta}$ e Σ_Z desconhecidos, então tiramos partido do facto de a distribuição condicional de \mathbf{Y} dado \mathbf{C} ser normal e dada por (5.26). Uma vez que isto se verifica, condicionando em \mathbf{C} , vemos que o metamodelo

$$\mathbf{Y} = \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) + \mathbf{C}\boldsymbol{\delta} + \mathbf{e} \quad (5.31)$$

é de facto o modelo não linear correcto para um vector resposta normal, onde

$$\tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) = (f(\mathbf{X}_{1.}), \dots, f(\mathbf{X}_{1.}), \dots, f(\mathbf{X}_{n.}), \dots, f(\mathbf{X}_{n.}))^T$$

(vector com N componentes) e

$$\mathbf{e} = (\varepsilon_{11}, \dots, \varepsilon_{1r}, \dots, \varepsilon_{n1}, \dots, \varepsilon_{nr})^T.$$

Utilizando o desenvolvimento em série de Taylor, em torno de $(\mathbf{X}, \boldsymbol{\theta}^*)$, até à primeira ordem, obtém-se

$$\tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) \approx \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*) + \tilde{\mathbf{F}}(\boldsymbol{\theta} - \boldsymbol{\theta}^*), \quad (5.32)$$

onde $\tilde{\mathbf{F}}$ é a matriz Jacobiana de $\tilde{\mathbf{f}}$, calculada em $(\mathbf{X}, \boldsymbol{\theta}^*)$. Por conseguinte, (5.31) pode ser reescrito como

$$\mathbf{G} \approx \tilde{\mathbf{F}}\boldsymbol{\lambda} + \mathbf{C}\boldsymbol{\delta} + \mathbf{e} \quad (5.33)$$

onde $\mathbf{G} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$ e $\boldsymbol{\lambda} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$. Aplicando [105, página 342] ao problema (5.33), obtém-se

$$\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) \approx (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T \mathbf{G} - (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T \mathbf{C} \hat{\boldsymbol{\delta}} \quad (5.34)$$

onde

$$\hat{\boldsymbol{\delta}} = (\mathbf{C}^T \tilde{\mathbf{P}} \mathbf{C})^{-1} \mathbf{C}^T \tilde{\mathbf{P}} \mathbf{G} \quad (5.35)$$

com

$$\tilde{\mathbf{P}} = \mathbf{I}_N - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T.$$

Substituindo (5.35) em (5.34), vem

$$\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) \approx (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T [\mathbf{I}_N - \mathbf{C}(\mathbf{C}^T \tilde{\mathbf{P}} \mathbf{C})^{-1} \mathbf{C}^T \tilde{\mathbf{P}}] \mathbf{G},$$

ou seja,

$$\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) \approx \mathbf{B} \mathbf{G} \quad (5.36)$$

onde

$$\mathbf{B} = (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T [\mathbf{I}_N - \mathbf{C}(\mathbf{C}^T \tilde{\mathbf{P}} \mathbf{C})^{-1} \mathbf{C}^T \tilde{\mathbf{P}}].$$

Uma vez que (5.26) se verifica, tem-se aproximadamente

$$\mathbf{G} | \mathbf{C} \sim N_N(\boldsymbol{\mu}_{G.C}, \tau^2 \mathbf{I}_N)$$

onde $\boldsymbol{\mu}_{G.C} = \boldsymbol{\mu}_{Y.C} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$. Utilizando este resultado, juntamente com (5.36), podemos aplicar [5, Teorema 17.2 d)] e vem

$$\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) | \mathbf{C} \sim N_{m,1}(\mathbf{B} \boldsymbol{\mu}_{G.C}, \mathbf{B} \mathbf{B}^T, \tau^2 \mathbf{I}_N),$$

ou seja,

$$\hat{\lambda}(\hat{\delta})|\mathbf{C} \sim N_m(\mathbf{B}\boldsymbol{\mu}_{G,C}, \tau^2\mathbf{B}\mathbf{B}^T).$$

Suponha-se que $\tilde{\mathbf{F}}$ é uma matriz de rank m , então $\tilde{\mathbf{P}}$ é a projecção ortogonal de \mathbb{R}^N em $\mathcal{R}(\tilde{\mathbf{F}})^\perp$ [107]. Uma vez que $\tilde{\mathbf{P}}$ é uma projecção ortogonal ($\tilde{\mathbf{P}} = \tilde{\mathbf{P}}^T$ e $\tilde{\mathbf{P}}^2 = \tilde{\mathbf{P}}$) e

$$(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\tilde{\mathbf{P}} = \tilde{\mathbf{P}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1} = 0$$

(($\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\tilde{\mathbf{P}} = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T(\mathbf{I}_N - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T) = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T - (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T = 0$ e, de forma análoga, $\tilde{\mathbf{P}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1} = 0$), obtém-se

$$\mathbf{B}\mathbf{B}^T = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}[\mathbf{I}_m + \tilde{\mathbf{F}}^T\mathbf{C}(\mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^T\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}].$$

Aplicando [91, página 121] ao problema linearizado (5.33), tem-se

$$\mathcal{E}[\hat{\lambda}(\hat{\delta})|\mathbf{C}] = \boldsymbol{\lambda}.$$

Por conseguinte, $\mathbf{B}\boldsymbol{\mu}_{G,C} = \boldsymbol{\lambda}$ e vem

$$\hat{\lambda}(\hat{\delta})|\mathbf{C} \sim N_m(\boldsymbol{\lambda}, \tau^2\mathbf{B}\mathbf{B}^T).$$

Mas $\hat{\lambda}(\hat{\delta}) = \hat{\boldsymbol{\theta}}(\hat{\delta}) - \boldsymbol{\theta}^*$ e $\boldsymbol{\lambda} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$, logo $\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\delta})|\mathbf{C}] = \boldsymbol{\theta}$ (dado \mathbf{C} , $\hat{\boldsymbol{\theta}}(\hat{\delta})$ é um estimador não enviesado de $\boldsymbol{\theta}$) e $\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\delta})|\mathbf{C}] = \tau^2\mathbf{B}\mathbf{B}^T$ e tem-se

$$\hat{\boldsymbol{\theta}}(\hat{\delta})|\mathbf{C} \sim N_m(\boldsymbol{\theta}, \tau^2\mathbf{B}\mathbf{B}^T). \quad (5.37)$$

Resta agora apresentar um estimador para τ^2 . Para isso, considere-se a seguinte proposição:

Proposição 5.2 Dado \mathbf{C} , os resíduos estimados no modelo não linear (5.31), $\hat{\mathbf{e}}$, satisfazem aproximadamente

$$\hat{\mathbf{e}}^T\hat{\mathbf{e}}|\mathbf{C} \sim \tau^2\chi_{N-m-q}^2$$

onde

$$\hat{\mathbf{e}} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \hat{\boldsymbol{\theta}}) - \mathbf{C}\hat{\delta}$$

é dado por (5.15).

Verificação Aplicando [88, Teorema 3.3] ao problema (5.33), vem

$$\hat{\mathbf{e}}^T \hat{\mathbf{e}} | \mathbf{C} \sim W_1(N - m - q, \tau^2, 0),$$

ou seja (ver [5, Teorema 17.6 b])

$$\hat{\mathbf{e}}^T \hat{\mathbf{e}} | \mathbf{C} \sim \tau^2 \chi_{N-m-q}^2,$$

onde $\hat{\mathbf{e}} = \mathbf{G} - \tilde{\mathbf{F}} \hat{\boldsymbol{\lambda}} - \mathbf{C} \hat{\boldsymbol{\delta}}$. Mas $\mathbf{G} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$, $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ e a aproximação (5.32) verifica-se com $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, logo $\hat{\mathbf{e}} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \hat{\boldsymbol{\theta}}) - \mathbf{C} \hat{\boldsymbol{\delta}}$. \diamond

Proposição 5.3 Dado \mathbf{C} , um estimador aproximado não enviesado para τ^2 é

$$\hat{\tau}^2 | \mathbf{C} = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{N - m - q} = \frac{1}{N - m - q} \sum_{i=1}^n \sum_{j=1}^r [Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) - \mathbf{C}_{ij} \hat{\boldsymbol{\delta}}]^2, \quad (5.38)$$

e, além disso

$$\hat{\tau}^2 | \mathbf{C} \sim (N - m - q)^{-1} \tau^2 \chi_{N-m-q}^2. \quad (5.39)$$

Verificação O resultado (5.39) sai directamente da proposição anterior. O não enviesamento resulta de uma propriedade da distribuição χ^2 : $E[\tau^2 | \mathbf{C}] = E[\hat{\mathbf{e}}^T \hat{\mathbf{e}}] / (N - m - q) = (N - m - q) \tau^2 / (N - m - q) = \tau^2$. \diamond

5.3.2 Razão entre variâncias e factor de perda

Quando Σ_Z é conhecida, então a razão entre variâncias generalizadas é dada pela razão mínima entre essas variâncias (5.17). Sendo assim, agora apenas se vai quantificar a eficiência do estimador controlado no caso em que Σ_Z é desconhecida (como se sabe, esse estimador é denominado por $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$).

Tal como nos trabalhos de Venkatraman e Wilson [116] e Porta Nova e Wilson [89], ao contrário do que é utilizado por Nozari, Arnold e Pegden [92] e Rubinstein e Marcus [99], considera-se que a generalização adequada das medidas de desempenho (razão entre variâncias e factor de perda), introduzidas por Lavenberg, Moeller e Welch [69], se baseia na matriz de dispersão não

condicionada dos coeficientes controlados do metamodelo de simulação (5.1). Assim, vamos em seguida encontrar a variância não condicionada de $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$.

Uma vez que (5.37) se verifica, tem-se

$$\begin{aligned}\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] &= \boldsymbol{\theta} \\ \mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] &= \tau^2 \mathbf{B}\mathbf{B}^T.\end{aligned}$$

Logo

$$\begin{aligned}\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})] &= \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] + \mathcal{D}[\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \\ &= \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] + \mathcal{D}[\boldsymbol{\theta}] \\ &= \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]]\end{aligned}\tag{5.40}$$

e

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] \approx \tau^2 (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} [\mathbf{I}_m + \tilde{\mathbf{F}}^T \mathbf{C} (\mathbf{C}^T \tilde{\mathbf{P}} \mathbf{C})^{-1} \mathbf{C}^T \tilde{\mathbf{F}} (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1}].$$

Seja $\mathbf{U} = (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T \mathbf{C}$ e $\mathbf{V} = \mathbf{C}^T \tilde{\mathbf{P}} \mathbf{C}$. Então pode-se escrever

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] \approx \tau^2 (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} + \tau^2 \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T.\tag{5.41}$$

Uma vez que $\mathbf{C} \sim N_{N,q}(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\Sigma}_C)$ e $\tilde{\mathbf{P}}$ é uma projecção ortogonal de \mathbb{R}^N num espaço de dimensão $N - m$ ($\mathcal{R}(\mathbf{F})^\perp$), aplica-se [5, Teorema 17.7] para obter

$$\mathbf{V} = \mathbf{C}^T \tilde{\mathbf{P}} \mathbf{C} \sim W_q(N - m, \boldsymbol{\Sigma}_C).$$

Utilizando o facto de $(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^T \tilde{\mathbf{P}} = \mathbf{0}$ e uma vez que $\tilde{\mathbf{P}}$ é definida positiva, aplica-se [5, Teorema 17.7 b.2.] e pode-se garantir que \mathbf{U} e \mathbf{V} são independentes. Tomando o valor esperado em (5.41) obtém-se

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2 (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} + \tau^2 \mathcal{E}[\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T]$$

Utilizando os resultados presentes em [92, Apêndice] (com \mathbf{X} e n substituídos por \mathbf{F} e N , respectivamente) obtém-se

$$\mathcal{E}[\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T] = \frac{q}{N - m - q - 1} (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1}.$$

Combinando os dois resultados anteriores, obtém-se

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2 \frac{N - m - 1}{N - m - q - 1} (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1}.$$

Mas $\tilde{\mathbf{F}}^T = [\mathbf{F}^T \dots \mathbf{F}^T]$, logo $(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} = \mathbf{F}^T \mathbf{F} / r$ e, por conseguinte,

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2 \frac{\mathbf{N} - \mathbf{m} - 1}{r(\mathbf{N} - \mathbf{m} - \mathbf{q} - 1)} (\mathbf{F}^T \mathbf{F})^{-1},$$

ou seja (ver (5.40)), tem-se o seguinte resultado assintótico

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})] \approx \tau^2 (\mathbf{F}^T \mathbf{F})^{-1} \frac{\mathbf{N} - \mathbf{m} - 1}{r(\mathbf{N} - \mathbf{m} - \mathbf{q} - 1)}.$$

Esta aproximação, juntamente com (5.6) e (5.15), permite obter a seguinte razão aproximada entre variâncias generalizadas:

$$\eta(\hat{\boldsymbol{\delta}}) = \frac{|\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]|}{|\mathcal{D}[\hat{\boldsymbol{\theta}}]|} \approx \frac{\mathbf{N} - \mathbf{m} - 1}{\mathbf{N} - \mathbf{m} - \mathbf{q} - 1} \frac{\tau^2}{\sigma^2} = \frac{\mathbf{N} - \mathbf{m} - 1}{\mathbf{N} - \mathbf{m} - \mathbf{q} - 1} (1 - \rho_{YC}^2). \quad (5.42)$$

Comparando a razão aproximada anterior com a razão mínima entre variâncias (5.17), observa-se uma degradação da redução máxima de variância denominada por factor de perda:

$$\text{FP}(\hat{\boldsymbol{\delta}}) = \frac{\mathbf{N} - \mathbf{m} - 1}{\mathbf{N} - \mathbf{m} - \mathbf{q} - 1}. \quad (5.43)$$

5.3.3 Regiões de confiança

- Caso em que Σ_Z é conhecida

Neste caso verifica-se (5.30), logo

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}^* \sim N_m \left(\mathbf{0}, \frac{1}{r} \tau^2 (\mathbf{F}^T \mathbf{F})^{-1} \right), \quad (5.44)$$

com τ^2 dado por (5.15). Por conseguinte [3, Teorema 3.3.3, página 54],

$$\frac{(\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}^*)^T r \mathbf{F}^T \mathbf{F} (\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}^*)}{\tau^2} \sim \chi_m^2. \quad (5.45)$$

Utilizando este resultado, é possível apresentar a seguinte região de confiança assintótica, com probabilidade de cobertura igual a $1 - \alpha$:

$$\left\{ \boldsymbol{\theta}^* : \frac{(\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}^*)^T r \mathbf{F}^T \mathbf{F} (\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}^*)}{\tau^2} \leq \chi_{m;\alpha}^2 \right\}. \quad (5.46)$$

Na prática, é conveniente apresentar um rectângulo de confiança para $\boldsymbol{\theta}$, constituído por intervalos de confiança simultâneos para θ_k , $k = 1, \dots, m$. Uma região deste tipo é mais fácil de visualizar e interpretar que uma região como (5.46). Uma vez que (5.30) se verifica, pode-se escrever

$$\mathbf{a}^T \hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \sim N \left(\mathbf{a}^T \boldsymbol{\theta}^*, \frac{\tau^2}{r} \mathbf{a}^T [\mathbf{F}^T \mathbf{F}]^{-1} \mathbf{a} \right)$$

onde \mathbf{a} é fixo e pertence a \mathbb{R}^m . Por conseguinte, um intervalo de confiança aproximado para $\mathbf{a}^T \boldsymbol{\theta}^*$, com probabilidade de cobertura igual a $(1 - \alpha)$, é dado por

$$\mathbf{a}^T \hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) \pm z_{\alpha/2} \left(\frac{\tau^2}{r} \mathbf{a}^T [\mathbf{F}^T \mathbf{F}]^{-1} \mathbf{a} \right)^{1/2}.$$

Se $\mathbf{a}^T = (0, \dots, 0, 1, 0, \dots, 0)$, onde o k -ésimo elemento de \mathbf{a} é igual a um e os restantes elementos iguais a zero, então um intervalo de confiança para $\boldsymbol{\theta}_k^*$ é o seguinte:

$$\hat{\boldsymbol{\theta}}_k(\boldsymbol{\delta}) \pm z_{\alpha/2} \frac{\tau}{r^{1/2}} \left[\hat{\mathbf{F}}^T \hat{\mathbf{F}} \right]_{kk}^{-1/2}.$$

Utilizando o método de Bonferroni obtém-se o rectângulo de confiança com nível de confiança aproximado de pelo menos $100(1 - \alpha)\%$, para p intervalos simultâneos ($1 \leq p \leq m$):

$$\hat{\boldsymbol{\theta}}_k(\boldsymbol{\delta}) \pm z_{\alpha/2p} \frac{\tau}{r^{1/2}} \left[\hat{\mathbf{F}}^T \hat{\mathbf{F}} \right]_{kk}^{-1/2} \quad k = 1, 2, \dots, p.$$

• Caso em que Σ_Z é desconhecida

Uma vez que se verifica (5.37), então utilizando [5, Teorema 3.10], pode-se garantir que condicionando em \mathbf{C} :

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)}{\tau^2} | \mathbf{C} \sim \chi_m^2 \quad (5.47)$$

Por outro lado, (5.39) pode ser reescrito como

$$(N - m - q) \frac{\hat{\tau}^2}{\tau^2} | \mathbf{C} \sim \chi_{N-m-q}^2. \quad (5.48)$$

Combinando (5.47) e (5.48) obtém-se

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)}{m \hat{\tau}^2} | \mathbf{C} \sim F_{m, N-m-q}.$$

Dado \mathbf{C} , uma região de confiança assintótica para $\boldsymbol{\theta}$, com probabilidade de cobertura condicional de pelo menos $1 - \alpha$, é dada por

$$\left\{ \boldsymbol{\theta}^* : \frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)}{m\hat{\tau}^2} \leq F_{m, N-m-q; \alpha} \right\}. \quad (5.49)$$

Uma vez que esta região de confiança tem cobertura condicional de pelo menos $1 - \alpha$, também tem cobertura não condicional de pelo menos $1 - \alpha$.

Seja $\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})$ a i -ésima componente do vector $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ e seja $\hat{\tau}^2[\mathbf{B}\mathbf{B}^T]_{ii}$ o estimador da variância correspondente (i -ésimo elemento da diagonal principal da matriz $\hat{\tau}^2\mathbf{B}\mathbf{B}^T$). Uma vez que $\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})$ é condicionalmente independente de $\hat{\tau}^2[\mathbf{B}\mathbf{B}^T]_{ii}$ dado \mathbf{C} , os resultados (5.37) e (5.39) implicam que

$$\frac{\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_i^*}{\hat{\tau} \sqrt{[\mathbf{B}\mathbf{B}^T]_{ii}}} | \mathbf{C} \sim t_{N-m-q},$$

onde t_{N-m-q} representa a distribuição t de Student com $N - m - q$ graus de liberdade. Como consequência e utilizando o método de Bonferroni, um rectângulo de confiança para $\boldsymbol{\theta}$, com probabilidade de cobertura condicional de pelo menos $1 - \alpha$, tem a forma

$$\hat{\boldsymbol{\theta}}_k(\hat{\boldsymbol{\delta}}) \pm t_{N-m-q; 1-\alpha/(2p)} \hat{\tau} [\mathbf{B}\mathbf{B}^T]_{kk}^{1/2}, \quad k = 1, \dots, p, \quad (5.50)$$

onde $1 \leq p \leq m$. Esta região de confiança tem probabilidade de cobertura condicional de pelo menos $1 - \alpha$, logo também tem probabilidade de cobertura não condicional de pelo menos $1 - \alpha$.

5.3.4 Testes de hipóteses

- Caso em que Σ_Z é conhecida

Suponha-se que se pretende testar a hipótese $H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. Uma vez que se verifica (5.44), se H_0 for verdadeira, isto é $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$, então é possível escrever

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0 \sim N_m \left(\mathbf{0}, \frac{\tau^2}{r} (\mathbf{F}^T \mathbf{F})^{-1} \right).$$

Por conseguinte, se H_0 for verdadeira, recorrendo a [3, Teorema 3.3.3, página 54] obtém-se

$$\frac{(\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}_0)^T \mathbf{r} \mathbf{F}^T \mathbf{F} (\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}_0)}{\tau^2} \sim \chi_m^2.$$

Como consequência, rejeita-se H_0 , com nível de significância α se

$$\frac{(\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}_0)^T \mathbf{r} \mathbf{F}^T \mathbf{F} (\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}) - \boldsymbol{\theta}_0)}{\tau^2} > \chi_{m;\alpha}^2. \quad (5.51)$$

Sempre que necessário substitui-se \mathbf{F} por $\hat{\mathbf{F}} = \mathbf{F}(\hat{\boldsymbol{\theta}})$.

• Caso em que Σ_Z é desconhecida

Para testar a hipótese $H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, procede-se como é descrito em seguida.

Condicionando em \mathbf{C} e uma vez que se verifica (5.37), utiliza-se [5, Teorema 3.10] e obtém-se

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)}{\tau^2} \sim \chi_m^2, \quad (5.52)$$

se H_0 é verdadeira. Combinando (5.52) e (5.48) tem-se, condicionando em \mathbf{C} ,

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)}{m\hat{\tau}^2} \sim F_{m, N-m-q}$$

onde $\hat{\tau}^2$ é dado por (5.38). Como resultado, rejeita-se H_0 , com nível de significância α , se

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)}{m\hat{\tau}^2} > F_{m, N-m-q; \alpha}. \quad (5.53)$$

5.4 Razão entre variâncias e factor de perda experimentais

Com o objectivo de estimar a razão entre variâncias e o factor de perda, para um determinado metamodelo, elabora-se uma meta-experiência com \mathbf{K} réplicas da experiência inicial. Para cada réplica k ($k = 1, \dots, \mathbf{K}$) da experiência inicial, calcula-se o estimador dos mínimos quadrados assintótico sem inclusão de variáveis de controlo, $\hat{\boldsymbol{\theta}}^{(k)}$, e o estimador assintótico controlado dos mínimos quadrados, $\hat{\boldsymbol{\theta}}^{(k)}(\hat{\boldsymbol{\delta}})$. Estes valores permitem obter estimativas não enviesadas das

variâncias de $\hat{\boldsymbol{\theta}}^{(k)}$ e $\hat{\boldsymbol{\theta}}^{(k)}(\hat{\boldsymbol{\delta}})$. Por exemplo, a partir da amostra aleatória $(\hat{\boldsymbol{\theta}}^{(1)}, \dots, \hat{\boldsymbol{\theta}}^{(K)})$, calcula-se o seguinte estimador não enviesado de $\mathcal{D}[\hat{\boldsymbol{\theta}}]$:

$$\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}] = \frac{1}{K} \sum_{k=1}^K \left(\hat{\boldsymbol{\theta}}^{(k)} - \bar{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}^{(k)} - \bar{\boldsymbol{\theta}} \right)^T,$$

onde

$$\bar{\boldsymbol{\theta}} = \frac{1}{K} \sum_{k=1}^K \hat{\boldsymbol{\theta}}^{(k)}.$$

Considerando os estimadores de variância calculados como foi explicado, obtém-se um estimador para a razão assintótica entre variâncias (5.42), denominado razão observada entre variâncias:

$$\hat{\eta}(\hat{\boldsymbol{\delta}}) = \frac{|\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]|}{|\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}]} \quad (5.54)$$

Para estimar a razão mínima entre variâncias (5.17), é preciso calcular estimadores não enviesados de σ_Y^2 , σ_{YC} e Σ_C , em cada ponto experimental. Por exemplo, se $\bar{Y}_i^{(k)}$ denotar a resposta média observada no i -ésimo ponto experimental, correspondente à réplica k , então uma estimativa para a variância σ_Y^2 , considerando os dados relativos ao i -ésimo ponto experimental, é:

$$\hat{\sigma}_{\bar{Y}_i}^2 = \frac{1}{K} \sum_{k=1}^K \left(\bar{Y}_i^{(k)} - \bar{\bar{Y}}_i \right)^2,$$

onde

$$\bar{\bar{Y}}_i = \frac{1}{K} \sum_{k=1}^K \bar{Y}_i^{(k)}. \quad (5.55)$$

Um estimador de σ_Y^2 , baseado em todos os n pontos experimentais, é obtido calculando

$$\bar{\sigma}_Y^2 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{\bar{Y}_i}^2.$$

Para calcular $\bar{\Sigma}_C$, seja $\bar{\mathbf{C}}_i^{(k)}$ o vector média de controlo correspondente ao i -ésimo ponto experimental, obtido na réplica k , então a matriz de variância-covariância de $\bar{\mathbf{C}}_i$ é estimada por

$$\hat{\Sigma}_C^i = \frac{1}{K} \sum_{k=1}^K \left(\bar{\mathbf{C}}_i^{(k)} - \bar{\bar{\mathbf{C}}}_i \right)^T \left(\bar{\mathbf{C}}_i^{(k)} - \bar{\bar{\mathbf{C}}}_i \right),$$

onde

$$\bar{\bar{\mathbf{C}}}_i = \frac{1}{K} \sum_{k=1}^K \bar{\mathbf{C}}_i^{(k)}.$$

O estimador correspondente, baseado nos n pontos experimentais é dado por

$$\bar{\Sigma}_C = \frac{1}{n} \sum_{i=1}^n \hat{\Sigma}_C^i.$$

De modo semelhante, obtém-se o estimador para o vector de covariâncias σ_{YC} :

$$\bar{\sigma}_{YC} = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{YC}^i,$$

com

$$\hat{\sigma}_{YC}^i = \frac{1}{K} \sum_{k=1}^K (\bar{Y}_{i.}^{(k)} - \bar{Y}_{i.}) (\bar{C}_{i.}^{(k)} - \bar{C}_{i.}).$$

Utilizando estes valores numéricos, calcula-se o seguinte estimador de (5.17), denominado razão mínima estimada entre variâncias:

$$\hat{\eta}(\delta) = 1 - \frac{\bar{\sigma}_{YC} \bar{\Sigma}_C^{-1} \bar{\sigma}_{YC}^T}{\hat{\sigma}_Y^2}. \quad (5.56)$$

Multiplicando (5.56) pelo factor de perda (5.43), obtém-se a razão prevista entre variâncias

$$\hat{\eta} = \hat{\eta}(\delta) \text{FP}(\hat{\delta}),$$

que pode ser comparada com a razão observada entre variâncias (5.54). Por último, o factor de perda observado

$$\widehat{\text{FP}}(\hat{\delta}) = \frac{\hat{\eta}(\hat{\delta})}{\hat{\eta}(\delta)}$$

pode ser comparado com o factor de perda teórico (5.43).

Capítulo 6

Aplicação

O objectivo principal deste capítulo é, através de exemplos de aplicação, ilustrar a aplicação da metodologia apresentada anteriormente. Quando se desenvolve um metamodelo, o primeiro impulso é escolhê-lo dentro da classe dos metamodelos lineares de regressão. Isto deve-se ao facto deste tipo de metamodelos ser simples de ajustar e manipular. No exemplo tratado na Secção 6.1, experimentou-se considerar, para metamodelos hipotéticos, diversos polinómios e obtiveram-se metamodelos não válidos. Concluiu-se que, neste caso, os metamodelos polinomiais não são boas aproximações para o modelo de simulação e, por último, para o próprio problema real. Isto é um problema comum que surge quando se desenvolvem metamodelos lineares em situações em que a resposta e as variáveis de decisão se relacionam de uma maneira intrinsecamente não linear. Acontece que, em problemas reais, essa é a situação mais comum. Por exemplo, em sistemas de filas de espera $M/M/s$, a taxa de chegadas, a taxa de serviço e o número de servidores relacionam-se de uma forma não linear. Existem outros exemplos análogos, entre os quais os dois problemas tratados neste capítulo: no primeiro, existe uma relação não linear entre o tempo médio de permanência no sistema e o tempo médio entre chegadas; e no segundo, a relação não linear é entre o tempo médio de espera na fila e o factor de utilização. Em ambos os problemas, conseguiram-se bons resultados construindo metamodelos não lineares.

Um inconveniente na construção de um metamodelo não linear é que, essa construção, é computacionalmente mais complicada e demorada que a construção de um metamodelo linear. Con-

tudo, este problema torna-se pouco relevante porque os computadores de hoje são rápidos, têm uma capacidade de cálculo elevada e existem pacotes de programas contendo boas realizações de métodos numéricos para minimizar funções não lineares (por exemplo, o método de Levenberg-Marquardt).

Existem metamodelos que são não lineares mas são linearizáveis, ou seja, podem ser transformados em modelos lineares. Por vezes, acontece que em vez de aplicar directamente a metodologia para metamodelos não lineares, lineariza-se o metamodelo hipotético e, em seguida, aplica-se a metodologia para metamodelos lineares (ver as Secções 6.1 e 6.2). Mesmo quando o metamodelo linearizado funciona bem, o metamodelo não linear correspondente pode, e por vezes deve, ser utilizado para conservar uma interpretação clara dos parâmetros (note-se que os parâmetros de um metamodelo linearizado não são, em geral, tão importantes como os parâmetros originais [107]).

Existem três razões para querer utilizar uma transformação de resultados [17]:

- (i) conseguir linearidade;
- (ii) obter erros com distribuição aproximadamente normal; e
- (iii) conseguir erros com variância constante.

Por vezes, a transformação para obter a linearidade consegue atingir também os outros objectivos (ver o exemplo da Secção 6.2). O problema da heterogeneidade da variância dos erros pode também ser ultrapassado utilizando mínimos quadrados ponderados (ver os exemplos das Secções 6.1 e 6.2).

6.1 Centro de inspecção e reparação de automóveis

O exemplo de aplicação, tratado nesta secção, é um centro de inspecção e reparação de automóveis onde estes são testados e, se for necessário, reparados para poderem andar em circulação. Quando um automóvel necessita de ser inspecionado, dirige-se a este centro onde primeiramente é testado numa secção de inspecção. Se for considerado que não está em condições para andar em circulação

(por exemplo, não trava convenientemente, o espelho retrovisor está partido, os valores de CO não estão dentro dos parâmetros legais, etc), então ele dirige-se para uma secção de reparação onde é arranjado mediante uma lista de reparações, indicada na secção de inspecção. Após a reparação, o automóvel é mandado de volta para a secção de inspecção, onde é novamente inspeccionado. Os automóveis que forem aprovados na secção de inspecção, quer à primeira vez ou após um ou mais trajectos até à secção de reparação, abandonam o centro com um certificado de que estão aptos para andar em circulação.

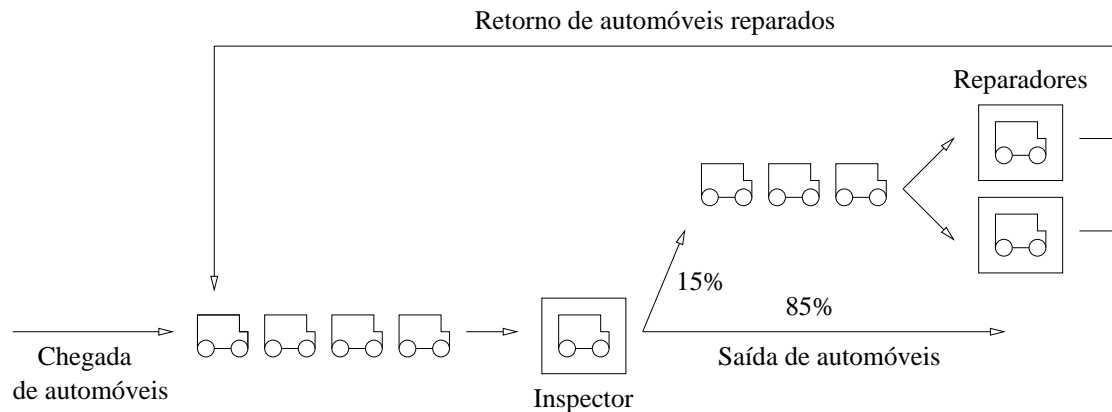


Figura 6.1: Centro de inspecção e reparação de automóveis.

A situação, aqui descrita, está representada na Figura 6.1. O tempo entre chegadas dos automóveis ao centro é normalmente distribuído, com média μ e desvio padrão igual a $\sqrt{15}$ minutos. Um único inspector aguarda a chegada de um automóvel à secção de inspecção e o tempo necessário para o inspeccionar é uniformemente distribuído entre 15 e 25 minutos. Na fila de inspecção, apenas está disponível espaço para seis automóveis. Em média, 85% dos automóveis passam na inspecção e abandonam o centro. Os outros 15% têm de dirigir-se à secção de reparação, onde dois reparadores (mecânicos) trabalham lado a lado. Após serem reparados, os automóveis têm de voltar à fila de inspecção. O tempo necessário para reparar um automóvel é exponencialmente distribuído, com média igual a 60 minutos. O sistema descrito anteriormente foi convenientemente modelado na linguagem de simulação SLAM II. O modelo gráfico obtido apresenta-se na Figura 6.2, onde os valores das variáveis XX são os seguintes: $XX(1) = 1, 5, 10, 15, 20, 23, 26, 29, 32, 35, 40, 50, 60, 90$; $XX(2) = 15$; $XX(3) = 15$; $XX(4) = 25$; e $XX(5) = 60$.

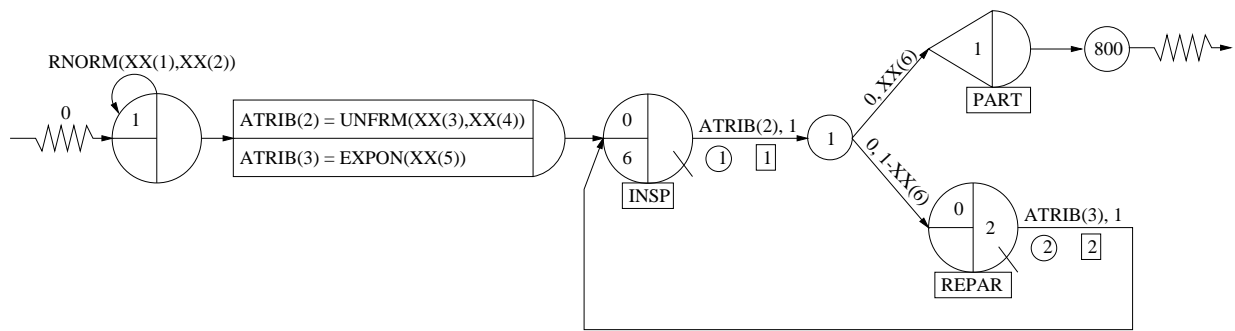


Figura 6.2: Modelo na linguagem SLAM II (centro de inspeção e reparação de automóveis).

Neste caso, o objectivo é expressar o tempo médio de permanência dos automóveis no sistema (resposta), como função do tempo médio entre chegadas dos automóveis (parâmetro de entrada). A variável dependente do metamodelo é a resposta e representa-se por Y . A variável independente do metamodelo é idêntica ao parâmetro de entrada μ e representa-se por X .

• Obtenção de resultados para metamodelação

Com o objectivo de obter os resultados para construção do metamodelo, começou-se por realizar 20 réplicas piloto do programa de simulação, todas elas com "comprimento" 2000, para cada um dos seguintes 10 valores distintos do parâmetro de entrada μ : $\{\mu_i : i = 1, 10\} = \{1, 10, 20, 30, 40, 50, 60, 70, 80, 90\}$. Cada uma das 20 réplicas começa com o sistema vazio, ou seja, nenhum carro à espera para ser inspeccionado. Para cada ponto experimental, utilizou-se o procedimento de Welch para determinar o viés inicial. As médias móveis de Welch baseiam-se nas 20 réplicas, de comprimento 2000, do modelo de simulação. Por exemplo, para o ponto experimental $\mu = 10$, ignoram-se as primeiras 100 observações e apenas se aproveitam as restantes 600, aproximadamente 85% das observações em cada réplica, para estimar a resposta Y do sistema (ver a Figura 6.3 e a Tabela 6.1).

Em seguida, representaram-se graficamente os pontos (μ_i, \bar{Y}_i) e observou-se que, a partir de $\mu = 60$, as distâncias entre valores consecutivos de \bar{Y}_i são bastante pequenas. Por conseguinte, basta considerar $\mu = 60$ e $\mu = 90$. O mesmo não acontece para valores consecutivos de \bar{Y}_i correspondentes a μ entre 20 e 40. Este problema foi ultrapassado, considerando valores de μ a variar de três em três minutos: $\mu = 20, 23, 26, 29, 32, 35, 40$. Entre $\mu = 1$ e $\mu = 20$, as distâncias entre

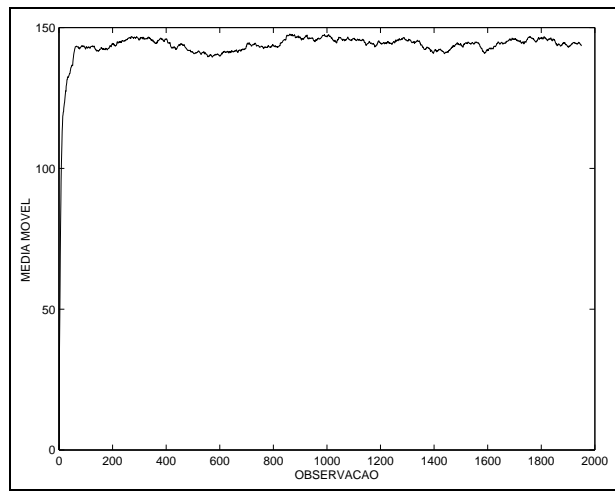


Figura 6.3: Para $\mu = 10$, média móvel com janela de Welch $w = 50$ baseada em 20 réplicas, onde cada réplica contém 2000 observações.

valores consecutivos de \bar{Y}_i , não são muito grandes, mas foi conveniente considerar mais alguns pontos para que se possa observar melhor o comportamento dos resultados e, conseqüentemente, escolher melhor o metamodelo hipotético. Finalmente, após esta busca computacional de desenho ótimo, consideraram-se 14 pontos experimentais para construir o metamodelo: $\{\mu_i : i = 1, 14\} = \{1, 5, 10, 15, 20, 23, 26, 29, 32, 35, 40, 50, 60, 90\}$.

Para cada novo ponto experimental, realizaram-se também 20 réplicas de comprimento 2000 e recorreu-se ao procedimento de Welch para detecção do viés inicial. Os resultados obtidos correspondentes ao procedimento de Welch encontram-se na Tabela 6.1.

Para efeitos de selecção, estimação e validação do metamodelo, consideraram-se $r = 30$ réplicas em cada um dos $n = 14$ pontos experimentais. Os resultados da simulação $\{(X_i, Y_{ij}, \hat{\sigma}_i) : i = 1, \dots, n, j = 1, \dots, r\}$ foram recolhidos com base na Tabela 6.1; X_i representa a média do tempo entre chegadas de automóveis ao centro na experiência i e Y_{ij} é a média do tempo de permanência dos automóveis no centro durante a j -ésima réplica da i -ésima experiência. Por exemplo, para $\mu_i = 10$, geraram-se 800 observações, mas apenas as últimas 700 foram utilizadas no desenvolvimento do metamodelo. Observe-se que: (i) para aplicar o teste de Rao, é preciso que $r > n$ (ver a página 90); (ii) quando a variância dos erros não é constante, então para obter uma estimativa apropriada de $\hat{\sigma}_i$, $i = 1, \dots, n$, é aconselhável que $r > 9$ (ver a página 65). Cada uma das 30 réplicas começa com o sistema vazio, ou seja, nenhum carro à espera para ser

Tabela 6.1: Detecção do viés inicial.

μ_i	Observações		Janela de Welch
	Ignoradas	Em execução	
1, 5, 10	100	700	50
15	150	1000	100
20	200	1400	150
23, 26, 29, 32, 35	200	1400	300
40, 50, 60	100	700	200
90	50	350	200

inspeccionado.

• **Seleccção do tipo de metamodelo**

Com o objectivo de identificar uma curva que possa aproximar a relação entre os dados e os resultados do programa de simulação, realizou-se uma inspecção visual representando graficamente os pares (X_i, Y_{ij}) (onde $i = 1, \dots, n$ e $j = 1, \dots, r$) e comparando o resultado com a representação gráfica de algumas relações funcionais. As curvas não lineares candidatas (incluídas no catálogo de curvas do Apêndice A), que se parecem com os resultados de simulação representados na Figura 6.4, encontram-se na Tabela 6.2. 'atan' é a função arco cuja tangente e as outras

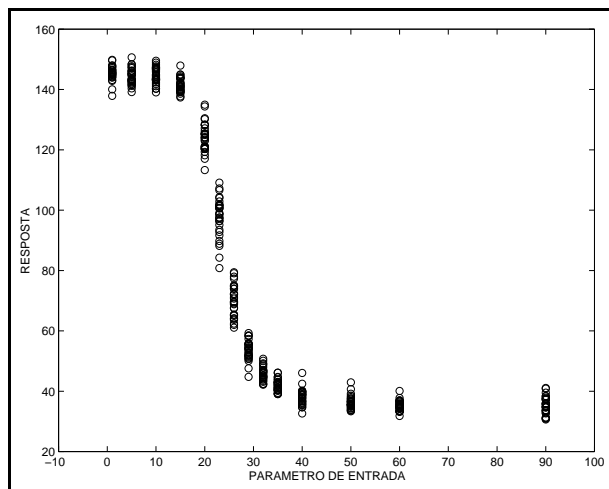


Figura 6.4: Resultados da simulação do centro de inspecção e reparação de automóveis.

Tabela 6.2: Curvas não lineares candidatas.

Metamodelo	Expressão
atan	$y = \theta_1 + \theta_2 \arctg(\theta_3 x + \theta_4)$
logístico	$y = \frac{\theta_1}{1 + \theta_2 e^{-\theta_3 x}} + \theta_4$
Weibull	$y = \theta_1 - (\theta_1 - \theta_2) e^{-(\theta_3 x)^{\theta_4}}$
MMF	$y = \theta_2 - \frac{\theta_2 - \theta_1}{1 + (\theta_3 x)^{\theta_4}}$

são três modelos de crescimento de sigmóide (ver [107, páginas 329-340]): 'Logístico' é o modelo logístico, 'Weibull' é a curva de crescimento de Weibull e 'MMF' é a família de Morgan-Mercer-Flodin. Também foram consideradas funções polinomiais de grau r , com $r = 2, \dots, 10$.

• Estimação do metamodelo

Na Figura 6.4 observa-se que a variância não é constante, para os vários pontos experimentais. E efectivamente obtém-se para *het* um valor bastante diferente de 1 (ver a quantidade (3.4)): $het = 3.946$. Por conseguinte, utilizou-se o método dos mínimos quadrados ponderados não lineares para os ajustes com curvas não lineares e o método dos mínimos quadrados ponderados no caso das curvas polinomiais. Os valores para os estimadores dos parâmetros desconhecidos dos metamodelos hipotéticos, presentes na Tabela 6.2, foram obtidos utilizando o método de Levenberg-Marquardt realizado na linguagem de programação MATLAB, com tolerância terminal igual a 10^{-6} e número máximo de cálculos da função igual a 600 (por omissão, é $100 \times$ número de parâmetros desconhecidos). No caso de ajustes utilizando polinómios de grau r , com $r = 4, \dots, 10$, obtiveram-se matrizes próximas do singular ou mal condicionadas. Uma vez que, neste caso, os resultados podem ser imprecisos, rejeitaram-se os metamodelos a eles correspondentes.

• Validação do metamodelo estimado

Com o objectivo de averiguar a adequação dos metamodelos restantes, utilizaram-se alguns métodos descritos na Secção 4.1. Por exemplo, no que diz respeito à capacidade previsiva dos metamodelos, utilizou-se a estatística PRESS e obtiveram-se os resultados presentes na Tabela 6.3. Os valores de SSE e PRESS para os metamodelos não lineares são semelhantes, em contraste

com o caso dos metamodelos polinomiais. Também, os valores de SSE são elevados para os metamodelos lineares, comparando com os não lineares. Com base nestes resultados, conclui-se que os metamodelos polinomiais têm falta de validade previsiva, logo não são boas aproximações, tanto para o modelo de simulação, como para o problema real.

Tabela 6.3: Diagnósticos do metamodelo (centro de inspeção e reparação de automóveis).

Metamodelo	SSE	MSE	PRESS	SSE/PRESS
atan	1554.23	3.86624	1608.21	0.966
logístico	1647.62	4.09856	1696.78	0.971
MMF	1366.68	3.39970	1418.26	0.964
Weibull	2224.96	5.53473	2272.22	0.979
pol2	39714.1	98.5462	15162.9	2.619
pol3	40041.4	99.6055	15105.4	2.651

Utilizou-se o teste F de Rao, para averiguar a falta de ajustamento dos metamodelos não lineares. Foi referido, na página 90, que se obtêm melhores resultados quando a distribuição das respostas é simétrica. Frequentemente, as respostas em pontos experimentais distintos têm distribuições diferentes, porque a variância não é constante ao longo de todos os pontos experimentais, ou então o grau de não simetria varia de ponto experimental para ponto experimental. Como consequência, averiguou-se a normalidade das respostas em cada ponto experimental, através da representação gráfica de probabilidade normal das respostas observadas $\{Y_{ij} : j = 1, \dots, r\}$. Todas as representações gráficas obtidas são aproximadamente lineares, mas a inclinação varia com o ponto experimental correspondente (ver, na Figura 6.5, os resultados para alguns pontos experimentais).

Por conseguinte, conclui-se que a resposta em cada ponto experimental tem distribuição marginal normal, com a variância dependendo do ponto experimental. Sendo assim e uma vez que $m < n < r$, aplicou-se o teste F de Rao e obtiveram-se os resultados expostos na Tabela 6.4.

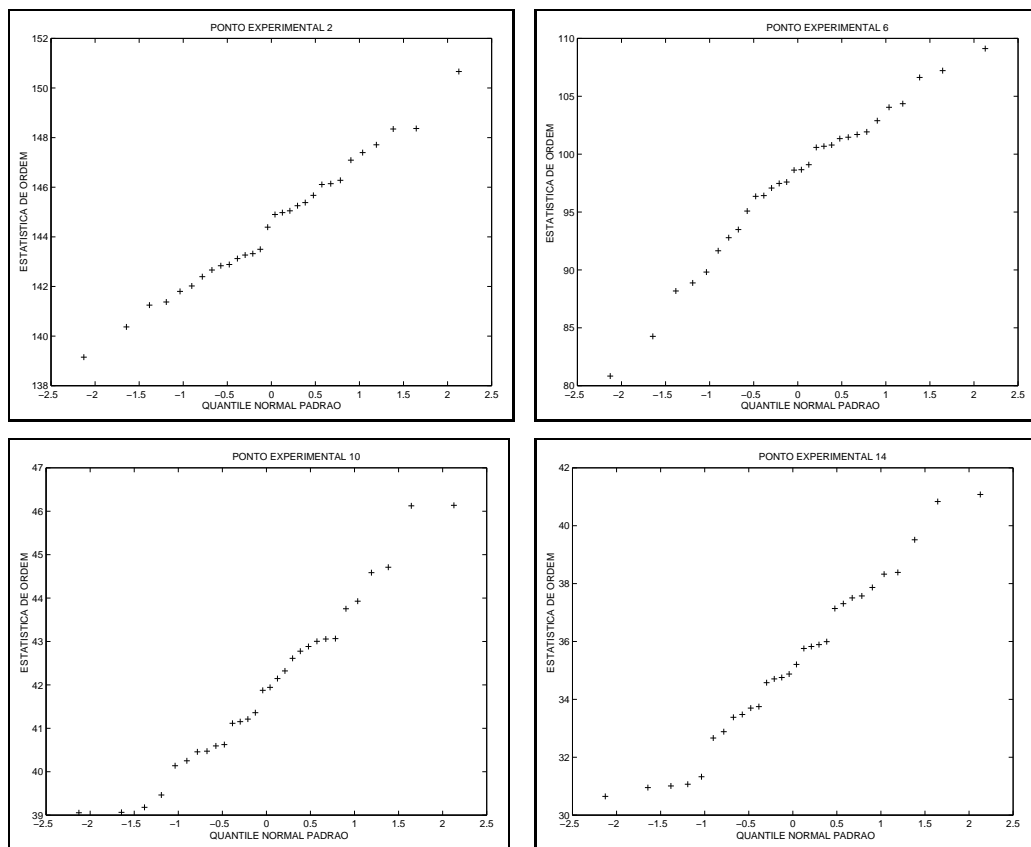


Figura 6.5: Amostras da resposta em papel de probabilidade normal: quantil normal $\Phi^{-1}[(j - 0.5)/r]$ versus a j -ésima estatística de ordem $Y_{i(j)}$ da resposta (pontos experimentais 2, 6, 10 e 14).

Comparando os valores de F_{Rao} com o valor crítico $F_{n-m, r-n+m; \alpha} = F_{10, 10; 0.05} = 2.978$, apenas o metamodelo 'MMF' não é rejeitado. Para além dos métodos até agora utilizados, a análise gráfica dos resíduos mostra que se verificam as hipóteses distribucionais associadas ao metamodelo 'MMF' (ver as Figuras 6.6 e 6.7).

Uma vez que não se conhecem os dados relativos ao problema real, apenas é possível validar o metamodelo em relação ao modelo de simulação. Para isso, utilizou-se o método de validação cruzada dupla descrito na Secção 4.2 e obtiveram-se os resultados apresentados na Tabela 6.5. Observa-se uma boa concordância entre os coeficientes obtidos com base nos subconjuntos 1 e 2. Também os valores dos coeficientes de determinação são semelhantes.

De acordo com os critérios utilizados para averiguar a validade dos metamodelos hipotéticos

Tabela 6.4: Teste de Rao.

Metamodelo	atan	logístico	MMF	Weibull
F_{Rao}	6.108	9.102	2.206	22.673

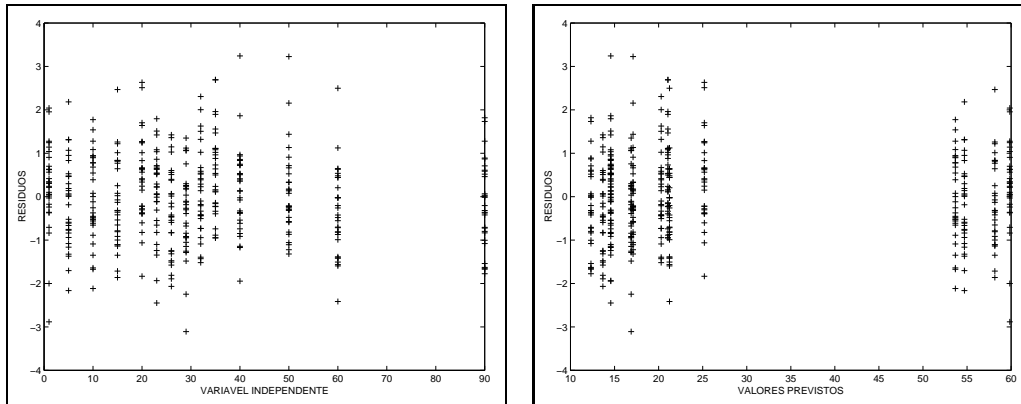


Figura 6.6: Gráficos dos resíduos ponderados (metamodelo 'MMF').

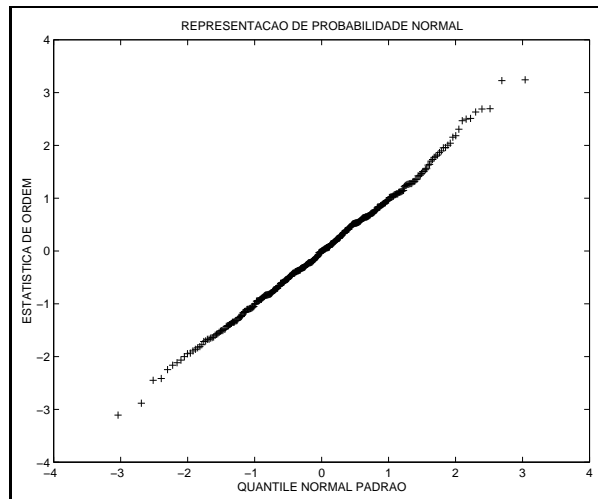


Figura 6.7: Papel de probabilidade normal (metamodelo 'MMF'): $\Phi^{-1}[(j - 0.5)/r]$ versus o j -ésimo resíduo ponderado ordenado.

Tabela 6.5: Teste de validação cruzada dupla, para o metamodelo 'MMF'.

Estatística	MMF	
	Subconjunto 1	Subconjunto 2
$\hat{\theta}_1$	144.58	145.29
$\hat{\theta}_2$	35.681	35.855
$\hat{\theta}_3$	0.0422	0.0420
$\hat{\theta}_4$	7.7327	7.7974
R^2_{const}	1.0003	0.9934
R^2_{val}	0.9866	1.0073
MSPR	1.1522	1.0130

estudados, o metamodelo considerado válido foi o 'MMF'. Os valores obtidos para estimar os parâmetros desconhecidos do metamodelo 'MMF', bem como as estimativas dos desvios padrões e os intervalos de confiança, com probabilidade de cobertura de pelo menos $1 - \alpha = 0.95$ (ver (3.26)), correspondentes a esses valores, com base em todo o conjunto de dados, encontram-se na Tabela 6.6. Observa-se que, as estimativas dos desvios padrões e as metades dos comprimentos dos intervalos de confiança assumem valores baixos, comparativamente com a grandeza dos valores numéricos dos estimadores. Como resultado, obteve-se neste estudo grande precisão dos estimadores de regressão.

Tabela 6.6: Intervalos de confiança de 95% para os parâmetros do metamodelo 'MMF' (centro de inspeção e reparação de automóveis).

Coeficiente de regressão	Estimador	Desvio padrão	Intervalo de confiança
θ_1	144.882	2.4852×10^{-1}	$144.882 \pm 9.3190 \times 10^{-1}$
θ_2	35.797	2.0540×10^{-1}	$35.797 \pm 7.7020 \times 10^{-1}$
θ_3	0.0421	1.2820×10^{-4}	$0.0421 \pm 4.8072 \times 10^{-4}$
θ_4	7.776	1.2185×10^{-1}	$7.776 \pm 4.5691 \times 10^{-1}$

6.2 Modelo de filas de espera $M/M/1$

Nesta secção, estuda-se o modelo de filas de espera $M/M/s$, com um único servidor e uma única fila de espera. As chegadas decorrem de acordo com um processo de Poisson, com uma taxa de chegadas constante λ e considera-se que os tempos de serviço seguem uma distribuição exponencial, com tempo médio de serviço constante e igual a $1/\mu$. Neste caso, o objectivo é expressar a média do tempo de espera na fila (resposta), como função do factor de utilização da fila (parâmetro de entrada). Estudaram-se doze ($n = 12$) configurações diferentes do sistema. A cada configuração corresponde um valor específico do factor de utilização, $\rho = \lambda/\mu$. Considerou-se sempre $\mu = 1$ e, conseqüentemente, $\rho = \lambda$. Os doze pontos experimentais consistem nas seguintes variações: $\{\lambda_i : i = 1, \dots, 12\} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.7, 0.75, 0.85, 0.9, 0.95\}$. Para cada um dos $n = 12$ pontos experimentais, foram executadas $r = 20$ réplicas. Cada uma das 20 réplicas começa com o sistema vazio, ou seja, nenhum cliente à espera para ser servido. Tal como na secção anterior, também aqui se recorreu ao procedimento de Welch para determinar o viés inicial e a duração de cada execução do programa de simulação (ver a Tabela 6.7). Em cada execução do programa de simulação, foram ignoradas aproximadamente 15% do número total de observações.

Tabela 6.7: Detecção do viés inicial.

ρ_i	Observações		Janela de Welch
	Ignoradas	Em execução	
0.10, 0.20, 0.30	500	3500	1000
0.40	1000	7000	1000
0.50	1500	10000	1000
0.55, 0.60, 0.70, 0.75	1500	10000	4000
0.85	2000	20000	8000
0.90	2500	20000	10000
0.95	3500	40000	10000

Na construção do metamodelo, X_i representa o factor de utilização na experiência i e Y_{ij} é a média do tempo de espera na fila durante a i -ésima execução do programa de simulação, com o factor de utilização X_i e sequência de números aleatórios r_{ij} , isto é, Y_{ij} é a resposta correspondente à j -ésima réplica da experiência i .

Com o objectivo de procurar metamodelos hipotéticos, comparou-se o diagrama de dispersão da resposta em função da variável de decisão (ver a Figura 6.8) com as curvas do catálogo presente no Apêndice A. As curvas não lineares candidatas, que parecem seguir mais de perto os resultados

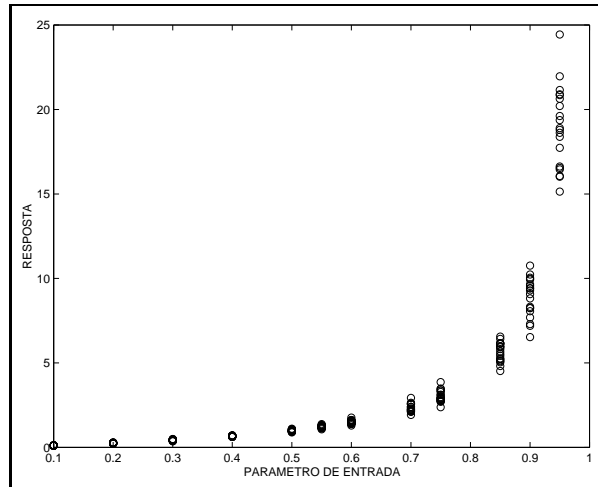


Figura 6.8: Resultados da simulação do sistema $M/M/1$.

do diagrama de dispersão, são apresentadas na Tabela 6.8 (ver também, nas Figuras 6.9 e 6.10, exemplos destas curvas candidatas para determinados valores dos parâmetros θ_1 e θ_2).

Tabela 6.8: Curvas candidatas.

Metamodelo	Expressão
Monómio	$y = \theta_1 x^{\theta_2}$
Hipérbole	$y = \frac{\theta_1 x}{1 + \theta_2 x}$

Embora o monómio presente na Tabela 6.8 seja uma função não linear, ele é linearizável. Considere-se o metamodelo correspondente a este monómio:

$$Y_{ij} = \theta_1 X_i^{\theta_2} + \epsilon_{ij},$$

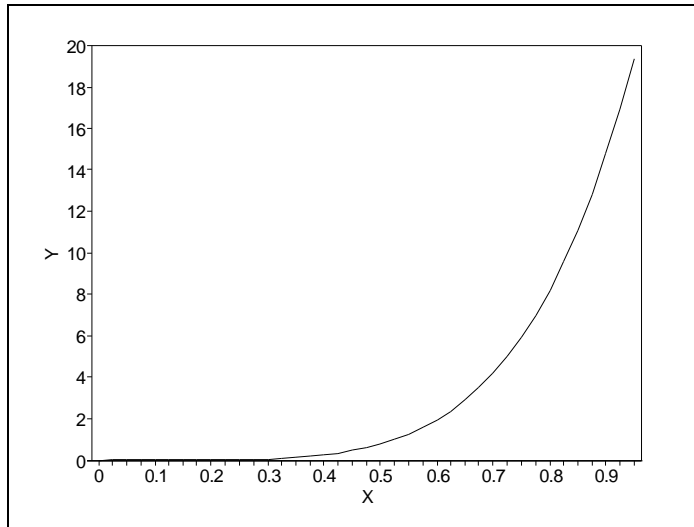


Figura 6.9: Monómio com $\theta_1 = 25$ e $\theta_2 = 5$.

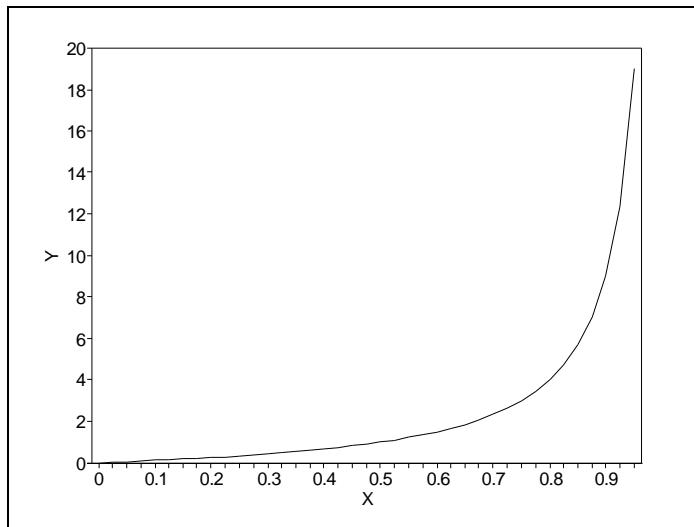


Figura 6.10: Hipérbole com $\theta_1 = 1$ e $\theta_2 = -1$.

Aplicando a transformação logarítmica, em ambos os membros da equação anterior, obtém-se

$$\log Y_{ij} = \log \theta_1 + \theta_2 \log X_i + e_{ij},$$

onde $i = 1, \dots, n$ e $j = 1, \dots, r$. $e_{ij} = \log(1 + \epsilon_{ij}/E[Y_{ij}])$ e $E[e_{ij}] \approx 0$ porque ϵ_{ij} é pequeno comparativamente com $E[Y_{ij}]$. Além disso, $\text{Var}[e_{ij}]$ é aproximadamente constante (neste caso $\text{het} \approx 2.32$). Fazendo uma simples mudança de variáveis no problema anterior surge a familiar equação de regressão linear

$$Z_{ij} = t_1 + t_2 W_i + e_{ij},$$

onde $Z_{ij} = \log Y_{ij}$, $t_1 = \log \theta_1$, $t_2 = \theta_2$, $W_i = \log X_i$. Como resultado, utilizou-se o método dos mínimos quadrados usuais para estimar t_1 e t_2 e, depois, aplicou-se a transformação inversa para obter $\theta_1 = 10^{t_1}$. Este método foi realizado na linguagem de programação MATLAB, por forma a obter as estimativas pretendidas.

Em relação ao metamodelo hiperbólico, utilizou-se o método dos mínimos quadrados ponderados não lineares porque a função não é linearizável e a variância das observações Y_{ij} não é constante. Obteve-se $\text{het} = 256.13$ e em particular observa-se, na Figura 6.8, que a variância das observações Y_{ij} aumenta, à medida que aumenta o valor da variável independente. Uma alternativa, a esta aproximação, é aplicar a transformação logarítmica em ambos os membros de (3.3) e, depois, aplicar o método dos mínimos quadrados usuais não lineares (ver a Secção 6.2.1). Neste caso, não é possível linearizar o problema, mas consegue-se estabilizar a variância. As estimativas para θ_1 e θ_2 foram obtidas utilizando o algoritmo de Levenberg-Marquardt, realizado na linguagem de programação MATLAB, com tolerância terminal igual a 10^{-6} e número máximo de cálculos da função igual a 600 (por omissão, é $100 \times$ número de parâmetros desconhecidos).

Os resultados numéricos da estimação encontram-se na Tabela 6.9.

Com o objectivo de averiguar a adequação dos metamodelos, começou-se por calcular os valores presentes na Tabela 6.10. Com base nestes resultados, observou-se que a hipérbole tem uma boa capacidade previsiva e melhor que a do monómio.

Foi descrita na Secção 4.1, uma alternativa à estatística PRESS: um método de validação cruzada. Os valores obtidos, através deste método, estão presentes na Tabela 6.11. O nível de significância global foi $\alpha_E = 0.05$, ou seja, testaram-se doze erros individuais com o nível

Tabela 6.9: Resultados da estimação.

Metamodelo	Coefficiente de regressão	Estimador	Desvio padrão
Hipérbole	θ_1	0.999622	5.558×10^{-3}
	θ_2	-0.999659	1.409×10^{-3}
Monómio	$\log\theta_1$	1.8820	2.224×10^{-1}
	θ_2	2.0514	2.269×10^{-1}

Tabela 6.10: Diagnósticos do metamodelo (sistema $M/M/1$).

Metamodelo	SSE	MSE	PRESS	SSE/PRESS
Hipérbole	95.393	0.4221	100.37	0.95
Monómio	62.229	0.2753	115.81	0.54

$\alpha_c = \alpha_E/12 = 0.00416$. O valor crítico é, neste caso, $t_{r-1; \alpha_c/2} = t_{19; 0.00208} = 3.256$. De acordo com os resultados obtidos, rejeita-se o monómio porque não tem capacidade previsiva.

Em seguida, realizou-se o teste F clássico de falta de ajustamento para ambos os metamodelos (ver a Tabela 6.12), considerando o valor crítico $F_{n-m, N-n; \alpha} = F_{10, 216; 0.05} = 1.875$. O valor obtido para a estatística F não é significativo no caso da hipérbole e é exageradamente grande no caso do monómio. Como resultado, rejeita-se o segundo metamodelo.

Em relação à análise de resíduos, no caso da hipérbole, verifica-se que as hipóteses distribucionais são satisfeitas (ver as Figuras 6.11 e 6.12). Por conseguinte, não existe razão para pensar que o metamodelo não é adequado.

Resta agora averiguar a validade do metamodelo com respeito ao modelo de simulação e ao problema real. Tal como na situação descrita na Secção 6.1, também aqui a validade em relação ao modelo de simulação é investigada, utilizando o método de validação cruzada dupla. Observa-se, na Tabela 6.13, uma boa concordância entre os valores das duas colunas da direita logo, o metamodelo é válido com respeito ao modelo de simulação.

Tabela 6.11: Um método de validação cruzada.

X_i	t_i	
	Hipérbole	Monómio
0.10	-0.4376	12.02
0.20	0.7134	0.4770
0.30	0.6999	-3.305
0.40	0.1366	-10.01
0.50	-0.8119	-8.635
0.55	-0.3409	7.779
0.60	0.01869	-6.295
0.70	0.3359	-3.209
0.75	0.3591	-1.785
0.85	-0.007776	1.901
0.90	-0.1378	4.158
0.95	0.02754	10.20
$\max_{i=1,\dots,n} t_i $	0.8119	12.0213

Tabela 6.12: Teste de falta de ajustamento (sistema $M/M/1$).

Fonte	Metamodelo	Graus de liberdade	Soma dos quadrados	Média dos quadrados	Teste F
Falta de ajustamento	Hipérbole	10	1.1469	0.11469	0.2749
	Monómio		60.306	6.0306	714.9
Erro puro	Hipérbole	228	95.1223	0.4172	
	Monómio		1.92329	8.4355×10^{-3}	

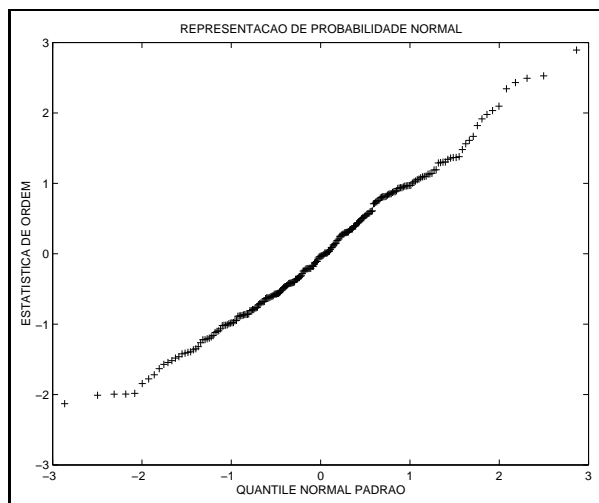


Figura 6.11: Papel de probabilidade normal (hipérbole): $\Phi^{-1}[(j - 0.5)/r]$ versus o j -ésimo resíduo ponderado ordenado.

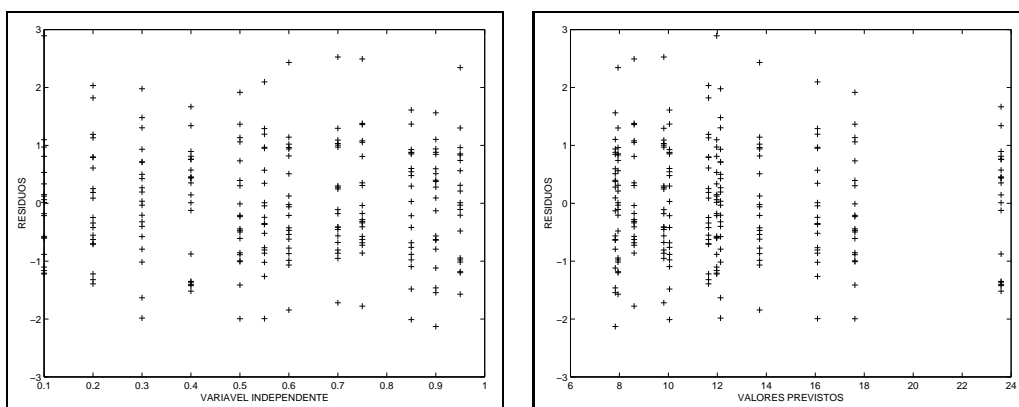


Figura 6.12: Gráficos dos resíduos ponderados (hipérbole).

Tabela 6.13: Teste de validação cruzada dupla, para o metamodelo hiperbólico.

Estatística	Hipérbole	
	Subconjunto 1	Subconjunto 2
$\hat{\theta}_1$	0.996627	1.00148
$\hat{\theta}_2$	-1.00034	-0.998794
R_{const}^2	0.9817	0.9413
R_{val}^2	0.9200	1.0045
MSPR	0.9342	0.9840

Tabela 6.14: Solução analítica *versus* metamodelo hiperbólico.

X_i	Solução analítica	Metamodelo	Erro relativo
0.10	0.11111	0.11106	4.161×10^{-4}
0.20	0.25000	0.24988	4.634×10^{-4}
0.30	0.42857	0.42835	5.242×10^{-4}
0.40	0.66666	0.66626	6.053×10^{-4}
0.50	1.0000	0.99928	7.187×10^{-4}
0.55	1.2222	1.2213	7.943×10^{-4}
0.60	1.4999	1.4987	8.888×10^{-4}
0.70	2.3333	2.3306	1.172×10^{-3}
0.75	3.0000	2.9958	1.399×10^{-3}
0.85	5.6666	5.6536	2.304×10^{-3}
0.90	9.0000	8.9691	3.434×10^{-3}
0.95	18.9999	18.8707	6.807×10^{-3}
% erro relativo médio = 1.63×10^{-3}			

Em relação à validação com respeito ao problema real, procede-se da forma descrita na Secção 4.3. Neste caso, o problema real tem solução analítica conhecida, o que facilita a investigação (ver, por exemplo, [44]). Sabe-se que a média do tempo de espera na fila, correspondente ao sistema $M/M/1$, é dada por $\lambda/\mu(\mu - \lambda)$. Como foi referido anteriormente, neste exemplo tem-se $\mu = 1$. Por conseguinte, os valores exactos são calculados a partir de

$$\frac{\lambda}{1 - \lambda}, \quad 0 \leq \lambda < 1$$

(por exemplo, para o ponto experimental $X_i = 0.1$, o valor exacto é $0.1/(1 - 0.1) = 0.111$). A percentagem de erro relativo médio obtida é baixa, donde se conclui que o metamodelo hiperbólico aproxima suficientemente bem o problema real (ver a Tabela 6.14).

Tendo em conta a averiguação da validade aqui descrita, o metamodelo hiperbólico é um metamodelo válido. Para finalizar, apresentam-se os intervalos de confiança, com probabilidade de cobertura de pelo menos $1 - \alpha = 0.95$ (ver (3.26)), para os parâmetros do metamodelo (ver a Tabela

6.15) e testa-se a hipótese $H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0 = (1.0, -1.0)^T$ versus $H_1 : \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0 = (1.0, -1.0)^T$, com nível de confiança de 95%. O valor obtido para (3.30) foi $\chi = 0.1047$, sendo o correspondente valor crítico $\chi_{m;\alpha}^2 = \chi_{2;0.05}^2 = 5.992$. Consequentemente, não se rejeita a hipótese H_0 .

Tabela 6.15: Intervalos de confiança de 95% para os parâmetros (sistema $M/M/1$).

Coeficiente de regressão	Estimador	Desvio padrão	Intervalo de confiança
θ_1	0.999622	5.5581×10^{-3}	$0.996217 \pm 1.8812 \times 10^{-2}$
θ_2	-0.999659	1.4094×10^{-3}	$-0.999659 \pm 4.7704 \times 10^{-3}$

6.2.1 Método das variáveis de controlo

Verificou-se, anteriormente, que o metamodelo hiperbólico é um metamodelo válido para explicar, tanto o modelo de simulação, como o problema real. Em particular ao longo dos pontos experimentais, tem-se a seguinte relação entre o tempo de espera médio na fila e o factor de utilização:

$$Y_{ij} = \frac{\theta_1 X_i}{1 + \theta_2 X_i} + \epsilon_{ij}, \quad (6.1)$$

onde $\epsilon_{ij} \sim N(0, \sigma_i^2)$, com $i = 1, \dots, n$, $j = 1, \dots, r$. Com o objectivo de estabilizar a variância, aplica-se a transformação logarítmica em ambos os membros de (6.1) e obtém-se

$$\log Y_{ij} = \log \frac{\theta_1 X_i}{1 + \theta_2 X_i} + v_{ij}, \quad (6.2)$$

onde $v_{ij} = \log(1 + \epsilon_{ij}/E[Y_{ij}])$, com $i = 1, \dots, n$, $j = 1, \dots, r$. $E[v_{ij}] \approx 0$ porque ϵ_{ij} é pequeno comparativamente com $E[Y_{ij}]$. Uma vez que o valor obtido $het = 2.32$ não está significativamente afastado de 1, conclui-se que a variância $Var[v_{ij}]$ pode ser considerada aproximadamente constante (note-se que foi observado que, antes da transformação de dados, se verificava $het = 256.13$).

Para melhorar a precisão dos resultados da estimação, escolheram-se as seguintes variáveis de controlo

$$C_{ijk} = \frac{t_{kij} - \mu_{ki}}{s_{ki}},$$

com $k = 1, 2, i = 1, \dots, n, j = 1, \dots, r$ e onde t_{1ij} é o tempo médio entre chegadas e t_{2ij} é o tempo médio de serviço, amostrados de distribuições exponenciais com médias e variâncias conhecidas.

O problema hipotético controlado é

$$\log Y_{ij} = \log \frac{\theta_1 X_i}{1 + \theta_2 X_i} + \delta_1 C_{1ij} + \delta_2 C_{2ij} + \nu_{ij}, \quad i = 1, \dots, 12, \quad j = 1, \dots, 20.$$

• Estimação do metamodelo

Para efeitos de estimação, utilizou-se o método dos mínimos quadrados não lineares. Os valores para os estimadores dos parâmetros desconhecidos foram obtidos utilizando o método de Levenberg-Marquardt realizado na linguagem de programação MATLAB, com tolerância terminal igual a 10^{-6} e número máximo de cálculos da função igual a 600 (ver a Tabela 6.16).

Tabela 6.16: Coeficientes de regressão estimados (sistema $M/M/1$).

Coeficientes de regressão	Estimador directo $\hat{\theta}$	Estimador controlado $\hat{\theta}(\hat{\delta})$
θ_1	0.998249	1.00011875
θ_2	-0.999231	-0.999134934

• Validação do metamodelo estimado

Após a estimação, procedeu-se à validação do metamodelo controlado. Os valores obtidos para SSE e PRESS estão bastante próximos:

$$\frac{\text{SSE}}{\text{PRESS}} = \frac{1.931}{2.050} = 0.942.$$

Também o teste F clássico de falta de ajustamento dá um bom resultado: uma vez que $F_{n-m, N-n; \alpha} = F_{8, 228; 0.05} \approx 1.95$, não se rejeita o metamodelo controlado (ver a Tabela 6.17).

Tabela 6.17: Teste de falta de ajustamento (sistema $M/M/1$, com variáveis de controlo).

Fonte	Graus de liberdade	Soma dos quadrados	Média dos quadrados	Teste F
Falta de ajustamento	8	0.008036	0.001004	0.119
Erro puro	228	1.923	0.008435	

Os resultados da validação, em relação ao modelo de simulação e ao problema real, encontram-se nas Tabelas 6.18 e 6.19 respectivamente. Com base nestes resultados, não há razão para rejeitar o metamodelo controlado. No caso da comparação do metamodelo com o problema real, os resul-

Tabela 6.18: Teste de validação cruzada dupla, para o metamodelo controlado.

Estatística	Metamodelo controlado	
	Subconjunto 1	Subconjunto 2
$\hat{\theta}_1(\hat{\delta})$	0.9735	1.0149
$\hat{\theta}_2(\hat{\delta})$	-1.0001	-0.9986
R_{const}^2	0.993	0.998
R_{val}^2	1.010	0.983
MSPR	8.595×10^{-3}	7.688×10^{-3}

tados foram obtidos do seguinte modo: uma vez que o metamodelo, com inclusão de variáveis de controlo, é da forma

$$\log Y_{ij} = \log \frac{1.0001X_i}{1 - 0.9991X_i} - 0.071745C_{1ij} - 0.83652C_{2ij}, \quad (6.3)$$

pode-se aplicar a transformação exponencial a ambos os membros de (6.3) e vem

$$Y_{ij} = \frac{1.0001X_i}{1 - 0.9991X_i} e^{-0.071745C_{1ij}} e^{-0.83652C_{2ij}}. \quad (6.4)$$

Por conseguinte, os erros relativos foram calculados utilizando a solução aproximada (6.4) e a solução exacta $X_i/(1 - X_i)$, com $i = 1, \dots, n$ e $j = 1, \dots, r$.

Tabela 6.19: Solução analítica e metamodelo controlado: uma comparação.

X_i	Solução analítica	Metamodelo	Erro relativo
0.10	0.11111	0.11085	2.3590×10^{-3}
0.20	0.25000	0.24939	2.4577×10^{-3}
0.30	0.42857	0.42746	2.5845×10^{-3}
0.40	0.66666	0.66483	2.7535×10^{-3}
0.50	1.0000	0.99701	2.9900×10^{-3}
0.55	1.2222	1.2184	3.1476×10^{-3}
0.60	1.4999	1.4950	3.3446×10^{-3}
0.70	2.3333	2.3242	3.9350×10^{-3}
0.75	3.0000	2.9868	4.4068×10^{-3}
0.85	5.6666	5.6310	6.2897×10^{-3}
0.90	9.0000	8.9223	8.6332×10^{-3}
0.95	18.9999	18.704	1.5598×10^{-2}
% erro relativo médio = 4.88×10^{-3}			

• **Intervalos de confiança e testes de hipóteses**

Uma vez que $F_{m,N-m-q;\alpha} = F_{2,236;0.05} = 3.034$ e $\hat{\tau}^2 = 8.184 \times 10^{-3}$, o elipsóide de confiança aproximada, com nível de confiança de 95% e centrada em $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, é dada por (ver (5.49)):

$$\left\{ \boldsymbol{\theta}^* : (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^T (\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*) \leq 4.966 \times 10^{-2} \right\},$$

onde

$$\mathbf{B}\mathbf{B}^T = \begin{bmatrix} 3.2891 \times 10^{-2} & 1.8495 \times 10^{-3} \\ 1.8495 \times 10^{-3} & 2.2099 \times 10^{-4} \end{bmatrix}.$$

O rectângulo de confiança aproximado correspondente, para $\boldsymbol{\theta}$, com probabilidade de cobertura de pelo menos $1 - \alpha = 0.95$, é dado na Tabela 6.20 (ver (5.50)); na construção dos intervalos de confiança, utilizou-se $t_{N-m-q;\alpha/(2m)} = t_{236;0.0125} \approx 2.256$.

Tabela 6.20: Intervalos de confiança de 95% para os parâmetros do metamodelo hiperbólico (sistema $M/M/1$, com variáveis de controlo).

Coeficiente de regressão	Estimador controlado	Desvio padrão	Intervalo de confiança
θ_1	$\hat{\theta}_1(\hat{\boldsymbol{\delta}})$	5.7478×10^{-3}	$1.00011875 \pm 3.34797 \times 10^{-3}$
θ_2	$\hat{\theta}_2(\hat{\boldsymbol{\delta}})$	4.7114×10^{-4}	$-0.999134934 \pm 2.74428 \times 10^{-4}$

Testou-se a hipótese $H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0 = (1.0, -1.0)^T$ versus $H_1 : \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0 = (1.0, -1.0)^T$, com nível de confiança de 95% e utilizando $F_{m,N-m-q;\alpha} = F_{2,236;0.05} = 3.034$, obteve-se: $(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^T (m\hat{\tau}^2\mathbf{B}\mathbf{B}^T)^{-1} (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0) = 0.3848$ (ver (5.53)). Por conseguinte, não se rejeita a hipótese H_0 .

• **Razão entre variâncias e factor de perda experimentais**

Com o objectivo de estimar as razões entre variâncias e o factor de perda experimentais, seguiu-se o procedimento descrito na Secção 5.4. Como consequência, elaborou-se uma meta-experiência com K réplicas independentes da experiência inicial, onde a experiência inicial consiste em r réplicas em cada um dos n pontos experimentais $\rho = 0.1, 0.2, \text{ etc.}$ Os resultados são apresentados na Tabela 6.21.

Tabela 6.21: Estimativas das razões entre variâncias e do factor de perda.

Coeficiente de regressão	Razões estimadas entre variâncias			Factor de perda		
	Mínima $\hat{\eta}(\boldsymbol{\delta})$	Actual		Real FP($\hat{\boldsymbol{\delta}}$)	Estimado $\widehat{FP}(\hat{\boldsymbol{\delta}})$	% Erro
		Prevista $\tilde{\eta}(\hat{\boldsymbol{\delta}})$	Observada $\hat{\eta}(\hat{\boldsymbol{\delta}})$			
$\boldsymbol{\theta}$	0.99	1.0072	1.6213	1.0174	1.6377	60.9%

Observa-se que a percentagem máxima de redução de variância generalizada, que pode ser atingida utilizando variáveis de controlo, é aproximadamente $100[1 - \hat{\eta}(\boldsymbol{\delta})]\% = 1\%$. Uma vez que não se conhece $\boldsymbol{\delta}$, ele tem de ser estimado. O estimador resultante $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ tem variância maior comparativamente com o estimador $\hat{\boldsymbol{\theta}}$ sem as variáveis de controlo. Como consequência, nesta situação é vantajoso utilizar variáveis de controlo apenas se o vector de coeficientes $\boldsymbol{\delta}$ for conhecido. O erro obtido para o factor de perda, cerca de 60.9%, é aceitável, uma vez que está próximo do obtido por Porta Nova e Wilson [89], 57.7%. Os referidos autores estudaram a modelação de uma rede de filas de espera, no contexto dos metamodelos lineares de simulação.

Capítulo 7

Conclusões e recomendações

Neste capítulo, faz-se um sumário das contribuições desta investigação sobre a construção de metamodelos de regressão não linear, para simulação de acontecimentos discretos. No decurso deste trabalho, também foram descobertos alguns caminhos interessantes, que podem ser seguidos para trabalho futuro nesta área. Na Secção 7.2, apresentam-se alguns tópicos, que se considera serem os mais promissores para futura investigação.

7.1 Sumário de investigação

O resultado principal desta investigação é o desenvolvimento de uma metodologia para construir metamodelos não lineares de simulação, no contexto da simulação de acontecimentos discretos. Foi importante desenvolver esta metodologia, porque os metamodelos não lineares são frequentemente construídos e utilizados de uma maneira ad hoc. A metodologia distingue rigorosamente estimação e validação de um metamodelo e, para além do metamodelo e do modelo de simulação, salienta a importância do papel do problema real no desenvolvimento de metamodelos.

No caso dos metamodelos não lineares de regressão, propuseram-se procedimentos detalhados para cada passo da metodologia. Deu-se especial atenção aos passos da estimação e validação de metamodelos. No passo da estimação, obteve-se o estimador dos mínimos quadrados de θ

(vector de parâmetros desconhecidos do metamodelo) e a sua distribuição, construíram-se regiões de confiança para θ e propuseram-se procedimentos para fazer testes de hipóteses. Todos estes resultados foram obtidos recorrendo à teoria estatística da regressão não linear. No passo da validação, foram aplicadas directamente, sem alteração alguma: (i) métodos gráficos de análise de resíduos (para averiguar as hipóteses distribucionais do metamodelo); e (ii) o teste F clássico de falta de ajustamento. Todos os outros procedimentos sugeridos no passo da validação (teste F de Rao, estatística PRESS, coeficiente de determinação, método de validação cruzada de Kleijnen, DFFITS, DFTETAS, média das diferenças absolutas percentuais, método de validação quadrada dupla e percentagem de erro relativo médio) tiveram de ser adaptados da teoria da regressão linear. As adaptações que parecem ser mais importantes, a metamodelos não lineares de simulação, foram: (i) a adaptação do teste de Rao; e (ii) a generalização do teste de validação cruzada de Kleijnen para metamodelos não lineares de simulação. Em particular, mostra-se que o referido teste de validação cruzada pode ser utilizado efectuando uma única vez a estimação do metamodelo (Passo 4 da metodologia).

De grande importância foi também o desenvolvimento de técnicas para utilizar múltiplas variáveis de controlo na estimação de um metamodelo não linear geral para uma resposta seleccionada de uma experiência de simulação. Esta técnica pode ser útil em muitas situações em que possam ser identificadas variáveis de controlo concomitantes eficazes. Na parte reservada à estimação de metamodelos, obteve-se o estimador dos mínimos quadrados de δ (vector de coeficientes de controlo) e o correspondente estimador dos mínimos quadrados controlado $\hat{\theta}(\hat{\delta})$. Considerando a hipótese que a resposta e as variáveis de controlo têm distribuição conjunta normal multivariada, construíram-se regiões de confiança para θ centradas em $\hat{\theta}(\delta)$ e em $\hat{\theta}(\hat{\delta})$ e propuseram-se procedimentos para fazer testes de hipóteses. Para avaliar o desempenho da técnica das variáveis de controlo generalizam-se adequadamente as medidas de desempenho denominadas por razão entre variâncias e factor de perda introduzidas por Lavenberg, Moeller e Welch [69] (no contexto dos metamodelos lineares de simulação).

Ao longo deste trabalho, salientou-se a importância dos metamodelos não lineares seguros em estudos de simulação com acontecimentos discretos. No exemplo discutido na Secção 6.1, é obtido um mau ajuste, quando se utilizam diversos metamodelos polinomiais, o que leva a desejar utilizar metamodelos mais flexíveis e mais precisos. Os metamodelos lineares são muito mais

simples de ajustar que os metamodelos não lineares, contudo não são capazes de assegurar um ajuste global a curvas de forma arbitrária. Os metamodelos não lineares de regressão são vantajosos, porque não têm esta limitação, permitindo um ajuste adequado de curvas complexas. No exemplo da Secção 6.1, são comparados entre si, no contexto da regressão não linear, diversas curvas de crescimento sigmóide e a função arco cuja tangente. No exemplo tratado na Secção 6.2, consideram-se e comparam-se duas curvas não lineares (uma delas é uma função hiperbólica). Neste caso, uma das curvas é linealizável e recorre-se à regressão linear para estimar os seus parâmetros desconhecidos.

Em todos os casos tratados neste trabalho, a variância dos erros não é constante: situação mais realista. Para ultrapassar esta dificuldade, seguiram-se dois tipos de abordagem:

(i) utilizou-se o método dos mínimos quadrados ponderados estimados; e

(ii) aplicou-se a técnica da transformação de resultados.

O primeiro caso tem a desvantagem de, como os pesos não se conhecem exactamente eles terem de ser estimados. Os estimadores destes pesos são valores que não são independentes das respostas. No segundo caso, não existe esse problema, mas os parâmetros do metamodelo linearizado não são em geral tão importantes como os parâmetros originais e pode acontecer que o metamodelo transformado já não goze da propriedade da aditividade dos erros.

É de extrema importância ter um metamodelo preciso e pronto a utilizar, em vez de um modelo mais caro e difícil de calibrar como o modelo de simulação. Com o objectivo de assegurar que um metamodelo específico é um bom substituto para o modelo de simulação, tem de ser efectuada uma série de testes de validação. Uma parte destes testes tem a ver com a adequação do metamodelo e a outra parte com a validação do metamodelo com respeito ao modelo de simulação. Sempre que possível, também convém ser testada a validade do metamodelo em relação ao problema real. Se algum destes testes falhar, o metamodelo é rejeitado. No caso tratado na Secção 6.1 e em relação à adequação do metamodelo, as estatísticas SSE e PRESS levam à rejeição dos metamodelos lineares. A generalização do teste de falta de ajustamento de Rao, para metamodelos não lineares, foi utilizado para refinar o processo de selecção da curva, permitindo eleger o metamodelo que melhor se ajusta: o metamodelo 'MMF', no exemplo. A validade do metamodelo 'MMF' em relação ao modelo de simulação foi provada utilizando o teste de validação cruzada dupla. No exemplo

tratado na Secção 6.2, o metamodelo linearizado é rejeitado com base nos dois testes efectuados. No entanto, a rejeição é mais evidente quando se aplica o teste de validação cruzada para metamodelos não lineares, proposto neste trabalho, que é a generalização do método de validação cruzada de Kleijnen. Com base no teste F de falta de ajustamento, também se rejeita o metamodelo linearizado. Neste problema é possível mostrar, e mostra-se, a validade do metamodelo em relação ao problema real. Finalmente, a validade com respeito ao modelo de simulação é mostrada através do teste de validação cruzada dupla, utilizado no exemplo da Secção 6.1.

A utilização de metamodelos não lineares requer um catálogo de curvas extenso e um software de regressão mais complexo e lento. A selecção de uma boa curva influencia dramaticamente a precisão do metamodelo resultante, como foi mostrado nos exemplos tratados no Capítulo 6. Contudo, se for fornecido um catálogo de curvas abrangente, a escolha é relativamente simples. O software de regressão e validação pode ser reutilizado repetidamente, assim que o utilizador forneça uma função para realizar a curva hipotética escolhida e um valor inicial. Finalmente, o aumento do tempo de computação, quando comparado com os procedimentos de regressão linear, torna-se menos importante com o aumento crescente do poder de cálculo dos computadores actuais. Todavia, o tempo de computação, necessário para desenvolver um metamodelo, pode ser diversas ordens de grandeza mais pequeno que o requerido para executar um programa de simulação. Um dos problemas no cálculo do estimador dos mínimos quadrados resulta da necessidade de utilizar um método numérico para minimizar funções não lineares. Este problema é facilmente ultrapassado porque, hoje em dia, existem boas realizações deste tipo de métodos, em pacotes de programas acessíveis ao utilizador. Por exemplo, neste trabalho utilizou-se o método de Levenberg-Marquardt realizado, disponível no pacote MATLAB.

7.2 Trabalho futuro

Não estão esgotadas as oportunidades de investigação na área de metamodelos não lineares de simulação. De facto, foi coberta uma pequena zona desta vasta área. Por exemplo, a validação do metamodelo em relação ao problema real ainda foi muito pouco investigada na literatura. Também

é interessante fazer a generalização desta metodologia para metamodelos em que a resposta tem diversas componentes, ou seja, metamodelos multivariados. Dois assuntos que merecem investigação, são a aplicação do método das variáveis de controlo enviesadas e a aplicação do método das variáveis de controlo correlacionadas, no contexto dos metamodelos não lineares de simulação. Seria útil também aumentar o catálogo de curvas, presente no Apêndice A. No caso dos metamodelos lineares, é importante ter uma distribuição de erros aproximadamente normal, porque a hipótese de normalidade permite obter inferências exactas em amostras pequenas. Contudo, com os metamodelos não lineares, depende-se quase inteiramente de obter inferências aproximadas, através da aplicação de resultados assintóticos a amostras finitas. Por conseguinte, torna-se importante fazer um estudo sobre a importância relativa das hipóteses de normalidade e da variância constante dos erros.

Apêndice A

Catálogo de curvas

Existem variadíssimas curvas que têm sido estudadas, ao longo do tempo, nos mais diversos domínios da ciência. O objectivo deste apêndice é apresentar um catálogo de curvas unidimensionais, encontradas mais frequentemente na bibliografia, que ajude o experimentador, na sua escolha do metamodelo hipotético [107, 6, 45, 11, 39, 53, 49, 28, 79]. Este catálogo não é exaustivo e, por conseguinte, é útil aumentá-lo posteriormente; por exemplo, é importante a inclusão de curvas bidimensionais para que não seja preciso trabalhar com projecções, no caso da regressão não linear bivariada. Com o objectivo de facilitar a consulta deste catálogo, encontram-se na tabela A.1 os diversos tipos de curvas e o local onde podem ser encontradas. A terminologia utilizada nas figuras, que caracteriza o tipo de linha utilizada no traçado de cada curva, é a seguinte:

<i>tipo de linha</i>	<i>designação</i>
<i>sólida</i>	<i>sól</i>
<i>tracejada</i>	<i>trac</i>
<i>pontilhada</i>	<i>pont</i>
<i>combinada com tracejado e pontos</i>	<i>comb</i>

Tabela A.1: Catálogo de curvas.

Tipo de curva	Figuras	Página
exponenciais	A.1	169
com potência de x desconhecida	A.2	170
com x como potência	A.3	170
com funções trigonométricas	A.4	171
funções racionais	A.5	172
de crescimento Sigmóide	A.6 e A.7	173 e 174
de densidade de produção	A.8	175
outras curvas	A.9	175

$$y = \theta_1 + \theta_2 e^{\theta_3 x^{\theta_4}}$$

$$y = \theta_1 + \theta_2 e^{\theta_3/x}$$

$$y = \theta_1(1 - e^{-\theta_2 x})$$

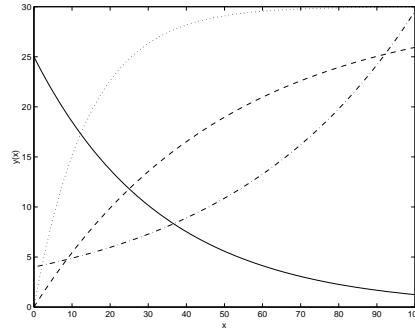
$$y = \theta_1(e^{-\theta_3 x} + e^{-\theta_2 x})$$

$$y = \theta_1(e^{-\theta_3 x} - e^{-\theta_2 x})$$

$$y = \theta_1(e^{-\theta_3(x-\theta_4)} - e^{-\theta_2(x-\theta_4)})$$

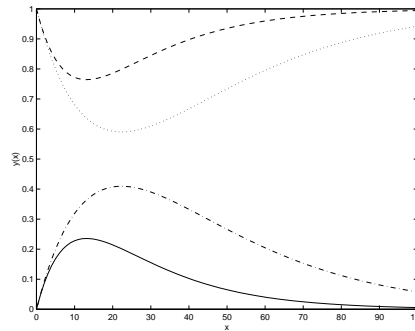
$$y = \theta_1(1 + e^{-\theta_2 x})$$

$$y = (e^{-\theta_2 x} - e^{-\theta_1 x})\theta_1/(\theta_1 - \theta_2)$$



$$y = 4e^{0.02x} \text{ (comb)}; y = 25e^{-0.03x} \text{ (sól)};$$

$$y = 30(1 - e^{-0.07x}) \text{ (pont)}; y = 30(1 - e^{-0.02x}) \text{ (trac)}.$$



$$y = 0.05(e^{-0.04x} - e^{-0.05x})/(0.05 - 0.04) \text{ (comb)}$$

$$y = 0.05(e^{-0.11x} - e^{-0.05x})/(0.05 - 0.11) \text{ (sól)}$$

$$y = 1 - 0.05(e^{-0.04x} - e^{-0.05x})/(0.05 - 0.04) \text{ (pont)}$$

$$y = 1 - 0.05(e^{-0.11x} - e^{-0.05x})/(0.05 - 0.11) \text{ (trac)}$$

Figura A.1: Curvas exponenciais.

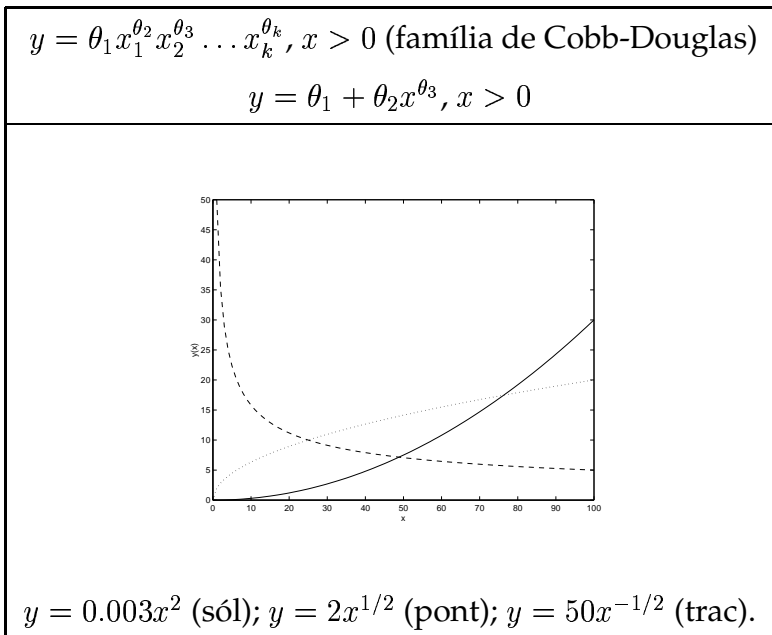


Figura A.2: Curvas com potência de x desconhecida.

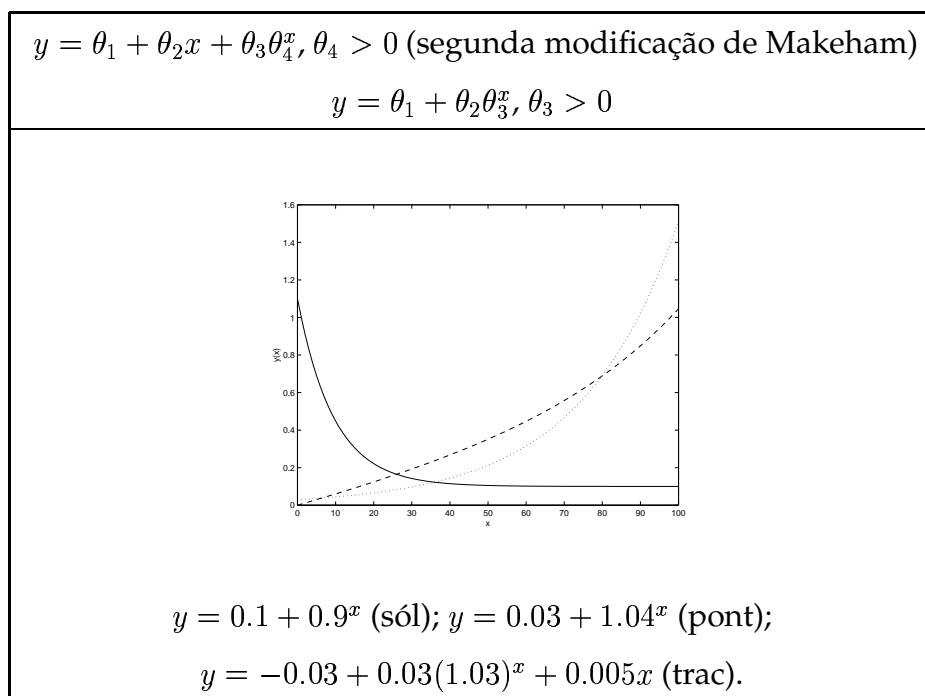
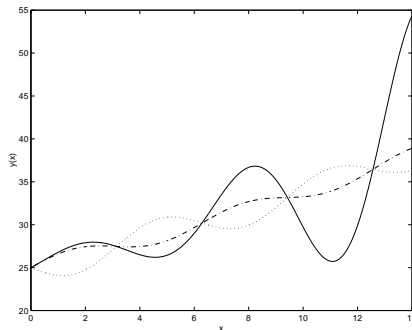


Figura A.3: Curvas com x como potência.

$$y = \theta_1 + \theta_2 e^{\theta_3 x} + (\theta_4 + \theta_5 e^{\theta_6 x}) \text{sen}(\theta_7 + \theta_8 x)$$

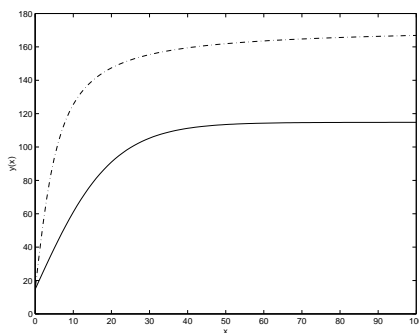
(onda sinusoidal exponencialmente amortecida)



$$y = 25e^{0.03x} + e^{0.2x} \text{sen}(x) \text{ (sól)}; y = 25e^{0.03x} - 2e^{-0.01x} \text{sen}(x) \text{ (pont)};$$

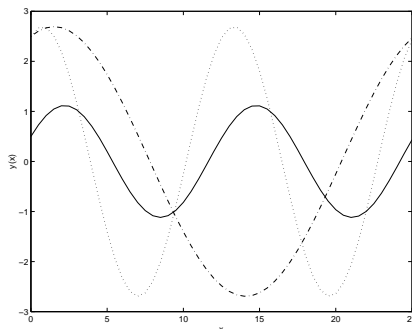
$$y = 25e^{0.03x} + e^{-0.01x} \text{sen}(x) \text{ (comb).}$$

$$y = \theta_1 + \theta_2 \text{arctg}(\theta_3 x) \text{ e } y = \theta_1 + \theta_2 \text{tgh}(\theta_3 x) \text{ (curvas de Wilcoxon)}$$



$$y = 14.8 + 100 \text{tgh}(0.05x) \text{ (sól)}; y = 14.8 + 100 \text{arctg}(0.2x) \text{ (comb).}$$

$$y = \theta_1 + \theta_2 \cos(\theta_4 x) + \theta_3 \text{sen}(\theta_4 x) \text{ (curva periódica)}$$



$$y = 0.5 \cos(0.5x) + \text{sen}(0.5x) \text{ (sól)}; y = 2.5 \cos(0.5x) + \text{sen}(0.5x) \text{ (pont)};$$

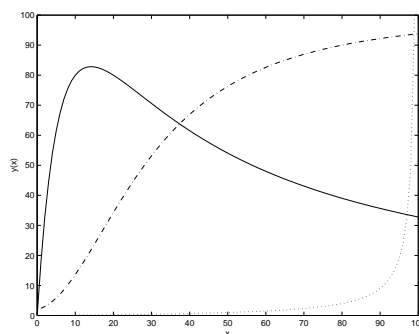
$$y = 2.5 \cos(0.25x) + \text{sen}(0.25x) \text{ (comb).}$$

Figura A.4: Curvas com funções trigonométricas.

$$y = (\theta_0 x) / (\theta_1 + x + \theta_3 x^2) \text{ (curva de Elliot e D. Peirsen)}$$

$$y = (\theta_1 x) / (\theta_2 + x) \text{ (curva de Michaelis-Menten)}$$

$$y = \theta_1 + \theta_2 \theta_3 x / (1 + \theta_3 x) \text{ (curva de Wilcoxson)}$$

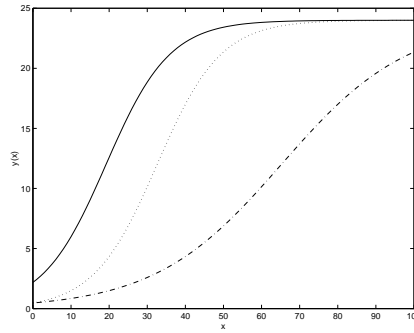


$$y = 200x / (10 + x + 0.05x^2) \text{ (sól)}; y = 0.01x / (1 - 0.01x) \text{ (pont)};$$

$$y = (2 + 0.3x + 0.1x^2) / (1 + 0.001x^2) \text{ (comb)}.$$

Figura A.5: Algumas funções racionais.

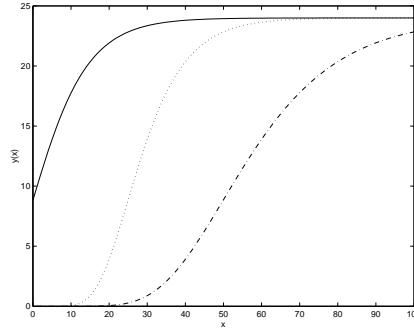
$$y = \theta_1 / (1 + \theta_2 e^{-\theta_3 x}) \text{ (curva logística)}$$



$$y = 24 / (1 + 10e^{-0.12x}) \text{ (sól); } y = 24 / (1 + 50e^{-0.12x}) \text{ (pont);}$$

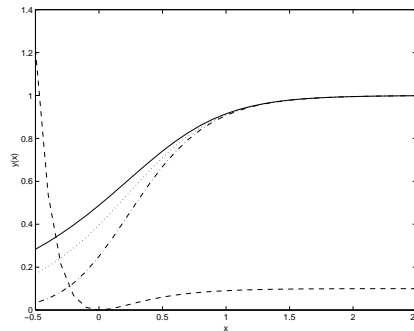
$$y = 24 / (1 + 50e^{-0.06x}) \text{ (comb).}$$

$$y = \theta_1 e^{-\theta_2 e^{-\theta_3 x}} \text{ (curva de crescimento de Gompertz)}$$



$$y = 24e^{-e^{-0.12x}} \text{ (sól); } y = 24e^{-20e^{-0.12x}} \text{ (pont); } y = 24e^{-20e^{-0.06x}} \text{ (comb).}$$

$$y = \theta_0 [1 + (\delta - 1)\theta_1 e^{-\theta_2 x}]^{1/(1-\delta)}, \delta \neq 1 \text{ (curva de Richard)}$$

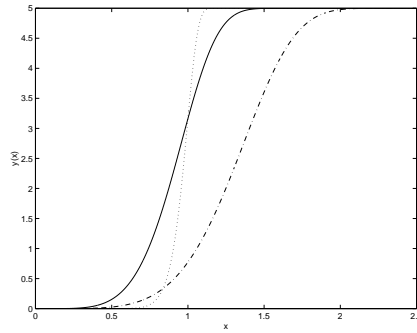


$$y = [1 + 2(\delta - 1)e^{-3x}]^{1/(1-\delta)} \text{ para } \delta = 3.5 \text{ (sól), } \delta = 2.5 \text{ (pont),}$$

$$\delta = 1.5 \text{ (comb), } \delta = 0.5 \text{ (trac).}$$

Figura A.6: Curvas de crescimento de Sigmóide.

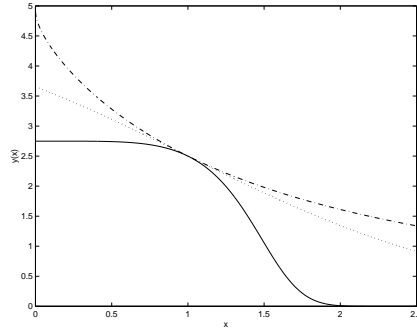
$$y = \theta_1(1 - e^{-[\theta_2(x-\theta_3)]^{\theta_4}}) \text{ e } y = \theta_1 - (\theta_1 - \theta_2)e^{(-\theta_3x)^{\theta_4}} \text{ (curvas de Weibull)}$$



$$y = 5 - 5e^{-x^5} \text{ (sól); } y = 5 - 5e^{-x^{15}} \text{ (pont); } y = 5 - 5e^{-(0.7x)^5} \text{ (comb)}$$

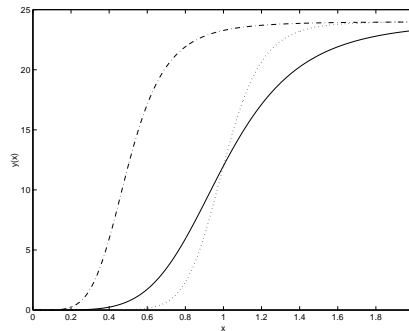
$$y = \theta_1/(1 + e^{\theta_2 + \theta_3g(x)}) \text{ (modelo logístico generalizado)}$$

$$g(x) = \beta_1x + \beta_2x^2 + \beta_3x^3 \text{ ou } g(x) = \frac{x^\lambda - 1}{\lambda}, \lambda \neq 0 \text{ (Box-Cox)}$$



$$y = 5/(1 + e^{(x^\lambda - 1)/\lambda}) \text{ com } \lambda = 5 \text{ (sól); } \lambda = 1 \text{ (pont); } \lambda = 0.2 \text{ (comb)}$$

$$y = \theta_2 - (\theta_2 - \theta_1)/(1 + (\theta_3x)^{\theta_4}) \text{ (família de Morgan-Mercer-Flodin)}$$



$$y = 24 - 24/(1 + x^5) \text{ (sól); } y = 24 - 24/(1 + x^{10}) \text{ (pont);}$$

$$y = 24 - 24/(1 + (2x)^5) \text{ (comb).}$$

Figura A.7: Curvas de crescimento de Sigmóide (continuação).

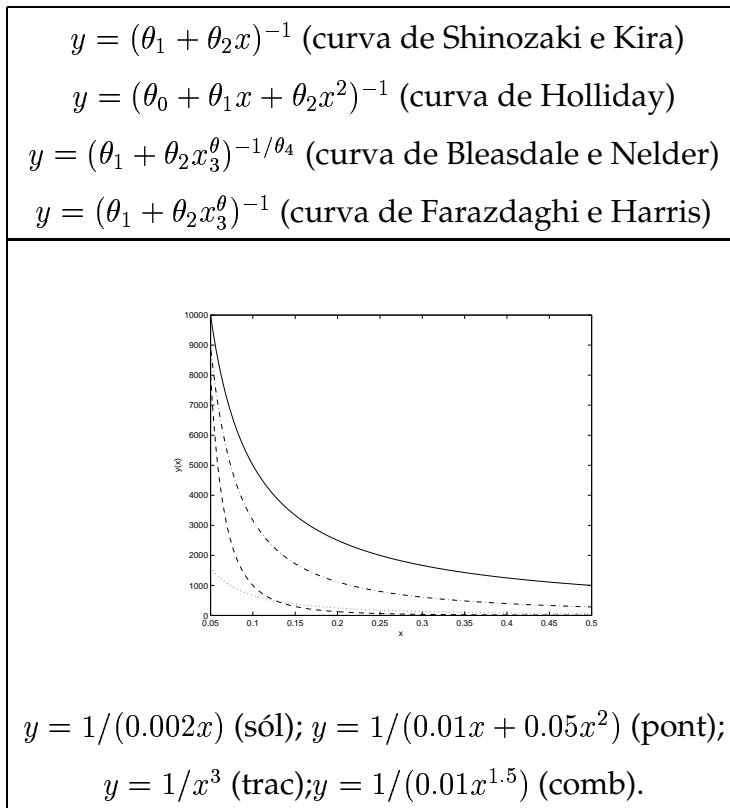


Figura A.8: Curvas de densidade de produção

$$y = \theta_1 + \theta_2 \log(x - \theta_3)$$

$$y = \theta_1 e^{\theta_2 x} x^{\theta_3}, x > 0$$

$$y = \theta_1(x + e^{-\theta_2 x}) - \theta_3 x e^{-\theta_2 x}$$

$$y = 1 - e^{-\theta_1 x} - \theta_1 x e^{-\theta_1 x}$$

$$y = \theta_1 + \theta_2(x - \theta_4)^{-1} + \theta_3(x - \theta_4)^{-2}$$

$$y = e^{-x}(\theta_1 + a_2 x + \theta_3 x^2 + x^3)/[x(\theta_4 + \theta_5 x + \theta_6 x^2 + x^3)]$$

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

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 + \theta_5/(x - 60)$$

Figura A.9: Função logarítmica e algumas curvas combinadas.

Bibliografia

- [1] C. Alexopoulos. *A review of advanced methods for simulation output analysis*. In D. A. Sadowski, J. D. Tew, S. Manivannan e A. F. Seila, editores, *Proceedings of the Winter Simulation Conference*, páginas 133–140, 1994.
- [2] D. M. Allen. *The relationship between variable selection and data augmentation and a method of prediction*. *Technometrics*, 16:125–127, 1974.
- [3] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York, NY, USA, 1984.
- [4] S. Andradóttir, J. M. Calvin e P. W. Glynn. *Increasing the frequency of regeneration for Markov processes*. In D. A. Sadowski, J. D. Tew, S. Manivannan e A. F. Seila, editores, *Proceedings of the Winter Simulation Conference*, páginas 320–323, 1994.
- [5] S. F. Arnold. *The Theory of Linear Models and Multivariate Analysis*. John Wiley and Sons, New York, NY, USA, 1981.
- [6] A. C. Atkinson e A. N. Donev. *Optimum Experimental Designs*. Clarendon Press, Oxford, New York, USA, 1992.
- [7] A. N. Avramidis e J. R. Wilson. *Integrated variance reduction strategies for simulation*. *Operations Research*, 1994.
- [8] A. B. Badiru e D. B. Sieger. *Neural network as a simulation metamodel in economic analysis of risky projects*. *European Journal of Operational Research*, 105:130–142, 1998.

- [9] R. R. Barton. *Metamodels for simulation input-output relations*. In J. J. Swain, D. Goldsman, R. C. Crain e J. R. Wilson, editores, *Proceedings of the Winter Simulation Conference*, páginas 289–299, 1992.
- [10] R. R. Barton. *Metamodeling: A state of the art review*. In J. D. Tew, S. Manivannan, D. A. Sadowski e A. F. Seila, editores, *Proceedings of the Winter Simulation Conference*, páginas 237–244, 1994.
- [11] D. M. Bates e D. G. Watts. *Nonlinear Regression Analysis and its Applications*. John Wiley and Sons, New York, NY, USA, 1988.
- [12] D. P. Bischak, W. D. Kelton e S. M. Pollock. *Weighted batch means for confidence intervals in steady-state simulations*. *Management Science*, 39(8):1002–1019, Agosto 1993.
- [13] R. B. Blackman e J. W. Tukey. *The measurement of power spectra from the point of view of communications engineering (part 1 and part 2)*. *Bell Sys. Tech. Journal*, 37:185–282, 485–569, 1958. ***.
- [14] M. Broadie e P. Glasserman. *Estimating security price derivatives using simulation*. *Management Science*, 42(2):269–277, Fevereiro 1996.
- [15] J. M. Calvin. *Covariance of regenerative mean and variance estimators for the Markov chains*. In *Proceedings of the Winter Simulation Conference*, páginas 473–475, Dezembro 1988. ***.
- [16] J. M. Calvin e M. K. Nakayama. *Using permutations in regenerative simulations to reduce variance*. *ACM Transactions on Modeling and Computer Simulation*, 8(2):153–193, Abril 1998.
- [17] R. J. Carrol e D. Ruppert. *Transformation and Weighting in Regression*. Chapman and Hall, New York, USA, 1988.
- [18] J. S. Carson e A. M. Law. *Conservation equations and variance reduction in queueing simulations*. *Operations Research*, 37:535–546, 1980.

- [19] D. Caughlin. *Automating the metamodeling process*. In S. Andradóttir, K. J. Healy, D. H. Withers e B. L. Nelson, editores, *Proceedings of the Winter Simulation Conference*, páginas 978–985, 1997.
- [20] J. M. Charnes e W. D. Kelton. *Multivariate autoregressive techniques for constructing confidence regions on the mean vector*. *Management Science*, 39(9):1112–1129, Setembro 1993.
- [21] R. C. H. Cheng e J. P. C. Kleijnen. *Improved design of queueing simulation experiments with highly heteroscedastic responses*. *Operations Research*, 47(5):762–777, Setembro 1999.
- [22] C. Chien. *Batch size selection for the batch means method*. In D. A. Sadowski, J. D. Tew, S. Manivannan e A. F. Seila, editores, *Proceedings of the Winter Simulation Conference*, páginas 345–352, 1994.
- [23] R. W. Conway, B. M. Johnson e W. C. Maxwell. *Some problems of digital systems simulation*. *Management Science*, 6:92–110, 1959. ***.
- [24] D. R. Cox e S. W. L. Smith. *Queues*. *John Wiley and Sons, New York, NY, USA*, 1961. ***.
- [25] M. A. Crane e D. L. Iglehart. *Simulating stable stochastic systems: I. general multiserver queues*. *Journal of the ACM*, 21:103–113, 1974. ***.
- [26] M. A. Crane e D. L. Iglehart. *Simulating stable stochastic systems: III. regenerative processes and discrete-event simulations*. *Operations Research*, 23:33–45, 1975. ***.
- [27] J. W. Crawford e T. J. Gallwey. *Bias and variance reduction in computer simulation studies*. *European Journal of Operational Research*, 124:571–590, 2000.
- [28] C. Daniel e F. S. Wood. *Fitting Equations to Data, Computer Analysis of Multifactor Data*. *John Wiley and Sons, New York, NY, USA, segunda edição*, 1980.

- [29] J. M. Donohue, E. C. Houck e R. H. Myers. *A sequential experimental design procedure for the estimation of first- and second-order simulation metamodels*. *ACM Transactions on Modeling and Computer Simulation*, 3(3):190–224, Julho 1993.
- [30] G. S. Fishman. *Problems in the statistical analysis of simulation experiments: The comparison of means and the length of sample records*. *Communications of the ACM*, 10:94–99, 1967. ***.
- [31] G. S. Fishman. *Estimating sample size in computing simulation experiments*. *Management Science*, 18:21–38, 1971. ***.
- [32] G. S. Fishman. *Concepts and Methods in Discrete Event Digital Simulation*. *John Wiley and Sons, New York, NY, USA, 1973*. ***.
- [33] G. S. Fishman. *Statistical analysis for queuing simulation*. *Management Science*, 20:363–369, 1973. ***.
- [34] G. S. Fishman. *Estimation in multiserver queueing simulations*. *Operations Research*, 22:72–78, 1974. ***.
- [35] G. S. Fishman. *Principles of Discrete Event Simulation*. *John Wiley and Sons, New York, NY, USA, 1978*. ***.
- [36] G. S. Fishman e P. J. Kiviat. *The analysis of simulation-generated time series*. *Management Science*, 13:525–557, 1967. ***.
- [37] L. W. Friedman e H. H. Friedman. *Validating the simulation metamodel: Some practical approaches*. *Simulation*, 45(3):144–146, 1985.
- [38] A. V. Gafarian, C. J. Anker e T. Morisaku. *Evaluation of commonly used rules for detecting steady-state in computer simulation*. *Naval Res. Logist. Quart.*, 25:511–529, 1978.
- [39] A. R. Gallant. *Nonlinear Statistical Models*. *John Wiley and Sons, New York, NY, USA, 1987*.

- [40] P. Glasserman e D. D. Yao. *Some guidelines and guarantees for common random numbers*. *Management Science*, 38(6):884–908, 1992.
- [41] P. W. Glynn e D. L. Iglehart. *Notes: Conditions for the applicability of the regenerative method*. *Management Science*, 39(9):1108–1111, Setembro 1993.
- [42] P. W. Glynn e W. Whitt. *Indirect estimation via $l = \lambda w$* . *Operations Research*, 37(1):82–103, Janeiro 1989.
- [43] P. W. Glynn e W. Whitt. *The asymptotic efficiency of simulation estimators*. *Operations Research*, 40(3):505–519, Maio 1992.
- [44] D. Gross e C. M. Harris. *Fundamentals of Queueing Theory*. *John Wiley and Sons*, New York, NY, USA, segunda edição, 1985.
- [45] L. C. Hamilton. *Regression with Graphics. A Second Course in Applied Statistics*. *Wadsworth*, Belmont, California, 1992.
- [46] J. M. Hammersley e D. C. Handscomb. *Monte Carlo Methods*. *Methuen*, London, 1964.
***.
- [47] S. G. Henderson e P. W. Glynn. *Can the regenerative method be applied to discrete-event simulation?* In P. A. Farrington, H. B. Nembhard, D. T. Sturrock e G. W. Evans, editores, *Proceedings of the Winter Simulation Conference*, páginas 367–373, 1999.
- [48] S. G. Henderson e S. P. Meyn. *Efficient simulation of multiclass queueing networks*. In S. Andradóttir, K. J. Healy, D. H. Withers e B. L. Nelson, editores, *Proceedings of the Winter Simulation Conference*, páginas 216–223, 1997.
- [49] D. M. Himmelblau. *Process Analysis by Statistical Methods*. *John Wiley and Sons*, New York, NY, USA, 1970.
- [50] R. D. Hurrion. *Using a neural network to enhance the decision making quality of a visual interactive simulation model*. *Journal of the Operational Research Society*, 43(4):333–341, 1992.

- [51] J. R. Hussey, R. H. Myers e E. C. Houck. *Correlated simulation experiments in first-order response surface design*. *Operations Research*, 35(5):744–758, Setembro 1987.
- [52] W. D. Kelton e A. M. Law. *An analytical evaluation of alternative strategies in steady-state simulation*. *Operations Research*, 32:169–184, 1984.
- [53] A. I. Khuri e J. A. Cornell. *Response Surfaces, Designs and Analyses*. *Marcel Dekker, New York*, 1987.
- [54] J. P. C. Kleijnen. *Statistical Techniques in Simulation*. *Marcel Dekker, New York*, 1974.
- [55] J. P. C. Kleijnen. *Design and analysis of simulations: practical statistical techniques*. *Simulation*, 28(3):81–90, 1977.
- [56] J. P. C. Kleijnen. *The role of statistical methodology in simulation*. In B. P. Zeigler, M. S. Elzas, G. J. Klir e T. I. Oren, editores, *Methodology in Systems Modeling and Simulation*. *North-Holland, Amsterdam*, 1979. ***.
- [57] J. P. C. Kleijnen. *Regression analysis for simulation practitioners*. *Journal of the Operational Research Society*, 32:35–43, 1981.
- [58] J. P. C. Kleijnen. *Experimentation with models: Statistical design and analysis techniques*. In F. E. Cellier, editor, *Progress in Modeling & Simulation*. *Academic Press, New York, NY, USA*, 1982.
- [59] J. P. C. Kleijnen. *Cross-validation using the t statistic*. *European Journal of Operational Research*, 13:133–141, 1983.
- [60] J. P. C. Kleijnen. *Statistical Tools for Simulation Practitioners*. *Marcel Dekker, New York*, 1987.
- [61] J. P. C. Kleijnen. *Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals*. *Management Science*, 38(8):1164–1185, Agosto 1992.
- [62] J. P. C. Kleijnen e B. Annink. *Vector computers, Monte Carlo simulation and regression analysis: An introduction*. *Management Science*, 38(2):170–181, Fevereiro 1992.

- [63] J. P. C. Kleijnen, R. Brent e R. Brouwers. *Small-sample behavior of weighted least squares in experimental design applications*. *Communications in Statistics B - Simulation and Computation*, 10(3):303–313, 1981.
- [64] J. P. C. Kleijnen e W. V. Groenendaal. *Simulation, A Statistical Perspective*. *John Wiley and Sons, New York, NY, USA*, 1992.
- [65] J. P. C. Kleijnen e R. G. Sargent. *A methodology for fitting and validating metamodels in simulation*. *European Journal of Operational Research*, 120:14–29, 2000.
- [66] J. P. C. Kleijnen, A. J. van den Burg e R. Th. van der Ham. *Generalization of simulation results, practicality of statistical methods*. *European Journal of Operational Research*, 3:50–64, 1979.
- [67] C. H. Kuei e C. N. Madu. *Polynomial metamodeling and Taguchi designs in simulation with application to the maintenance float system*. *European Journal of Operational Research*, 72:364–375, 1994.
- [68] C. Kwon e J. D. Tew. *Strategies for combining antithetic variates and control variates in designed simulation experiments*. *Management Science*, 40(8):1021–1034, Agosto 1994.
- [69] S. S. Lavenberg, T. L. Moeller e P. D. Welch. *Statistical results on control variables with application to queueing network simulation*. *Operations Research*, 30:182–202, 1982.
- [70] S. S. Lavenberg e P. D. Welch. *A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations*. *Management Science*, 27(3):322–335, Março 1981.
- [71] A. M. Law. *Confidence interval in discrete event simulation: A comparison of replication and batch means*. *Naval Research Logistics Quarterly*, 24:667–678, 1977. ***.
- [72] A. M. Law e W. D. Kelton. *Confidence intervals for steady-state simulations: I. a survey of fixed sample size procedures*. *Operations Research*, 32(6):1221–1239, 1984.

- [73] A. M. Law e W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, NY, USA, segunda edição, 1991.
- [74] R. W. Lawless, L. H. Williams e C. G. Richie. *A sensitivity analysis tool for simulation with application to disaster planning*. *Simulation*, páginas 217–223, 1972.
- [75] P. L'Ecuyer. *Efficiency improvement and variance reduction*. In J. D. Tew, S. Manivannan, D. A. Sadowski e A. F. Seila, editores, *Proceedings of the Winter Simulation Conference*, páginas 122–132, 1994.
- [76] Y. H. Lee e H. S. Oh. *Detecting truncation point in steady-state simulation using chaos theory*. In J. D. Tew, S. Manivannan, D. A. Sadowski e A. F. Seila, editores, *Proceedings of the Winter Simulation Conference*, páginas 353–360, 1994.
- [77] M. Livny, B. Melamed e A. K. Tsiolis. *The impact of autocorrelation on queuing systems*. *Management Science*, 39(3):322–339, Janeiro 1993.
- [78] W. W. Loh. *On the Method of Control Variates*. *PhD thesis, Department of Operations Research, Stanford University, Stanford CA, 1994*. <http://www-or.stanford.edu/simlib/theses.html>.
- [79] R. L. Mason, R. F. Gunst e J. L. Hess. *Statistical Design & Analysis of Experiments*. John Wiley and Sons, New York, NY, USA, 1989.
- [80] V. B. Melas. *On the efficiency of the slitting and roulette approach for sensitivity analysis*. In S. Andradóttir, K. J. Healy, D. H. Withers e B. L. Nelson, editores, *Proceedings of the Winter Simulation Conference*, páginas 269–274, 1997.
- [81] D. J. Morrice. *The effects of batching on the power of the test for frequency domain methodology*. In B. L. Nelson, W. D. Kelton e G. M. Clark, editores, *Proceedings of the Winter Simulation Conference*, páginas 1018–1022, 1991.
- [82] D. J. Morrice e L. W. Schruben. *A frequency domain metamodeling approach to transient sensitivity analysis*. *IIE Transactions*, páginas 229–244, Março 2001.

- [83] R. H. Myers. *Classical and Modern Regression with Applications*. Duxbury Press, Boston, 1990.
- [84] J. W. Neill e D. E. Johnson. *Testing for lack of fit in regression – a review*. *Communications in Statistics A - Theoretical Methods*, 13(4):485–511, 1984.
- [85] B. L. Nelson. *Control variate remedies*. *Operations Research*, 38(6):974–992, Dezembro 1990.
- [86] B. L. Nelson e B. W. Schmeiser. *Decomposition of some well-known variance reduction techniques*. *Journal of Statistical Computation and Simulation*, 23:183–209, 1986.
- [87] J. Neter, W. Wasserman e M. H. Kutner. *Applied Linear Regression Models*. Richard D. Irwin Inc., segunda edição, 1989.
- [88] A. M. Porta Nova. *A Generalized Approach to Variance Reduction in Discrete-Event Simulation Using Control Variables*. *PhD thesis, Mechanical Engineering Department, University of Texas, Austin, Maio 1985*.
- [89] A. M. Porta Nova e J. R. Wilson. *Estimation of multiresponse simulation metamodels using control variates*. *Management Science*, 35(11):1316–1333, Novembro 1989.
- [90] A. M. Porta Nova e J. R. Wilson. *Selecting control variates to estimate multiresponse simulation metamodels*. *European Journal of Operational Research*, 71:80–94, 1993.
- [91] A. Nozari. *On Simulation Experiments for Multipopulation Problems*. *PhD thesis, Pennsylvania State University, Agosto 1982*.
- [92] A. Nozari, S. F. Arnold e C. D. Pegden. *Control variates for multipopulation simulation experiments*. *IIE Transactions*, 16:159–169, Junho 1984.
- [93] A. B. Owen. *Latin supercube sampling for very high-dimensional simulations*. *ACM Transactions on Modeling and Computer Simulation*, 8(1):71–102, Janeiro 1998.
- [94] R. P. Panis, R. H. Myers e E. C. Houck. *Combining regression diagnostics with simulation metamodels*. *European Journal of Operational Research*, 13:85–94, 1994.

- [95] K. Pawlikowski. *Steady-state simulation of queueing processes: A survey of problems and solutions*. ACM Computing Surveys, 22(2):123–170, Junho 1990.
- [96] A. A. B. Pritsker. *The GASP IV Simulation Language*. John Wiley and Sons, New York, NY, USA, 1974. ***.
- [97] C. R. Rao. *Some problems involving linear hypotheses in multivariate analysis*. Biometrika, 46:49–58, 1959.
- [98] M. R. Reynolds e M. L. Deaton. *Comparisons of some tests for validation of stochastic simulation models*. Communications in Statistics B - Simulation and Computation, 1981.
- [99] R. Y. Rubinstein e R. Marcus. *Efficiency of multivariate control variates in Monte Carlo simulation*. Operations Research, 33(3):661–677, May-June 1985.
- [100] M. I. Santos e A. M. Porta Nova. *The main issues in nonlinear simulation metamodel estimation*. In P. A. Farrington, H. B. Nembhard, D. T. Sturrock e G. W. Evans, editores, Proceedings of the Winter Simulation Conference, páginas 502–509, 1999.
- [101] M. I. Santos e A. M. Porta Nova. *Validation of nonlinear simulation metamodels*. In M. H. Hamza, editor, Proceedings of Applied Simulation and Modeling, páginas 421–425, 2001.
- [102] B. W. Schmeiser, M. R. Taaffe e J. Wang. *Biased control-variate estimation*. IIE Transactions, 33(3):219–228, Março 2001.
- [103] T. J. Schriber e R. W. Andrews. *Two ARMA-based confidence-interval procedures for the analysis of simulation output*. American Journal of Mathematical and Management Sciences, 1985. ***.
- [104] L. W. Schruben. *Confidence interval estimation using standardized time series*. Operations Research, 31(6):1090–1108, 1983.
- [105] S. R. Searle. *Linear Models*. John Wiley and Sons, New York, NY, USA, 1971.

- [106] G. A. F. Seber. *Linear Regression Analysis*. John Wiley and Sons, New York, NY, USA, 1977.
- [107] G. A. F. Seber e C. J. Wild. *Nonlinear Regression*. John Wiley and Sons, New York, NY, USA, 1989.
- [108] A. F. Seila. *Output analysis for simulation*. In B. L. Nelson, W. D. Kelton e G. M. Clark, editores, *Proceedings of the Winter Simulation Conference*, páginas 28–36, 1991.
- [109] Ashish Sen e Muni Srivastava. *Regression Analysis, Theory, Methods, and Applications*. Springer-Verlag, New York, 1990.
- [110] N. H. Shih e W. T. Song. *Estimating simulation metamodels using correlated control variates*. *Communications in Statistics B - Simulation and Computation*, 24(1):79–90, 1995.
- [111] J. Shirish e J. D. Tew. *Validation and statistical analysis procedures under the common random number correlation-induction strategy for multipopulation simulation experiments*. *European Journal of Operational Research*, 85:205–220, 1995.
- [112] M. Stone. *Cross-validation choice and assessment of statistical predictions*. *Journal of the Royal Statistical Society Serie B*, 36:111–147, 1974.
- [113] J. D. Tew e J. R. Wilson. *Validation of simulation analysis methods for the Schruben-Margolin correlation-induction strategy*. *Operations Research*, 40(1):87–103, 1992.
- [114] G. Tokol, D. Goldsman, D. Ockerman e J. Swain. *Standardized time series l_p -norm variance estimators for simulations*. *Management Science*, 44:234–245, Fevereiro 1998.
- [115] Turnquist e Sussman. *Toward guidelines for designing experiments in queuing simulation*. *Simulation*, 32:169–184, 1984. ***.
- [116] S. Venkatraman e J. Wilson. *The efficiency of control variates in multiresponse simulation*. *Operations Research Letters*, 5(1):37–42, Junho 1986.
- [117] D. J. Watson. *The estimation of leaf areas*. *Journal of Agriculture and Science*, 27(474), 1937.

- [118] P. D. Welch. *The statistical analysis of simulation results*. In S. S. Lavenberg, editor, *Computer Performance Modeling Handbook*, páginas 268–328. Academic Press, New York, NY, USA, 1983.
- [119] H. White. *Nonlinear regression on cross-section data*. *Econometrica*, 48(3):721–746, Abril 1980.
- [120] K. P. White. *An effective truncation heuristic for bias reduction in simulation output*. *Simulation*, 69(6):323–334, Dezembro 1997.
- [121] W. Whitt. *The efficiency of one long run versus independent replications in steady-state simulation*. *Management Science*, 37(6):645–666, Junho 1991.
- [122] J. G. Wilson. *A note on variance reduction group maintenance policies*. *Management Science*, 42(3):452–456, Março 1996.
- [123] J. R. Wilson. *Variance reduction techniques for digital simulation*. *American Journal of Mathematical and Management Sciences*, 4:277–312, 1984.
- [124] W. Yang e W. Liou. *Combining antithetic variates and control variates in simulation experiments*. *ACM Transactions on Modeling and Computer Simulation*, 6(4):243–260, Outubro 1996.
- [125] W. N. Yang e B. L. Nelson. *Using common random numbers and control variates in multiple-comparison procedures*. *Operations Research*, 39(4):583–591, Julho 1991.
- [126] C. W. Zobel e K. P. White. *Determining a warm-up period for a telephone network routing simulation*. In P. A. Farrington, H. B. Nembhard, D. T. Sturrock e G. W. Evans, editores, *Proceedings of the Winter Simulation Conference*, páginas 662–665, 1999.