

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA QUÍMICA**

**NÍCOLAS ARTIFON DORNELES**

**AVALIAÇÃO DE DIFERENTES ALTERNATIVAS PARA A DETERMINAÇÃO DE  
REGIÕES DE CONFIANÇA DE ESTIMATIVAS DE PARÂMETROS**

Porto Alegre, RS

2020

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA QUÍMICA

NÍCOLAS ARTIFON DORNELES

AVALIAÇÃO DE DIFERENTES ALTERNATIVAS PARA A DETERMINAÇÃO DE  
REGIÕES DE CONFIANÇA DE ESTIMATIVAS DE PARÂMETROS

Dissertação de Mestrado apresentada como  
requisito parcial para obtenção do título de  
Mestre em Engenharia Química.

Orientador: Prof. Dr. Marcio Schwaab

Porto Alegre, RS

2020

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA QUÍMICA

A comissão examinadora, abaixo assinada, aprova a Dissertação: *Avaliação de diferentes alternativas para a determinação de regiões de confiança de estimativas de parâmetros*. Elaborada por Nicolas Artifon Dorneles como requisito parcial para obtenção do título de Mestre em Engenharia Química.

Comissão examinadora:

---

Prof. Dr. Pedro Rafael Bolognese Fernandes

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Kese Pontes Freitas Alberton

---

Dr.<sup>a</sup> Simoní Da Ros

Dorneles, Nicolás Artifon

Avaliação de diferentes alternativas para a determinação de regiões de confiança de estimativas de parâmetros / Nicolás Artifon Dorneles. -- 2020.

159 f.

Orientador: Marcio Schwaab.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Programa de Pós-Graduação em Engenharia Química, Porto Alegre, BR-RS, 2020.

1. Aproximação Gauss-Newton. 2. Bootstrap. 3. Região do contorno. 4. Região do perfil. 5. Regiões de confiança.  
I. Schwaab, Marcio, orient. II. Título.

Este trabalho é dedicado à minha avó,  
Paulina Theresinha Charão Artifon, a vó  
Paula.

## **AGRADECIMENTOS**

Aos meus pais, Jairo e Silviane, pela estrutura familiar sólida que proporcionaram, permitindo que esta e outras conquistas fossem possíveis e também pelo incentivo para que eu sempre continuasse me aprimorando.

Aos meus avós, Antonio e Clotildes, pelo carinho recebido e pelas queixas de não os visitar com tanta frequência devido a este trabalho.

À minha namorada, Ana Carolina, pelo apoio e incentivo durante todo o tempo de titulação e pelos finais de semana e feriados que passamos em casa para que fosse possível estudar e na elaborar esta dissertação.

Ao meu orientador, Prof. Marcio Schwaab, pela disposição e pelos ensinamentos e esclarecimentos quase que semanais e pelas “lives”, necessárias para manter a continuidade deste trabalho, apesar dos problemas técnicos.

Aos meus queridos amigos, pelo apoio e por tornarem a busca por titulações uma competição amistosa e saudável.

À Capes, pelo apoio financeiro.

A todos que acreditam que a educação de qualidade é a ferramenta que transforma o mundo.

## RESUMO

---

A estimação de parâmetros consiste em determinar o conjunto ótimo dos valores dos parâmetros de modelo específico, o qual corresponde ao menor valor possível para o desvio entre as variáveis medidas experimentalmente e as previstas pelo modelo. Como os valores mensurados experimentalmente estão sujeitos a incertezas, o mesmo acontecerá com os valores estimados, fazendo-se necessário determinar a incerteza destes. Em modelos com um único parâmetro, podem ser definidas faixas ou intervalos de confiança contendo valores estatisticamente iguais ao valor ótimo. Já para modelos que possuem mais de um parâmetro, a abordagem mais adequada é determinar as suas regiões de confiança, pois a provável presença da correlação entre os parâmetros estimados invalida a análise somente por intervalos de confiança. Nesse trabalho foram avaliadas diversas metodologias para a determinação de regiões de confiança. Inicialmente foi avaliado o impacto que o uso da aproximação de Gauss-Newton gera no cálculo das regiões de confiança elípticas quando comparada com o cálculo da matriz Hessiana completa. O uso da aproximação comprovou-se justificável nos casos estudados, já que não foram observadas diferenças significativas entre as regiões resultantes pelo emprego dos dois métodos. Também foi avaliada a qualidade das regiões de confiança obtidas pelo método da razão de verossimilhança. Foi utilizado o método do Contorno que percorre o limite da região de confiança determinando de forma quase exata o seu traçado. Este método funciona bem para uma grande maioria de modelos, mas apresenta algumas dificuldades quando as regiões possuem variações bruscas na sua geometria. Também foi usado o método do Perfil, que consiste na determinação de intervalos de confiança baseados na razão de verossimilhança, seguido de uma interpolação para determinação das regiões de confiança. Entretanto, esta metodologia não conseguiu delimitar regiões de confiança com geometria mais complexa. Por fim, foi utilizado o método de *Bootstrap*, o qual consiste em estimar novos valores ótimos a partir de perturbações normalmente distribuídas nas variáveis experimentais. As regiões geradas por este método possuem pontos que se concentram majoritariamente nas proximidades do ponto ótimo, embora também ocorram dispersões. Apesar do traçado da região de confiança não ficar evidente, estes pontos podem ser utilizados para a avaliação da distribuição de probabilidades das estimativas dos parâmetros.

**Palavras chaves:** Regiões de confiança. Região do perfil. Aproximação de Gauss-Newton. Bootstrap. Região do contorno.

## ABSTRACT

---

Parameter estimation consists of determining the optimal set of parameter values for a specific model, which minimize the difference between the variables measured experimentally and those predicted by the model. As the values measured experimentally are subject to uncertainties, the same will happen with the estimated values, making it necessary to determine their uncertainty. In models with a single parameter, confidence ranges or intervals can be defined containing values statistically equal to the optimum value. For models that have more than one parameter, the most appropriate approach is to determine their confidence regions, since the likely presence of the correlation between the estimated parameters invalidates the analysis only by confidence intervals. In this work, several methodologies for determining regions of trust were evaluated. Initially it was evaluated the impact that the use of the Gauss-Newton approximation generates when considering it in the calculation of the elliptical confidence regions when compared with the calculation of the complete Hessian matrix. The use of the approximation proved to be justifiable, since no significant differences were observed between the regions obtained by using the two methods. The quality of the confidence regions obtained by the likelihood ratio method was also evaluated. For this work, the Contour method was chosen, which crosses the border of the confidence region, determining its contour almost exactly. This method works well for a large majority of models, but presents some difficulties when the regions have sudden variations in their geometry. The Profile method was also used, which consists of determining the confidence interval based on the likelihood ratio, followed by an interpolation to determine the confidence regions. However, this methodology was unable to delimit regions of trust with more complex geometry. Finally, the Bootstrap method was used, which consists of estimating new optimal values from perturbations normally distributed in the experimental variables. The regions generated by this method have points that are mainly concentrated in the vicinity of the optimum point, while some of them are more dispersed. Although the tracing of the confidence region is not evident, these points can be used to assess the probability distribution of the parameter estimates.

**Keywords:** Confidence regions. Contour confidence region. Gauss-Newton method. Bootstrap confidence region. Profile confidence region.

## LISTA DE TABELAS

---

Tabela 1: Números de publicações encontradas envolvendo o uso das regiões de confiança para cada década desde 1960.....	32
Tabela 2: Valores das variáveis independentes e dependentes para o modelo linear da Equação (3.1), Modelo 1.....	34
Tabela 3: Valores das variáveis independentes e dependentes para o modelo não-linear da Equação (3.2), Modelo 2.A.....	35
Tabela 4: Valores das variáveis independentes e dependentes para o modelo não-linear da Equação (3.2), Modelo 2.B.....	35
Tabela 5: Valores das variáveis independentes e dependentes para o modelo não-linear da Equação (3.3), Modelo 3. ....	36
Tabela 6: Valores das variáveis independentes e dependentes para o modelo linear com três parâmetros da Equação (3.4), Modelo 4. ....	37
Tabela 7: Valores das variáveis independentes e dependentes para o modelo não-linear com três parâmetros da Equação (3.5), Modelo 5.A. ....	38
Tabela 8: Valores das variáveis independentes e dependentes para o modelo não-linear com três parâmetros da Equação (3.5), Modelo 5.B. ....	38
Tabela 9: Valores do número de experimentos e de parâmetros de cada modelo utilizado neste trabalho. ....	39
Tabela 10: Valores dos limites inferior e superior utilizados para a distribuição das partículas do enxame no espaço dos parâmetros.....	42
Tabela 11: Valores dos desvios padrão dos parâmetros, de $\tau_{MAX}$ e do valor limite da distribuição $F$ de Fisher para o Modelo 1. ....	66
Tabela 12: Valores dos desvios padrão dos parâmetros, de $\tau_{MAX}$ e do valor limite da distribuição $F$ de Fisher para o Modelo 2.A.....	71
Tabela 13: Valores dos desvios padrão dos parâmetros, de $\tau_{MAX}$ e do valor limite da distribuição $F$ de Fisher para o Modelo 2.B.....	75
Tabela 14: Valores dos desvios padrão dos parâmetros, de $\tau_{MAX}$ e do valor limite da distribuição $F$ de Fisher para o Modelo 4. ....	83
Tabela 15: Valores dos desvios padrão dos parâmetros, de $\tau_{MAX}$ e do valor limite da distribuição $F$ de Fisher para o Modelo 5.A.....	88

Tabela 16: Valores das variáveis independentes e dependentes para o cálculo dos <i>splines</i> .....	115
Tabela 17 Comparação entre os valores da matriz Hessiana analítica e numérica do Modelo 2.B com perturbações iguais a $10^{-7}$ , $10^{-8}$ , e $10^{-9}$ .....	121

## LISTA DE FIGURAS

---

Figura 1: Comparação entre os valores de saída experimentais e os preditos para o Modelo1. ....	51
Figura 2: Regiões de confiança do Modelo 1 com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia). ....	52
Figura 3: Comparação entre os valores de saída experimentais e os preditos para o Modelo 2.A. ....	54
Figura 4: Regiões de confiança do Modelo 2.A com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia). ....	55
Figura 5: Comparação entre os valores de saída experimentais e os preditos para o Modelo 2.B. ....	56
Figura 6: Regiões de confiança do Modelo 2.B com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia). ....	57
Figura 7: Comparação entre ao desvio da linearidade do parâmetro $\theta_1$ do Modelo 2.A, Figura 7.a, e do Modelo 2.B, Figura 7.b. ....	58
Figura 8: Comparação entre ao desvio da linearidade do parâmetro $\theta_2$ do Modelo 2.A, Figura 8.a, e do Modelo 2.B, Figura 8.b. ....	59
Figura 9: Comparação entre os valores de saída experimentais e os preditos para o Modelo 3. ....	61
Figura 10: Regiões de confiança do Modelo 3 com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia). ....	62
Figura 11: Comparação entre as regiões de confiança do Modelo 1 determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.1) (linha cheia). ....	65
Figura 12: Representação dos perfis do Modelo 1 (pontos), dos limites dos perfis (linha tracejada) e da elipse gerada pelo método (linha pontilhada). ....	66

Figura 13: Comparação entre as regiões de confiança do Modelo 1 determinadas pelo método do Perfil (linha pontilhada) e pela Equação (2.34) com os valores da matriz da Equação (4.1) (linha cheia).....	67
Figura 14: Comparação entre as regiões de confiança do Modelo 1 determinadas pelo método Bootstrap (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.1) (linha tracejada). ....	68
Figura 15: Comparação entre as regiões de confiança do Modelo 2.A determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.3) (linha cheia).....	69
Figura 16: Comparação entre as regiões de confiança do Modelo 2.A determinadas pelo método do Perfil (linha pontilhada) e pelo método do Contorno (linha cheia). ....	71
Figura 17: Comparação entre as regiões de confiança geradas pelo método do Perfil de acordo com a implementação deste trabalho (linha tracejada) e com a disponível na literatura (pontos). ....	72
Figura 18: Comparação entre as regiões de confiança do Modelo 2.A determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada). .	74
Figura 19: Comparação entre as regiões de confiança do Modelo 2.B determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.5) (linha cheia).....	75
Figura 20: Comparação entre as regiões de confiança do Modelo 2.B determinadas pelo método do Perfil (linha pontilhada) e pelo método do Contorno (linha cheia). ....	76
Figura 21: Comparação entre as regiões de confiança do Modelo 2.B determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada). .	77
Figura 22: Comparação entre as regiões de confiança do Modelo 3 determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.7) (linha cheia). ....	78
Figura 23: Comparação entre as regiões de confiança do Modelo 3 determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada). ....	79
Figura 24: Comparação entre os valores de saída experimentais e os preditos para o Modelo 4. ....	81

Figura 25: Comparação entre as regiões de confiança do Modelo 4 determinadas pelo método do Contorno (pontos) e pela Equação (2.34) (linha cheia). .....	82
Figura 26: Comparação entre as regiões de confiança do Modelo 4 determinadas pelo método do Perfil (linha pontilhada) e pela Equação (2.34) (linha cheia). .....	84
Figura 27: Comparação entre as regiões de confiança do Modelo 4 determinadas pelo método Bootstrap (pontos) e pela solução analítica (linha tracejada). .....	85
Figura 28: Comparação entre os valores de saída experimentais e os preditos para o Modelo 5.A. ....	86
Figura 29: Comparação entre as regiões de confiança do Modelo 5.A determinadas pelo método do Contorno (pontos) e pela Equação (2.34) (linha cheia). .....	88
Figura 30: Comparação entre as regiões de confiança do Modelo 5.A determinadas pelo método do Perfil (linha pontilhada) e pelo método do Contorno (linha cheia). .....	89
Figura 31: Comparação entre as regiões de confiança do Modelo 5.A determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada). .	90
Figura 32: Comparação entre os valores de saída experimentais e os preditos para o Modelo 5.B. ....	91
Figura 33: Comparação entre as regiões de confiança do Modelo 5.B determinadas pelo método do Contorno (pontos) e pela Equação (2.34) (linha cheia). .....	93
Figura 34: Comparação entre as regiões de confiança do Modelo 5.B determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada). .	94
Figura 35: Comparação entre o <i>spline</i> linear e <i>spline</i> cúbico para o conjunto de dados da Tabela 16. ....	117
Figura 36: Regiões de confiança do Modelo 2.B com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia). .....	119

## LISTA DE SÍMBOLOS

---

$s$	Desvio padrão amostral.
$\sigma$	Desvio padrão do experimento.
$\lambda$	Direção do algoritmo Gauss-Newton.
$S(\theta)$	Função objetivo.
$V_{ey}$	Matriz de covariância dos desvios experimentais da variável dependente.
$V_{ex}$	Matriz de covariância dos desvios experimentais da variável independente.
$V_{\theta}$	Matriz de covariância dos parâmetros.
$S_J$	Matriz de covariâncias dos pseudoparâmetros.
$J$	Matriz de sensibilidade do modelo.
$H_{\theta}$	Matriz Hessiana da função objetivo em relação aos parâmetros.
$p$	Melhor ponto encontrado pela partícula.
$p_{global}$	Melhor ponto encontrado pelo enxame.
$RDN$	Número aleatório.
$RDL$	Número aleatório limitado.
$B$	Número de amostras geradas aleatoriamente.
$NE$	Número de experimentos realizados.
$N_{Iter}$	Número de iterações.
$NP$	Número de parâmetros do modelo.
$NY$	Número de variáveis dependentes.
$N$	Número de variáveis.
$r_1, r_2$	Números aleatórios do Enxame de Partículas.
$C_1$	Parâmetro cognitivo.
$C_2$	Parâmetro social.
$\Delta x$	Perturbação na direção do parâmetro.
$W$	Peso de inércia.
$\alpha$	Significância estatística.
$h$	Tamanho do intervalo.
$F$	Valor da distribuição $F$ de Fisher.

$\chi^2$	Valor da distribuição <i>qui-quadrado</i> .
$t$	Valor da distribuição <i>t-Student</i> .
$\tilde{S}(\theta_p)$	Valor da função objetivo para o parâmetro $\theta_p$ quando a sua derivada é anulada na etapa de minimização.
$\tilde{\theta}$	Valor do parâmetro gerado pelo método do perfil.
$\tau, \delta$	Valor do parâmetro na escala da distribuição <i>t-Student</i> .
$\hat{\theta}_{(g)}$	Valor do parâmetro quando o experimento $g$ é desconsiderado.
$\tilde{\theta}_g$	Valor do pseudoparâmetro quando o experimento $g$ é desconsiderado.
$\bar{\theta}_J$	Valor médio do pseudoparâmetro quando o experimento $g$ é desconsiderado.
$\bar{\theta}_B$	Valor médio dos pseudoparâmetros.
$z^e$	Valores experimentais.
$z^m$	Valores gerados pelo modelo.
$W_{max},$ $W_{min},$	Valores máximo e mínimo do peso de inércia.
$z^*$	Valores reais.
$V_e$	Variância dos dados experimentais.
$y^m$	Variável dependente calculada pelo modelo.
$y,$	Variável dependente do modelo.
$y^e$	Variável dependente obtida experimentalmente.
$x^m$	Variável independente calculada pelo modelo.
$x$	Variável independente do modelo.
$x^e$	Variável independente obtida experimentalmente.
$v$	Velocidade da partícula do Enxame de Partículas.
$\theta$	Vetor com os parâmetros do modelo.
$\hat{\theta}$	Vetor com os parâmetros ótimos.
$\nabla S_\theta$	Vetor gradiente da função objetivo em relação aos parâmetros.
$e$	Vetor unitário.

## SUMÁRIO

---

Capítulo 1: Introdução .....	1
Capítulo 2: Revisão Bibliográfica .....	5
2.1. Modelos .....	5
2.2. Estimacão de parâmetros .....	7
2.3. Função objetivo.....	8
2.4. Minimização.....	11
2.5. Regiões de confiança .....	19
Capítulo 3: Metodologia.....	33
3.1. Modelos e dados experimentais .....	33
3.2. Estimacão de parâmetros .....	39
3.3. Regiões de confiança .....	43
Capítulo 4: Resultados .....	49
4.1. Análise da região de confiança elíptica dos parâmetros .....	49
4.2. Métodos para determinacão de regiões de confiança não elípticas ....	63
Capítulo 5: Conclusão .....	96
REFERÊNCIAS.....	99
APÊNDICES .....	104
A.1 Aproximacão Numérica da Matriz Hessiana .....	104
A.2 Equacão da Região de Confiança Elíptica.....	107
A.3 Splines .....	109
A.4 Análise da aproximacão numérica da matriz Hessiana.....	118
A.5 Pseudocódigos .....	123
A.6 Artigo .....	135

## Capítulo 1: Introdução

---

As atividades que envolvem análise de dados tendem a crescer, tanto no setor acadêmico como na área industrial. A interpretação destes conjuntos de dados é essencial para o estabelecimento de novas propostas que descrevam ou melhorem o esclarecimento de fenômenos naturais ou que aumentem a eficácia de processos industriais. Para tal, é necessário propor modelos capazes de representar da melhor maneira possível o conjunto de dados estabelecidos. Um modelo adequado é aquele que, a partir de um novo valor de entrada, prediz valores de saída que se relacionam com os valores experimentais utilizados para estabelecer este modelo. Em áreas que estabelecem a proposta de um modelo para representar o conjunto de dados, os valores conhecidos são os de entrada e saída e se deseja determinar o valor das constantes. Estas, chamadas de parâmetros, representem da melhor maneira a relação entre as variáveis de entrada e saída e sua obtenção estabelece um problema de estimação de parâmetros.

Em problemas de estimação de parâmetros, procura-se o valor para determinado conjunto de parâmetros de um modelo de forma que este represente de forma adequada a relação entre os dados experimentais. Para que este valor seja encontrado, são realizadas sucessivas iterações que modificam o valor dos parâmetros até que os valores preditos se ajustem da melhor forma aos dados experimentais. O ajuste entre estes valores é possível pelo uso de uma classe de funções denominadas função objetivo, as quais medem a “distância” entre as predições do modelo e os valores obtidos experimentalmente. Uma função objetivo comumente utilizada, bastante robusta e simples é a função dos mínimos quadrados que compara a soma do quadrado das diferenças entre os valores experimentais e os valores preditos pelo modelo. Porém, seu uso pode não ser adequado em situações nas quais as incertezas associadas aos experimentos são diferentes ou quando as variáveis do modelo possuem diferentes dimensões ou escalas. Uma alternativa é a função dos mínimos quadrados ponderados, que pondera a soma do quadrado das diferenças entre os valores experimentais e os valores preditos pelo valor da variância de cada experimento.

O procedimento que determina o valor dos parâmetros que resultam no menor valor para a função objetivo é denominado minimização. Esta etapa consiste em,

através de uma metodologia específica, reduzir o valor da função objetivo até o menor valor possível ajustando os parâmetros do modelo. Existem diferentes abordagens para este cálculo, sendo que a escolha dependerá de questões como tipo de modelo, precisão dos resultados e preferência do usuário. Há trabalhos realizados por MARSILI-LIBELLI e CHECCHI, (2005); e MARSILI-LIBELLI, GUERRIZIO e CHECCHI, (2003) preferem o uso de técnicas de busca direta, como o método geométrico dos poliedros flexíveis. Já outros, como GUELY; SIARRY, (1993); e RUDER, (2016), utilizam técnicas mais precisas e específicas para problemas de otimização como os métodos do tipo gradiente.

Sendo a classe de métodos de minimização mais clássica, os métodos de gradiente do tipo Newton, utilizam uma estimativa inicial dos parâmetros e os valores do gradiente e da matriz Hessiana da função objetivo para determinar novos valores de forma iterativa até que seja alcançado o menor valor possível. Uma alternativa bastante utilizada para evitar dificuldades relacionadas ao cálculo da matriz Hessiana é a aproximação de Gauss-Newton que propõe uma simplificação para a obtenção da matriz Hessiana.

Entre as dificuldades envolvendo o uso dos métodos de Newton, podem-se destacar a existência de mínimos locais, que dificultam a convergência do algoritmo para o ponto ótimo global, a dependência de boas estimativas iniciais para que o método convirja e a necessidade do cálculo das derivadas. Desta forma, uma alternativa interessante é o uso de métodos estocásticos de minimização da função objetivo, como os métodos do Enxame de Partículas, (KENNEDY; EBERHART, 1995), e do Algoritmo Genético (HOLLAND, 1975), que realizam a busca do valor ótimo de forma global em larga escala na região de interesse e que apresentam características parcialmente aleatórios. Outra grande vantagem destes métodos é que os mesmos são aplicáveis a diversos modelos não-lineares, sendo robustos, pois não é necessário o cálculo das derivadas, as quais podem ser de difícil obtenção ou não existirem.

Entretanto, somente a determinação dos valores ótimos dos parâmetros do modelo não é suficiente, já que dados experimentais estão sujeitos a erros resultantes das imprecisões das técnicas de medidas e dos fatores não controlados. Assim, a determinação das incertezas paramétricas é de fundamental importância para permitir a avaliação da qualidade das estimativas obtidas para cada parâmetro (SCHWAAB;

PINTO, 2007a). A avaliação das incertezas é geralmente feita através da análise do intervalo de confiança de cada parâmetro, mas, devido à correlação paramétrica, principalmente em modelos não lineares, a análise deve ser realizada através das regiões de confiança dos parâmetros.

A determinação das regiões de confiança dos modelos lineares é simples e possui solução analítica, sendo representada por elipses ou elipsoides. Sua equação é obtida através da expansão em série de Taylor da função objetivo e de sua relação com as distribuições  $\chi^2$  (*qui-quadrado*) e  $F$  de Fisher (DONALDSON; SCHNABEL, 1987). Por outro lado, as regiões de confiança dos modelos não-lineares, tipicamente, não possuem uma solução exata, sendo a solução elíptica uma aproximação que pode diferir completamente do seu real comportamento. A sua obtenção geralmente envolve o uso de métodos computacionais intensivos, isto é, com um alto custo computacional.

Nesta dissertação foram avaliadas cinco metodologias para determinação de regiões de confiança. Inicialmente, foi avaliada a diferença entre as regiões elípticas de confiança calculadas com o uso da aproximação de Gauss-Newton e quando o valor completo da matriz Hessiana é empregado. O objetivo principal é comprovar se o uso da aproximação de Gauss-Newton é justificável ou se é necessária a obtenção de sua matriz Hessiana completa.

Para tal foram obtidas as regiões de confiança de modelos lineares e não-lineares através de três diferentes métodos: o método do Contorno, o método do Perfil e o método *Bootstrap*. O método do Contorno procura, a partir de um valor ótimo, um ponto sobre o limite da região de confiança e procura percorrê-la até retornar a este ponto de origem. O método do Perfil se trata da reparametrização do valor dos parâmetros do modelo na escala da distribuição *t-Student* para que eles indiquem a linearidade de modelo, (BATES; WATTS, 1988); contudo, estes valores podem ser utilizados para delimitar os intervalos de confiança de cada parâmetro e a região de confiança através de cálculos específicos de interpolação. O método *Bootstrap*, (EFRON, 1979), consiste em, a partir dos valores experimentais e da medida das suas incertezas, gerar novos valores por reamostragem para os utilizar na confecção da região de confiança.

Este trabalho está dividido em cinco capítulos. O Capítulo 1 se trata deste, no qual são resumidos alguns conceitos abordados neste trabalho, bem como as metodologias usadas para determinação das regiões de confiança.

No Capítulo 2, são apresentados os conceitos básicos e o desenvolvimento matemático utilizados neste trabalho através de uma revisão na literatura sobre a estimação de parâmetros; função de máxima verossimilhança, função dos mínimos quadrados ponderados e mínimos quadrados; minimização da função objetivo pelos métodos de Newton e pelos métodos estocásticos; região de confiança elíptica e outros métodos utilizados para determinar as regiões de confiança.

No Capítulo 3, são apresentados os modelos matemáticos e seus respectivos conjuntos de dados utilizados para o estudo das regiões de confiança. Os métodos de minimização utilizados nas estimações de parâmetros são indicados e têm suas configurações detalhadas. Também são explicados mais especificamente o funcionamento das técnicas e as configurações utilizadas para gerar as regiões de confiança elípticas e pelos métodos do Contorno, Perfil e *Bootstrap*.

No Capítulo 4, são demonstrados e discutidos os resultados das regiões de confiança elípticas calculadas com a matriz Hessiana completa e com a aproximação de Gauss-Newton. Também, os métodos do Contorno, Perfil e *Bootstrap* são analisados e comparados quanto a sua capacidade em delimitar a região de confiança de sete diferentes modelos. As vantagens e desvantagens de cada um são apontadas de acordo com as regiões geradas.

No Capítulo 5, há a discussão final dos resultados obtidos nos Capítulo 4 quanto a capacidade dos métodos em delimitar as regiões de confiança dos parâmetros para modelos com diferentes características. São discutidas algumas melhorias que poderiam ser implementadas aos métodos e, por fim, propõem-se algumas sugestões para eventuais trabalhos futuros.

No final desta dissertação, foram estruturados apêndices com explicações e equacionamentos que complementam algumas abordagens presentes no texto principal. Também há a apresentação do artigo desenvolvido, como pré-requisito para apresentação deste trabalho, no qual alguns conceitos detalhados no Capítulo 2 foram utilizados.

## Capítulo 2: Revisão Bibliográfica

---

Neste capítulo serão abordados alguns dos conceitos necessários para o pleno entendimento das metodologias consideradas por esta dissertação.

### 2.1. Modelos

Um modelo pode ser definido como uma “tentativa de explicar a realidade” (SCHWAAB; PINTO, 2007a). Também pode se referir a observações de eventos aos quais busca-se uma forma de explicá-los através de metodologias científicas, existindo inúmeras maneiras de explicar determinado evento observado, seja através de comparações ou associações com um acontecimento ou pela aplicação de métodos estabelecidos através de conceitos físico-matemáticos. Estão presentes nas mais diversas áreas do conhecimento, tanto humanas, como ciências sociais, administração, sociologia, como nas exatas (matemática, física, química, biologia), e são imprescindíveis para descrever os fenômenos observados.

Embora as áreas possam diferir quanto ao seu entendimento de modelo: para a física, um modelo pode ser entendido como a versão idealizada e simplificada de um sistema físico; para a química, um modelo tende a ser a representação gráfica de compostos e do comportamento entre as substâncias; e para a biologia, pode explicar um ecossistema através das interações entre os seres vivos, é consenso entre todas que um modelo é estabelecido após a observação do comportamento de um sistema.

Um modelo matemático se trata da representação física ou equacional de determinado conceito ou comportamento (PINTO; LAGE, 2001). Representações expositivas de gráficos e figuras matemáticas, bem como os jogos lúdicos utilizados no ensino da matemática básica são exemplos de modelos através de representações físicas.

As representações de fenômenos e experimentos através de equações são os exemplos mais importantes de modelos matemáticos. Usados intensamente nas áreas da computação e engenharia, consistem nas relações entre conjuntos de variáveis para descreverem previsões quantitativas sobre o comportamento de um sistema (SCHWAAB; PINTO, 2007a).

Os modelos teóricos são aqueles que têm sua base no emprego de teorias e pressupostos teóricos para descrever um sistema existente como, por exemplo, as

leis de conservação. A partir de um caminho lógico, tenta-se aplicar os postulados estabelecidos às condições impostas pelo modelo para que seja possível explicar determinado fenômeno (PINTO; LAGE, 2001).

Modelos empíricos são elaborados através de observações de um conjunto representativo de dados que descrevem determinado experimento, aliado ao emprego de pressupostos teóricos para tentar explicá-los. O emprego deste tipo de modelo pode gerar informações errôneas ou que desviam das demais do conjunto a que pertencem.

Também existem os modelos determinísticos e os estocásticos. Enquanto que em um modelo determinístico, cada dado experimental resultará em uma resposta bem definida, os estocásticos geram um conjunto de possíveis resultados, distribuídos probabilisticamente.

Os modelos podem apresentar valores desconhecidos denominados parâmetros, que estão relacionados com o modelo proposto e o conjunto de valores que descreve o sistema. A determinação destes valores é conhecida como estimação de parâmetros.

De acordo com DONALDSON e SCHNABEL, (1987), modelos lineares podem ser definidos conforme a Equação (2.1), onde é possível observar que o modelo é constituído pela soma de diferentes termos de variáveis independentes de entrada e parâmetros, ambos elevados na primeira potência. Adaptando este conceito, pode-se definir o modelo linear nos parâmetros, na qual apenas os parâmetros são elevados à primeira potência, enquanto o termo das variáveis independentes pode estar condicionado por uma função que não envolva os parâmetros.

$$f(x, \theta) = \sum_{j=1}^{NP} g_j(x) \theta_j \quad (2.1)$$

Onde  $f(x, \theta)$  representa uma equação em função das variáveis independentes e dos parâmetros do modelo,  $x$  representa o vetor das variáveis independentes (ou funções de variáveis independentes) do modelo,  $g_j(x)$  representa uma equação genérica em função das variáveis independentes,  $\theta_j$  corresponde ao valor  $j$ -ésimo parâmetro e  $NP$  o número de parâmetros do modelo.

Já os modelos não-lineares diferem deste conceito devido ao fato de que, pelo menos uma de suas derivadas em respeito aos parâmetros tem dependência de um dos parâmetros. Problemas envolvendo o crescimento de microrganismos, cinética de reações químicas e economia são exemplos bem característicos deste tipo de modelo. (BATES; WATTS, 1988).

## 2.2. *Estimação de parâmetros*

Em diversos problemas de engenharia, deparam-se com casos nos quais são necessários o uso de equações e modelos matemáticos para descrever um problema. Muitas dessas situações podem não possuir um modelo bem definido capaz de descrever os eventos. Assim, faz-se necessário propor um modelo que possa reproduzir os resultados observados em eventos. Dessa forma, é estruturado um modelo que considera os dados reais e os compara com os valores correspondentes aos seus obtidos através do uso do modelo.

Qualquer tipo de informação medida apresenta desvios quando comparada com a fornecida por um modelo, seja por questões de precisão de equipamento, manuseio ou condições climáticas (SCHWAAB; PINTO, 2007a; VAN DEN BOS, 2007). Havendo sempre um erro pertinente quando comparamos um valor experimental com o predito pelo modelo, é importante que exista uma maneira de aproximar estas duas informações, ou seja, reduzir o erro que está associado ao problema. Assim, para um modelo são propostas constantes que possam ser ajustadas em função dos desvios entre os valores experimentais e dos preditos, denominadas parâmetros (BENYAHIA *et al.*, 2013).

As técnicas de estimação de parâmetros consistem em metodologias para determinar estes valores desconhecidos, através da aproximação dos dados reais e os valores preditos pelo modelo. Na determinação do melhor conjunto de parâmetros que descreve determinado sistema, a etapa inicial é a propor o modelo que será utilizado, ou seja, garantir que o mesmo seja capaz de representar matematicamente os dados experimentais do problema (BENYAHIA *et al.*, 2013). Para tal, buscam-se sempre os melhores valores de parâmetros, que reduzem a diferença entre as variáveis medidas experimentalmente e as determinadas pelo modelo.

A aproximação entre valores experimentais e os obtidos pelo modelo é garantida através do uso de uma classe específica de funções denominada função

objetivo (ALMEIDA; CASTEL-BRANCO; FALCÃO, 2002), a qual mensura o desvio existente entre os dados calculados e medidos por meio de sucessivas variações nos valores dos parâmetros, realizadas pela minimização, e comparações através dos dados de saída calculados a cada uma delas.

Ao realizar-se um experimento, temos ciência das variáveis de entrada, as quais, idealmente, são controladas e conhecidas pelo usuário, e se observam as variáveis de saída. Após a conclusão do mesmo e da escolha de um modelo, pode-se inferir um conjunto de valores iniciais para os parâmetros e os utilizar para a obtenção de valores de saída do modelo. A obtenção destes valores depende de estudos de casos semelhantes disponíveis na literatura científica, ou pela experiência do pesquisador. Utilizando a função objetivo, comparam-se os valores calculados aos valores reais, obtendo-se o desvio entre o experimento e o modelo, e através de métodos iterativos, varia-se o valor dos parâmetros até que se obtenha o menor valor possível para a função objetivo, os quais corresponderam ao valor ótimo dos parâmetros.

### 2.3. Função objetivo

A função objetivo quantifica a diferença entre os valores experimentais e os preditos pelo modelo. Os métodos de estimação de parâmetros podem ser resumidos como um extenso número de comparações entre estes valores, sendo que para cada uma, o valor da função objetivo é reduzido pela etapa de minimização.

Por ser amplamente utilizada em situações nas quais são necessárias a confiança elevada nos resultados, é importante que esta função possa estar baseada em critérios estatísticos rigorosos para representar o problema com maior confiabilidade estatística possível.

A função de máxima verossimilhança é a etapa inicial para a determinação de estimadores, consistindo em uma função de densidade de probabilidades. A relação entre as variáveis de um problema de estimação é quantificada por esta função, ou seja, os valores de entrada e saída de um experimento estão associados à esta função, e estão sujeitos a erros, Equação (2.2).

$$P(z^e, z^*, V_e) \tag{2.2}$$

Onde  $z^e$  representa os valores experimentais,  $z^*$  representa os valores reais desconhecidos e  $V_e$  representa a medida da incerteza experimental.

Considerando que os erros experimentais estejam normalmente distribuídos e que as variáveis independentes e dependentes sejam descritas pela função normal, o problema de estimação consistirá em maximizar a Equação (2.3).

$$L(z^m, \theta) = \frac{(2\pi)^{-N/2}}{\sqrt{\det(V_e)}} \exp\left(-\frac{1}{2}(z^e - z^m)^T V_e^{-1}(z^e - z^m)\right) \quad (2.3)$$

Onde  $z^m$  representa os valores gerados pelo modelo,  $N$  é o número de variáveis e  $V_e$  representa a matriz de covariância experimental.

Como descrito por SCHWAAB e PINTO, (2007), o valor máximo da função representada pela Equação (2.3) também corresponde ao máximo de seu logaritmo.

$$\ln(L(z^m, \theta)) = \ln\left(\frac{(2\pi)^{-N/2}}{\sqrt{\det(V_e)}}\right) - \frac{1}{2}(z^e - z^m)^T V_e^{-1}(z^e - z^m) \quad (2.4)$$

O primeiro termo da Equação (2.4) é constante, então maximizar a função da Equação (2.3), corresponde a maximizar o segundo termo desta equação que, por sua vez, corresponde a minimizar a Equação (2.5).

$$S(z^m, \theta) = (z^e - z^m)^T V_e^{-1}(z^e - z^m) \quad (2.5)$$

Onde  $S$  representa a função objetivo.

Realizando os experimentos de forma independente, não haverá correlação entre os diferentes experimentos, então, pode-se escrever uma equação para cada um realizado.

$$S(z^m, \theta) = \sum_{i=1}^{NE} (z_i^e - z_i^m)^T V_{ei}^{-1}(z_i^e - z_i^m) \quad (2.6)$$

Onde  $NE$  representa o número de experimentos.

Também é possível separar as variáveis do problema em variáveis de entrada e saída, considerando que estas não estão experimentalmente correlacionadas entre si.

$$S(\theta) = \sum_{i=1}^{NE} \left( y_i^e - y_i^m(x_i^m, \theta) \right)^T V_{eyi}^{-1} \left( y_i^e - y_i^m(x_i^m, \theta) \right) + \sum_{i=1}^{NE} \left( x_i^e - x_i^m \right)^T V_{exi}^{-1} \left( x_i^e - x_i^m \right) \quad (2.7)$$

Onde  $x^e$  e  $x^m$  são os vetores com os valores das variáveis independentes estabelecidas para o experimento e calculadas pelo modelo,  $y^e$  e  $y^m$  são vetores com os valores das variáveis dependentes medidas experimentalmente e calculadas pelo modelo,  $V_{ex}$  e  $V_{ey}$  as matrizes de covariância dos desvios experimentais relacionados às variáveis independentes e às variáveis dependentes, respectivamente.

A estimação dos parâmetros do modelo e dos valores de entrada resultaria em um problema com número elevado de variáveis desconhecidas, sendo assim, geralmente é proposto o descarte do termo envolvendo o desvio nas variáveis independentes (SCHWAAB; PINTO, 2007a). Essa consideração é realizada pois é possível medi-las com precisão ao executar determinado experimento, principalmente quando realizados em escala laboratorial, devido à melhor facilidade em controlar as condições do experimento.

$$S(\theta) = \sum_{i=1}^{NE} \left( y_i^e - y_i^m(x_i^m, \theta) \right)^T V_{eyi}^{-1} \left( y_i^e - y_i^m(x_i^m, \theta) \right) \quad (2.8)$$

Caso as medidas nas variáveis dependentes sejam realizadas de forma que estas não estejam correlacionadas, pode-se escrever a função objetivo de mínimos quadrados ponderados.

$$S(\theta) = \sum_{i=1}^{NE} \sum_{j=1}^{NY} \frac{\left( y_i^e - y_i^m(x_i^m, \theta) \right)^2}{\sigma_{ij}^2} \quad (2.9)$$

Onde  $NY$  é o número de variáveis dependentes e  $\sigma_{ij}^2$  é variância associada à medida da variável  $j$  no experimento  $i$ .

O uso da função dos mínimos quadrados ponderados consiste na atribuição de pesos aos valores de saída conforme a precisão destes, visto que os valores do desvio padrão são influenciados pelos valores das variáveis independentes. A escolha destes pesos pode ser estipulada pela análise do erro relativo ou através de correlações empíricas (ALMEIDA; CASTEL-BRANCO; FALCÃO, 2002).

Por fim, tendo-se o erro das variáveis dependentes constante, a função de mínimos quadrados ponderados pode ser simplificada à função de mínimos quadrados.

$$S(\theta) = \sum_{i=1}^{NE} \left( y_i^e - y_i^m(x_i^m, \theta) \right)^2 \quad (2.10)$$

BARD, (1974), destaca algumas situações nas quais o uso da função de mínimos quadrados pode ser insuficiente: problemas com múltiplas variáveis de entradas ocasionam em problemas ao utilizar esta função objetivo devido às diferenças dimensionais ou de ordem de magnitude; algumas estimações, dentro de um mesmo experimento, são mais confiáveis que outras, e é necessário garantir que a estimação de parâmetros não será influenciada preferencialmente por um deles.

As avaliações da função objetivo são realizadas no processo iterativo denominado minimização, o qual busca, através da redução do valor da função, determinar o seu valor ótimo, ou seja, o menor valor possível para a função objetivo.

#### 2.4. Minimização

Após a escolha da melhor função objetivo para o problema em estudo, deve-se encontrar o menor valor possível para esta função, relacionando os dados experimentais com os preditos pelo modelo e, assim, realizar a minimização da função objetivo (RUDER, 2016). Para tal, podem ser utilizados métodos baseados no cálculo de derivadas, métodos de busca direta ou métodos estocásticos; ainda é possível os combinar para aproveitar melhor as qualidades de cada um (DA ROS *et al.*, 2013; NHAMAGE *et al.*, 2017; SCHWAAB; PINTO, 2007a). O mecanismo básico para que seja possível otimizar a função objetivo envolve iterações que modificam os valores dos parâmetros sequencialmente, garantindo que os valores que satisfaçam a condição do mínimo sejam encontrados.

Pode-se propor, para a resolução de um problema, diversos métodos de otimização, entre eles, destacam-se: métodos tipo Newton ou gradiente, que são caracterizados pelo emprego de técnicas do cálculo diferencial para encontrar os pontos onde as derivadas são nulas; métodos numéricos ou estocásticos, que utilizam informações prévias dos problemas para a sua otimização através de sucessivas iterações (DA ROS *et al.*, 2013); métodos gráficos, que utilizam o gráfico da função para a determinação do ponto ótimo (FENG; BAI; ZHENG, 2007); métodos de busca direta, que propõem a substituição do vetor gradiente da função objetivo por aproximações. A escolha do método a ser utilizado é baseada na interpretação das características do problema, como tipo de função, linearidade e dimensões.

Entre os métodos baseados no cálculo de derivadas, os métodos do tipo gradiente são os mais simples e preferidos para se implementar, pois convergem rapidamente para a maioria dos problemas bem condicionados (ANDREI, 2006). O funcionamento destes métodos consiste inicialmente na escolha de um ponto inicial para os valores dos parâmetros. Após esta escolha, o algoritmo segue determinada direção de acordo com o cálculo do vetor gradiente, se o valor da função objetivo aumentar, ele avança na direção oposta e tende a seguir a direção do gradiente negativo até convergir ao ponto de valor mínimo (ANDREI, 2006; GUELY; SIARRY, 1993; RUDER, 2016).

$$\theta_{k+1} = \theta_k - \lambda_k \nabla S_{\theta} \quad (2.11)$$

Onde  $\theta_k$  representa o valor do parâmetro na iteração  $k$ ,  $\theta_{k+1}$  representa o valor do parâmetro na iteração  $k + 1$ ,  $\lambda_k$  representa o tamanho do passo do método na iteração  $k$  e  $\nabla S_{\theta}$  corresponde ao vetor gradiente da função objetivo em relação aos parâmetros.

Na literatura existem modificações no método do gradiente que visam a melhora das características do algoritmo, visto que, para conjuntos de dados extensivos, característicos em problemas de *Machine Learning*, o método simplificado pode demorar a convergir, não conseguindo resolver o problema de estimação. Essencialmente, as vertentes do algoritmo diferem na quantidade de informações que são utilizadas para calcular o gradiente da função objetivo. RUDER, (2016), apresenta algumas variantes do método do gradiente, entre elas têm-se: o Método do Gradiente em Grupo e o Método do Gradiente Estocástico.

O Método do Gradiente em Grupo realiza o cálculo do gradiente da função objetivo para todos os valores do conjunto de dados a cada iteração, assegurando a determinação do ponto ótimo, mas podendo convergir a um ótimo local. Outra desvantagem é que este método pode ser computacionalmente lento para amostras de elevado tamanho, devido ao número de cálculos necessários para a convergência ao ótimo.

Já o Método do Gradiente Estocástico consiste em calcular o valor do gradiente da função objetivo para cada um dos valores experimentais a cada iteração, permitindo uma exploração maior do espaço de otimização e garantindo melhores chances de convergir ao ótimo global. A variação estocástica do método do gradiente aumenta o número de avaliações do conjunto de dados iniciais, mas evita cálculos repetidos, diminuindo o tempo total de convergência. Contudo, esta contribuição dificulta a determinação do valor exato do mínimo.

$$\theta_{k+1} = \theta_k - \lambda_k \nabla S_{\theta}(\theta, x_i, y_i) \quad (2.12)$$

TARANTOLA, (2005), aponta que os métodos do tipo gradiente tentam balancear dois pressupostos contraditórios. O primeiro é de que, a cada iteração, o método utiliza informações de pontos próximos, ou seja, locais, para otimizar a função; sendo que esta informação não representa todo o domínio da função objetivo. O segundo implica que a cada iteração realizada, o método dá o maior passo possível, distanciando-se da informação local utilizada para dar o passo.

O Método de Newton propõe uma adaptação ao método do gradiente utilizando, além do o vetor gradiente da função objetivo, a sua matriz Hessiana para que se possa calcular os pontos onde o gradiente da função é nulo (HIMMELBLAU, 1972). Assim, o problema de minimização passa a depender das derivadas de segunda ordem da função objetivo, conforme a Equação (2.13).

$$\theta_{k+1} = \theta_k - \lambda_k \cdot H_{\theta}^{-1} \cdot \nabla S_{\theta} \quad (2.13)$$

onde  $H_{\theta}$  é a matriz Hessiana dos valores dos parâmetros estimados.

O princípio do método é obter a direção e o tamanho do passo através do produto do vetor gradiente com a inversa da matriz Hessiana e, assim, iniciar as

iterações até que o ponto ótimo seja encontrado. A busca pelo ponto de mínimo é rápida, pois a convergência do método é quadrática sendo que, a melhor condição de uso do método é para valores iniciais próximos ao ponto de otimização, podendo apresentar baixo desempenho ou não convergir se a estimativa inicial não for boa ou se o problema possuir um número elevado de dimensões (TARANTOLA, 2005).

Outra dificuldade no emprego do método de Newton é a inversão da matriz Hessiana, cujo cálculo pode resultar em um alto gasto computacional e, determinar sua inversa, pode ser complicado ou impossível em problemas com alta correlação entre os parâmetros (SCHWAAB, 2007). Também, faz-se necessário que a matriz Hessiana seja definida positiva, caso contrário não haverá minimização da função objetivo (MONTEIRO, 2013).

Uma alternativa viável para o cálculo da matriz Hessiana em problemas de estimação de parâmetros foi proposta por Gauss. Considerando as suposições da Equação (2.10) para uma única variável dependente, ou seja, sem desvios na variável independente e sem correlação entre os experimentos, pode-se representar a matriz Hessiana conforme a Equação (2.14).

$$H_{\theta} = -2 \sum_{i=1}^{NE} \left( y_i^e - y_i^m(x_i^m, \theta) \right)^T V_{ey}^{-1} \left( \frac{\partial^2 y_i^m}{\partial \theta_i \partial \theta_j} \right) + 2 \sum_{i=1}^{NE} \left( \frac{\partial y_i^m}{\partial \theta_i} \right)^T V_{ey}^{-1} \left( \frac{\partial y_i^m}{\partial \theta_j} \right) \quad (2.14)$$

A aproximação, denominada Gauss-Newton, omite o primeiro termo do lado direito da Equação (2.14), envolvendo a derivada de segunda ordem dos valores do modelo em relação aos parâmetros. O emprego deste artifício se deve à consideração de que os desvios entre os valores dos experimentos e os preditos pelo modelo são pequenos, dessa maneira é possível minimizar a função objetivo apenas com o cálculo das derivadas de primeira ordem do modelo.

$$H_{\theta} \approx 2 \sum_{i=1}^{NE} \left( \frac{\partial y_i^m}{\partial \theta_i} \right)^T V_{ey}^{-1} \left( \frac{\partial y_i^m}{\partial \theta_j} \right) \quad (2.15)$$

Embora o uso dos métodos baseados em derivadas garanta rapidamente a convergência ao ponto ótimo, muitos modelos não lineares não podem ser estimados por métodos analíticos devido à impossibilidade da inversão de suas matrizes

Hessianas. Para estes modelos, alguns cuidados na minimização devem ser tomados como: o valor do estimativa inicial de parâmetros, o formato da função, observando a existência de mínimos locais e as iterações entre os parâmetros, que podem resultar em grandes regiões de confiança (DA ROS *et al.*, 2013). Assim, alternativamente, existem técnicas que substituem o cálculo do vetor gradiente da função objetivo para contornar os problemas relacionados à convergência. Entre estas técnicas, os métodos de busca direta estabelecem alternativas para o cálculo do vetor gradiente (LEWIS *et al.*, 2000).

Introduzidas por HOOKE e JEEVES, (1961), as técnicas de busca em linha consistem em avaliações sequenciais do valor da função objetivo durante a sua minimização até que se obtenha o menor valor possível (LEWIS *et al.*, 2000). A diferença entre este tipo de método e os baseados em derivadas é devido ele não necessitar das informações das derivadas da função objetivo para definir o valor ótimo. A busca do algoritmo se dá comparando o valor atual da função objetivo com o menor valor encontrado até o momento. O avanço do algoritmo é obtido através do incremento da função objetivo ou pelo uso de abordagens geométricas (BAEYENS; HERREROS; PERÁN, 2016). Entre estes eles, destacam-se o Método de Hooke-Jeeves (*pattern seach*) e o Método de Nelder-Mead (*polytope method*) ou poliedros flexíveis.

As principais observações a respeito do uso destes métodos estão relacionadas com o seu funcionamento. Os métodos que não seguem uma abordagem matemática rigorosa são lentos e são melhor condicionados para problemas com conjunto de dados pequenos. Contudo, por dispensarem o cálculo de derivadas, conseguem minimizar vários modelos, falhando poucas vezes em determinar o valor mínimo da função objetivo (BAEYENS; HERREROS; PERÁN, 2016).

Outra alternativa aos métodos baseados em derivadas são os métodos heurísticos, que dispensam o cálculo de derivadas para convergirem, utilizando conjuntos de dados aleatórios e metodologias específicas para o problema de otimização como o Enxame de Partículas e o Algoritmo Genético.

A ideia principal dos métodos heurísticos é, a partir de um conjunto inicial de pontos gerados aleatoriamente, modificá-los de acordo com sua metodologia, geralmente de acordo com comportamentos da natureza, de forma que esse conjunto

de pontos possa se aproximar do ponto ótimo a cada iteração. Esses métodos vêm sendo amplamente utilizados, pois são flexíveis a diversos problemas e restrições, sendo aplicáveis em várias áreas e, embora não assegurem a otimização ao ponto de mínimo exato, fornecem uma ótima aproximação do seu real valor (BENYAHIA *et al.*, 2013; DA ROS *et al.*, 2013).

O método do Enxame de Partículas vem sendo utilizado em diversos problemas devido ao seu alto rendimento e flexibilidade em resolver problemas não lineares e não diferenciáveis (SHI; EBERHART, 1998). O conceito da técnica baseia-se em replicar o comportamento social de animais em situações como busca de alimentos, abrigo ou reprodução, onde cada partícula é direcionada a um sentido baseado em sua experiência e na troca de informações com o enxame (CHEN, 2009; WANG; YANG, 2010).

O funcionamento do modelo é dividido em três partes, a primeira relativa à inércia da partícula; a segunda relativa à atração da partícula ao melhor ponto já encontrado; e a terceira relativa à atração com o melhor ponto de todo o grupo (SCHWAAB *et al.*, 2008). A abordagem de busca do Enxame de Partículas é descrita nas Equações (2.16) e (2.17) abaixo.

$$v_{i,d}^{k+1} = v_{i,d}^k + c_1 r_1 (p_{i,d}^k - x_{i,d}^k) + c_2 r_2 (p_{global,d}^k - x_{i,d}^k) \quad (2.16)$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^{k+1} \quad (2.17)$$

Onde os índices  $k$ ,  $i$  e  $d$  mostrados na equação descrevem, respectivamente, a iteração, a partícula e a direção de busca;  $v$  é a velocidade da partícula e  $x$  a posição dela no espaço de busca;  $c_1$  e  $c_2$  são definidas como constantes positivas denominadas parâmetro cognitivo e social, respectivamente;  $r_1$  e  $r_2$  são dois números aleatórios com distribuição uniforme entre 0 e 1;  $p_i$  é o melhor ponto encontrado pela partícula  $i$  e  $p_{global}$  é o melhor valor encontrado pelo enxame.

Cada partícula do enxame é tratada como um ponto no espaço de  $N$  dimensões, ocupando uma posição  $x$ ,  $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ . A melhor posição encontrada pela partícula é armazenada em  $p$ ,  $p_i = (p_{i1}, p_{i2}, \dots, p_{iN})$  e a melhor posição do enxame é armazenada em  $p_{global}$  (SHI; EBERHART, 1998).

A primeira parte da equação está associada à velocidade da partícula, também chamada de comportamento habitual (ÖZSOY; ÖRKÇÜ, 2016). Se a Equação (2.16) considerasse apenas este termo, a partícula se moveria apenas em uma direção, até encontrar os limites estabelecidos para o problema de otimização, ou seja, a única maneira de encontrar o ponto ótimo seria se ele estivesse na mesma trajetória da partícula.

Caso o primeiro termo da Equação (2.16) fosse removido, a área de busca seria restringida pelo resultado das iterações, podendo convergir para mínimos locais. Assim, essa solução é influenciada pela população inicial do enxame. Ao inserir o primeiro termo nessa equação, a tendência é que as partículas explorem o espaço de busca, tendo maior facilidade em encontrar o mínimo global. Dessa forma, SHI e EBERHART, (1998), propõem que haja um balanceamento entre a capacidade de busca local e a de busca global introduzindo o peso de inércia à equação.

O segundo termo da Equação (2.16) é a aproximação linear da partícula pelo melhor ponto encontrado por ela, enquanto o terceiro termo da mesma equação representa a aproximação linear da partícula com o melhor ponto encontrado pelo enxame. O comportamento destes termos é controlado pelos parâmetros social e cognitivo, os quais definirão qual será o movimento preferencial do enxame. Valores baixos desses parâmetros permitem que as partículas se dispersem pela região, contribuindo para a exploração do espaço, enquanto valores elevados asseguram que as mesmas se movimentem bruscamente à direção resultante, contribuindo para a convergência ao ponto de mínimo (ÖZSOY; ÖRKÇÜ, 2016; WANG; YANG, 2010).

Existem adaptações para o Enxame de Partículas, como a proposta por SHI e EBERHART, (1998), na qual é considerado o termo de peso de inércia,  $W$ , Equação (2.18), responsável por modificar o termo da velocidade prévia da partícula.

$$v_{i,d}^{k+1} = Wv_{i,d}^k + c_1r_1(p_{i,d}^k - x_{i,d}^k) + c_2r_2(p_{global,d}^k - x_{i,d}^k) \quad (2.18)$$

Onde  $W$  corresponde ao peso de inércia.

A inserção do peso de inércia proporciona ao o algoritmo o equilíbrio entre a sua capacidade de exploração global e local. Valores maiores para o parâmetro peso de inércia garantem que o enxame não convirja a um mínimo local e valores mais baixos asseguram que ele explore menos a região, ou se concentre em áreas

próximas ao mínimo (WANG; YANG, 2010). A estratégia mais adequada para escolha do valor do peso de inércia segundo ÖZSOY e ÖRKÇÜ, (2016), é assegurar a exploração mais eficiente do espaço disponível às partículas com o menor número de interações possíveis. Assim, comumente, o peso de inércia é definido como uma função decrescente em respeito ao número de iterações.

$$W = W_{\max} - \frac{W_{\max} - W_{\min}}{NIter} i \quad (2.19)$$

Onde  $W_{\max}$  corresponde ao valor inicial do peso de inércia,  $W_{\min}$  corresponde ao valor final do peso de inércia,  $i$  representa o número de iterações do algoritmo e  $NIter$  representa o número total de iterações do algoritmo.

Outro método heurístico, cuja vantagem é não necessitar do cálculo das derivadas do modelo, é o Algoritmo Genético e suas adaptações, que são baseados no conceito de evolução das espécies, na qual é possível que a população sofra mutações ou que os indivíduos do conjunto sejam cruzados.

O mecanismo do algoritmo começa com a geração de uma população aleatória de pontos na qual se avalia quais são os melhores; para problemas de otimização considera-se o correspondente ao menor valor da função objetivo (WHITLEY, 1994). Após, a ideia é garantir que a população seguinte esteja melhor adaptada, ou seja, apresente melhores valores para a função objetivo.

A continuidade do algoritmo é devida à geração de um novo conjunto de indivíduos baseada na anterior e, para isso, estes indivíduos são submetidos a eventos que simulam adaptações de uma população ao ambiente como: elitismo, cruzamento, mutação e geração espontânea.

O elitismo considera que os melhores pontos de um conjunto são aptos, e por isso os clona para o novo conjunto, (HARIK; LOBO; GOLDBERG, 1999), garantindo que a população não piore na próxima geração. O cruzamento consiste no cruzamento entre dois pontos, denominados pais, para que se possa gerar um novo, denominado filho. Existem várias técnicas para misturar os indivíduos pais, podendo ser por média aritmética, média geométrica, ou, a mais utilizada, *Blend*, na qual se utiliza um número aleatório com distribuição normal (SILVA, 2006).

O processo de mutação copia o indivíduo da geração anterior e o perturba para que o valor de sua função objetivo possa ser modificado, enquanto a geração espontânea gera novos indivíduos aleatórios, melhorando a busca inicial, assegurando que o algoritmo não convirja para um ótimo local.

Usualmente, os parâmetros que podem ser manipulados no algoritmo são as taxas de cada evento, ou seja, qual evento acontecerá com maior incidência. SILVA, (2006), estabelece as condições que devem ser evitadas, como taxas elevadas de mutação, que tornam a busca essencialmente aleatória; taxas muito elevadas de cruzamento podem comprometer a permanência de indivíduos pais bons e gerar filhos ruins, enquanto taxas baixas tornam o algoritmo lento; e taxas muito altas de clonagem resultam na substituição de pontos bons da população.

Usualmente o Enxame de Partículas apresenta como principal vantagem ao Algoritmo Genético a implementação mais fácil e poucos parâmetros do modelo a serem ajustados, bem como sua robustez e rapidez em encontrar o ponto ótimo (WANG; YANG, 2010). Contudo, essa preferência pode variar para o tipo de problema que está sendo estudado; os Algoritmos Genéticos, por exemplo, apresentam melhor desempenho para solucionar os problemas do tipo Caixeiro Viajante, cujo funcionamento se adapta melhor à dinâmica do problema.

## 2.5. *Regiões de confiança*

Até agora foi discutida a obtenção do melhor conjunto de parâmetros que representa um modelo se baseando em suas variáveis de entrada e saída. Contudo, apenas determinar o melhor conjunto de parâmetros através do valor da função objetivo não é o suficiente (CHECCHI; GIUSTI; MARSILI-LIBELLI, 2007). Considerar que os valores experimentais estão sujeitos a erros é fundamental para problemas de estimação de parâmetros e, dessa forma, o uso de ferramentas estatísticas para determinar a faixa de abrangência dos valores determinados para os parâmetros se faz necessário.

Para um único parâmetro podemos inferir uma faixa de valores que simboliza, estatisticamente, que os valores destes limites são idênticos com confiança de  $1 - \alpha$ ; este intervalo é denominado intervalo de confiança.

$$\left( \hat{\theta} - t_{(NE-NP; 1-\frac{\alpha}{2})} s_{\theta}; \hat{\theta} + t_{(NE-NP; 1-\frac{\alpha}{2})} s_{\theta} \right) \quad (2.20)$$

Onde  $t_{(NE-NP; 1-\alpha/2)}$  é o valor da distribuição *t-Student* com  $NE - NP$  graus de liberdade e significância igual a  $\alpha$  (geralmente igual a 0,05, de forma que o intervalo tenha 95% de confiança),  $s_{\theta}$  é o desvio padrão das estimativas dos parâmetros.

Quando o modelo apresenta mais parâmetros, existem correlações entre eles, desta forma, o conceito intervalo de confiança deve ser expandido. Assim, a região de confiança é um espaço que delimita todos os conjuntos de valores dos parâmetros que, quando substituídos no modelo, ainda garantem a qualidade do ajuste aos dados experimentais (BENYAHIA *et al.*, 2013), já que o conjunto de parâmetros contidos na região de confiança é estatisticamente igual ao conjunto ótimo; permitindo-se supor que o modelo utilizado é uma boa representação para a relação entre os valores de saída experimentais e os de entrada.

Segundo (SCHWAAB; PINTO, 2007b), a região de confiança para problemas de estimação de parâmetros pode ser definida do como um hiper-elipsoide nas dimensões delimitadas pelos parâmetros. A disposição no espaço deste elipsoide é condicionada pela correlação entre os parâmetros, sendo que seus eixos são paralelos aos dos parâmetros quando estes não são correlacionados e, à medida que existe uma correlação, a elipsoide demonstra determinada inclinação.

O método mais comum e usual para determinar a região de confiança dos parâmetros é pela aproximação quadrática da função objetivo em torno do ponto de mínimo, dessa maneira, é possível reescrever a função objetivo como uma série de Taylor truncada no segundo termo.

$$S(\theta) = S(\hat{\theta}) + (\theta - \hat{\theta})^T \nabla S_{\theta} + \frac{1}{2} (\theta - \hat{\theta})^T H_{\theta} (\theta - \hat{\theta}) \quad (2.21)$$

Onde  $\theta$  é o vetor dos parâmetros e  $\hat{\theta}$  é o vetor dos parâmetros no ponto ótimo;  $S$  é a função objetivo avaliada para os valores de  $\theta$  e  $\hat{\theta}$ ;  $\nabla S_{\theta}$ , é o vetor gradiente da função objetivo e  $H_{\theta}$  é a matriz Hessiana da função objetivo em relação aos parâmetros.

Como a aproximação é realizada no ponto mínimo, o gradiente da função objetivo é nulo, ou seja, este termo com as derivadas de primeira ordem pode ser desconsiderado da equação.

$$S(\theta) = S(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^T H_{\theta}(\theta - \hat{\theta}) \quad (2.22)$$

A matriz Hessiana pode ser relacionada com a matriz de covariância dos parâmetros de acordo com a Equação (2.23) (SCHWAAB; PINTO, 2007a).

$$V_{\theta} \approx 2H_{\theta}^{-1} \quad (2.23)$$

Quando o cálculo da matriz Hessiana em relação aos parâmetros é substituído pela aproximação proposta por Gauss, Equação (2.15), a matriz de covariância dos parâmetros pode ser calculada da seguinte forma (KANG; BATES, 1990):

$$V_{\theta} = \left[ \sum_{i=1}^{NE} \left( \frac{\partial y_i^m}{\partial \theta_j} \right)^T V_{ey}^{-1} \left( \frac{\partial y_i^m}{\partial \theta_j} \right) \right]^{-1} \quad (2.24)$$

onde  $V_{ey}$  é a matriz de covariância dos erros experimentais e  $V_{\theta}$  é a matriz de covariância dos parâmetros. Por fim a função objetivo é escrita da seguinte maneira.

$$S(\theta) - S(\hat{\theta}) = (\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) \quad (2.25)$$

Assumindo que os desvios entres os valores obtidos experimentalmente e os preditos pelo modelo são normalmente distribuídos e que o modelo é linear, as estimativas dos parâmetros também têm distribuição normal. Dessa forma, como o lado direito da Equação (2.25) apresenta a soma dos desvios em relação ao ponto ótimo elevados ao quadrados e ponderados pelas respectivas (co)variâncias, é correto afirmar que a Equação (2.25) pode ser descrita pela distribuição *qui-*

quadrado,  $\chi^2$ , com  $NP$  graus de liberdade e confiança igual a  $1 - \alpha$ , (CUI; YAN; SONG, 2015), sendo que, pode-se escrever o lado esquerdo dessa equação como:

$$S(\theta) - S(\hat{\theta}) \leq \chi^2(NP, 1 - \alpha) \quad (2.26)$$

e para o lado direito da Equação (2.26) como:

$$(\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) \leq \chi^2(NP, 1 - \alpha) \quad (2.27)$$

As duas equações anteriores consideram que a matriz de covariância dos parâmetros é conhecida com precisão, o que, quase sempre não é verdade, visto que seria necessário um número grande de observações dos experimentos. Para prosseguir com o equacionamento, as Equações (2.26) e (2.27) são divididas pela função objetivo e ponderadas pelos seus respectivos graus de liberdade, dessa forma, a razão entre valores com distribuição *qui-quadrado* segue a distribuição  $F$  de Fisher (YOO *et al.*, 2017).

$$\frac{S(\theta) - S(\hat{\theta}) / NP}{S(\hat{\theta}) / (NE - NP)} \leq F(NP, NE - NP, 1 - \alpha) \quad (2.28)$$

$$\frac{(\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) / NP}{S(\hat{\theta}) / (NE - NP)} \leq F(NP, NE - NP, 1 - \alpha) \quad (2.29)$$

Onde  $F(NP, NE - NP, 1 - \alpha)$  representa a distribuição cumulativa  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$ .

Reescrevendo a função objetivo conforme a Equação (2.8) e expandido a matriz de covariância dos parâmetros em termos da matriz de sensibilidade de do modelo em relação aos parâmetros e da matriz de covariância das variáveis dependentes.

$$\frac{(\theta - \hat{\theta})^T (J^T V_y^{-1} J) (\theta - \hat{\theta}) / NP}{(y^e - y^m)^T V_y^{-1} (y^e - y^m) / NE - NP} \leq F(NP, NE - NP, 1 - \alpha) \quad (2.30)$$

Onde  $J$  corresponde a matriz de sensibilidade do modelo, com suas derivadas de primeira ordem.

Supondo que os valores das covariâncias das variáveis dependentes são desconhecidos, admite-se que as variâncias são constantes e as covariâncias são nulas.

$$\frac{(\theta - \hat{\theta})^T \frac{(J^T J)}{\sigma^2} (\theta - \hat{\theta}) / NP}{(y^e - y^m)^T \frac{I}{\sigma^2} (y^e - y^m) / NE - NP} \leq F(NP, NE - NP, 1 - \alpha) \quad (2.31)$$

Onde  $\sigma^2$  corresponde à variância experimental, que muitas vezes não é conhecida com qualidade. A Equação (2.31) pode ser simplificada e escrita como:

$$\frac{(\theta - \hat{\theta})^T (J^T J) (\theta - \hat{\theta}) / NP}{(y^e - y^m)^T (y^e - y^m) / NE - NP} \leq F(NP, NE - NP, 1 - \alpha) \quad (2.32)$$

Reescrevendo o denominador como a estimativa da variância,  $s^2$ , e aproximando o produto das matrizes de sensibilidade como o produto da estimativa da variância com a inversa da matriz de covariância dos parâmetros, chega-se a:

$$\frac{s^2 (\theta - \hat{\theta})^T V_\theta^{-1} (\theta - \hat{\theta}) / NP}{s^2} \leq F(NP, NE - NP, 1 - \alpha) \quad (2.33)$$

Assim, podemos definir a região de confiança sendo função dos parâmetros do ponto mínimo, da matriz de covariância dos parâmetros, do número de parâmetros e da distribuição  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$ .

$$\left(\theta - \hat{\theta}\right)^T V_{\theta}^{-1} \left(\theta - \hat{\theta}\right) \leq NP F(NP, NE - NP, 1 - \alpha) \quad (2.34)$$

DONALDSON e SCHNABEL, (1987), salientam que, devido às aproximações e considerações realizadas, a Equação (2.34) só é válida para modelos lineares, visto que o desvio entre as variáveis do modelo e as medidas segue a distribuição normal e que a matriz Hessiana definida pela aproximação de Gauss-Newton seja válida. Para modelos não-lineares, a qualidade da aproximação da região de confiança não é exata e, assim, dependerá do desvio da linearidade do modelo. Dessa maneira, os resultados podem levar a conclusões errôneas sobre os parâmetros estimados e sobre a região de confiança (SCHWAAB *et al.*, 2008).

Analogamente, podemos realizar os mesmos procedimentos para o lado esquerdo da Equação (2.25), obtendo a equação da região da Razão de Verossimilhança (DONALDSON; SCHNABEL, 1987).

$$S(\theta) - S(\hat{\theta}) \leq s^2 NP.F(NP, NE - NP, 1 - \alpha) \quad (2.35)$$

Para modelos lineares, as Equações (2.34) e (2.35) são idênticas, formando a mesma região de confiança. Já para modelos não-lineares, o erro da Equação (2.35) está associado ao fato de que a Equação (2.26) não segue a distribuição *qui-quadrado*.

O algoritmo do Enxame de Partículas também pode ser utilizado para delimitar a região de confiança dos parâmetros. A principal ideia do método é, a cada iteração realizada, aproximar as partículas do valor ótimo da função objetivo, sendo que, para tal, é recomendado o uso de vários pontos. Aproveitando as partículas geradas durante a etapa de minimização da função e escolhendo apenas as que possuem o valor da função objetivo menor do que o seu valor nos limites da região é possível as utilizar para delimitar o espaço interno da região de confiança (KODURU; WELCH; DAS, 2007).

A eficiência do emprego do Enxame de Partículas para delimitar as regiões de confiança dos parâmetros dependerá da escolha dos valores do peso de inércia,

responsável condicionar a velocidade do algoritmo (SHI; EBERHART, 1998). Para valores pequenos de peso de inércia, o termo primeiro termo da equação terá valor pequeno, restringindo a área de busca às proximidades do ponto de mínimo, dessa forma as partículas não serão capazes de definir a região de confiança nos locais afastados do ponto ótimo. Caso o termo de inércia tenha valores elevados, a busca do enxame será realizada em áreas afastadas do ponto de mínimo, dificultando a convergência do problema. Assim, a melhor alternativa para que a região de confiança seja bem definida é assegurar que o valor do peso de inércia possa ser alterado durante as iterações das partículas, garantindo que a região seja explorada em sua totalidade e convirja sem dificuldades (KESSENTINI; BARCHIESI, 2015). Para tal, pode ser utilizada a Equação (2.19).

Outros como, LOBRY, ROSSO e FLANDROIS, (1991), propuseram um método para delimitar a região de confiança baseado na interpolação de pontos distribuídos uniformemente e o ponto mínimo. O algoritmo consiste na construção de uma esfera unitária gerada com pontos aleatórios que é transladada às dimensões do problema de estimação e, através de sucessivas interpolações lineares, adequam-se os pontos ao contorno da região de confiança. As regiões que possuem contornos bem definidos são facilmente representadas pelo método, enquanto as que possuem ângulos abruptos ou geometria com várias mudanças de direção podem ser difíceis de representar ou apresentarem erros na delimitação da região (DORNELES, 2018).

Outro algoritmo heurístico para determinar as regiões de confiança dos parâmetros é o proposto por KLEPPER e HENDRIX, (1994), denominado UCPR. O método consiste em comparar o valor da função objetivo de um ponto gerado aleatoriamente como a média dos valores da função objetivo de um conjunto gerado aleatoriamente. Caso o valor da função objetivo seja menor que a média, este ponto substituirá o pior ponto do conjunto, caso contrário, será gerado um novo ponto. O método delimita diferentes configurações de regiões de confiança sem muito esforço, contudo o mesmo pode ser computacionalmente exaustivo devido à significativa quantidade de avaliações da função objetivo realizadas (DORNELES, 2018).

Um dos métodos que é empregado neste trabalho para delimitar a região de confiança de parâmetros é método *profile*, ou perfil, proposto por BATES e WATTS, (1988). Os autores o propõem como uma alternativa ao cálculo da região pela

aproximação linear elíptica através da sumarização da informação contida na região de confiança dos parâmetros em seus perfis. (WATTS, 1994).

BATES e WATTS, (1988), dividem o seu método em três partes: perfis (*t-plots*), traços (*traces*) e os esboços (*pair sketches*). No trabalho que introduz o método, há uma ótima explicação sobre a concepção dos perfis e dos traços bem como, há um pseudoalgoritmo explicando os cálculos, contudo, para o desenvolvimento dos esboços, a explicação se torna vaga. Outros trabalhos envolvendo os autores também explicam detalhadamente as duas primeiras etapas, mas não mostram o desenvolvimento da terceira.

Para a concepção dos perfis, BATES e WATTS, (1988), introduzem o cálculo da função objetivo do perfil,  $\tilde{S}(\theta_p)$ . Esta função objetivo deve ser a mesma utilizada para o problema de otimização, a diferença é observada na etapa de minimização. Para tal é utilizada uma adaptação do método de Gauss-Newton na qual, a derivada de primeira ordem em respeito ao parâmetro do perfil,  $\tilde{\theta}_p$ , será igualada a zero no vetor gradiente da função objetivo. Após o cálculo da nova função objetivo, seu valor é utilizado para gerar as reparametrizações dos parâmetros na escala da distribuição *t-Student* conforme a equação abaixo.

$$\tau(\theta_i) = \text{sign}(\theta_i - \hat{\theta}_i) \cdot \sqrt{\frac{NE - NP}{S(\hat{\theta})}} \cdot \sqrt{\tilde{S}(\theta_i) - S(\hat{\theta})} \quad (2.36)$$

Onde  $\tau(\theta_i)$  representa o parâmetro na escala *t-Student* correspondente ao parâmetro  $\theta_i$ ,  $\text{sign}(\theta_i - \hat{\theta}_i)$  representa o sinal do valor resultante da operação  $\theta_i - \hat{\theta}_i$ ,  $S(\hat{\theta})$  representa o valor da função objetivo calculada no ponto ótimo e  $\tilde{S}(\theta_i)$  representa o valor da função objetivo para o parâmetro  $\theta_i$  quando a sua derivada é anulada na etapa de minimização. O intervalo de validade para os valores dos parâmetros reparametrizados com nível de confiança igual a  $1 - \alpha$  é definido pela expressão a seguir.

$$-t\left(NE - NP, 1 - \frac{\alpha}{2}\right) \leq \tau(\theta_i) \leq t\left(NE - NP, 1 - \frac{\alpha}{2}\right) \quad (2.37)$$

Para os modelos lineares, este tratamento matemático sugerido por BATES e WATTS, (1988), é idêntico a reparametrizar os parâmetros diretamente, conforme a equação a seguir.

$$\delta(\theta_i) = \frac{\theta_i - \hat{\theta}_i}{\sigma_i} \quad (2.38)$$

Onde  $\delta(\theta_i)$  representa o valor do parâmetro  $\theta_i$  parametrizado e  $\sigma_i$  o desvio padrão em relação ao parâmetro  $\theta_i$ .

Assim, quando comparamos os valores resultantes das Equações (2.36) e (2.38) para um modelo linear, obtêm-se duas curvas idênticas descritas por retas entre os intervalos superiores e inferiores da Equação (2.37). Ao realizarmos o mesmo procedimento para uma função não-linear, obtêm-se uma reta, resultante da Equação (2.38) e uma curva, resultante da Equação (2.36). Essa análise, a mais importante do método, serve como medida da linearidade de um modelo através da diferença entre os valores destas duas equações (BATES; WATTS, 1988, 1991).

O estudo da linearidade utilizando os perfis de um modelo é apontado por WATTS, (1994), o qual afirma que o método propõe uma ótima forma de descrever o comportamento. Outro trabalho, SULIEMAN; MCLELLAN; BACON, (2001), utiliza o cálculo dos perfis para demonstrar a não linearidade de modelos cinéticos quando comparados com os valores da Equação (2.38).

Outra aplicação utilizando os perfis é a capacidade de relacioná-los, formando os traços. Estes consistem em relacionar os valores dos parâmetro resultante do cálculo da função objetivo do perfil, por exemplo: para um modelo com dois parâmetros, quando o parâmetro do perfil é  $\theta_i$ , o resultado da minimização é o par  $\tilde{\theta}_i, \theta_j$ ; da mesma forma que, quando o parâmetro do perfil é  $\theta_j$ , o resultado da minimização é o par  $\theta_i, \tilde{\theta}_j$ ; assim, os traço são construídos pelas curvas  $\tilde{\theta}_i(\theta_j)$  versus  $\tilde{\theta}_j(\theta_i)$  (BATES; WATTS, 1991).

Os perfis resultam em retas que se interceptam na coordenada correspondente ao ponto ótimo e o ângulo entre elas é função da matriz de correlações dos parâmetros (BATES; WATTS, 1991). Para modelos com correlação alta entre os parâmetros, as duas retas estarão próximas e com inclinações semelhantes, enquanto que, para modelos com parâmetros pouco correlacionados, a

inclinação entre as retas será próxima à  $90^\circ$ . Já para os modelos não-lineares, os perfis terão curvatura que dependerá do grau de não-linearidade do modelo. A importância dessa etapa do método, segundo BATES e WATTS, (1988), é a análise do comportamento da região de confiança através dos traços, os quais se interceptam de maneira análoga à inclinação da região.

Os traços gerados também são fundamentais para a última etapa do método do perfil, a geração do esboço, o qual é a representação da região de confiança. BATES e WATTS, (1988), abordam que, sabendo a localização dos pontos onde a região de confiança intercepta os traços, têm-se quatro pontos pertencentes ao contorno da curva. Somado a isso, há o fato que a curva da região é tangente a estes pontos, pois estes independem da direção do parâmetro  $\theta_i$ . De posse destas informações, é proposto um algoritmo relacionando os valores dos parâmetros com o de suas versões reparametrizadas através do emprego de uma série de interpolações.

Outros tipos de abordagem empregados na literatura para delimitar as regiões de confiança dos parâmetros são as técnicas de reamostragem. Esses métodos consistem em, a partir dos valores das variáveis dependentes de um modelo, relacioná-los de maneira a gerar um novo conjunto de dados sem a necessidade de realizar novos experimentos. Segundo SILVA FILHO, (2010), os métodos não necessitam que a distribuição dos dados seja normal e também é utilizado como justificativa para que menos experimentos sejam realizados. Ressalva-se que, independentemente do método utilizado, uma amostra de tamanho expressivo, assim como experimentos bem planejados, contribui para que se possa obter resultados com maior confiabilidade.

O *Jackknife* é um dos métodos de reamostragem de dados que é utilizado para a obtenção das regiões de confiança dos parâmetros. Foi introduzido por QUENOUILLE, (1949), para explicar os erros de estimadores e tem como principal ideia estimar os valores dos parâmetros quando o valor de um dos pontos  $[x_g, y_g]$  é desconsiderado do cálculo (SIMONOFF; TSAI, 1986).

Após as etapas de estimação de parâmetros, para um conjunto com  $NE$  valores experimentais, são estipuladas todas as combinações possíveis excluindo um destes do conjunto inicial. Por exemplo, para o conjunto de tamanho  $NE$ , são obtidos  $NE$  subconjuntos de tamanho  $NE - 1$ . Para cada subconjunto, é realizada a estimação de parâmetros de forma que se obtenha  $NE$  conjuntos de parâmetros,  $\hat{\theta}_{i,(g)}$ .

Em seguida, os valores dos parâmetros obtidos quando um experimento é excluído dos valores experimentais são utilizados para a obtenção dos pseudovalores dos parâmetros, pseudoparâmetros, calculados conforme a Equação (2.39).

$$\tilde{\theta}_{i,g} = NE \cdot \hat{\theta}_i - (NE - 1) \hat{\theta}_{i,(g)} \quad (2.39)$$

Onde,  $\hat{\theta}_{i,(g)}$  é a estimativa do valor do parâmetro  $i$  quando o experimento  $g$  é desconsiderado e  $\tilde{\theta}_{i,g}$  é o valor do pseudoparâmetro  $i$  quando o experimento  $g$  é desconsiderado.

Os valores dos pseudoparâmetros obtidos são sumarizados em uma estimativa do valor dos parâmetros pelo cálculo da média aritmética.

$$\bar{\theta}_{i,J} = \frac{1}{NE} \sum_{g=1}^{NE} \tilde{\theta}_{i,g} \quad (2.40)$$

Onde  $\bar{\theta}_{i,J}$  é o valor médio dos pseudoparâmetros em relação ao parâmetro  $i$ .

Após a confecção dos pseudoparâmetros, é necessário adaptar a matriz de covariância, visto que o conjunto de valores de saída do modelo foi alterado; assim é definida a matriz de covariâncias dos pseudovalores.

$$S_J = \frac{1}{NE - 1} \sum_{g=1}^{NE} (\tilde{\theta}_g - \bar{\theta}_J) (\tilde{\theta}_g - \bar{\theta}_J)^T \quad (2.41)$$

Onde  $S_J$  é a matriz de covariância dos pseudovalores.

A região de confiança dos parâmetros é definida pela adaptação da Equação (2.34), sendo que também é considerado que os pseudovalores dos parâmetros seguem a distribuição normal.

$$(\theta - \bar{\theta}_J)^T S_J^{-1} (\theta - \bar{\theta}_J) \leq \frac{NP}{NE - NP} \frac{NE - 1}{NE} F(NP, NE - NP, \alpha) \quad (2.42)$$

Algumas adaptações ao *Jackknife* são apresentadas por SIMONOFF e TSAI, (1986), as quais apresentam casos como o *Jackknife* Linear, no qual na Equação (2.42) são consideradas o cálculo das derivadas do modelo, e o *Jackknife* Linear Ponderado, que considera o número de experimentos como peso da equação.

Outro método de reamostragem é o *Bootstrap*, (EFRON, 1979), que consiste em, a partir do conjunto de variáveis dependentes preditas de tamanho NE após a estimação de parâmetros, gerar novas amostras de mesmo tamanho, escolhendo aleatoriamente e com reposição os valores iniciais; sendo assim possível obter pseudoparâmetros para cada iteração e, analogamente ao *Jackknife*, calcular a suas médias (SILVA FILHO, 2010).

$$\bar{\theta}_B = \frac{1}{B} \sum_{j=1}^B \theta_{(j)} \quad (2.43)$$

Onde  $B$  é o número de amostras geradas aleatoriamente,  $\theta_j$  é o valor dos pseudoparâmetros de cada conjunto e  $\bar{\theta}_B$  é o valor médio dos pseudoparâmetros.

A maioria dos trabalhos que apresentam o *Bootstrap* o utiliza para a construção de intervalos de confiança, já JOSHI, SEIDEL-MORGENSTERN e KREMLING, (2006) , através de réplicas de experimentos, utilizam o desvio entre eles para gerar diferentes conjuntos de parâmetros e para construir os conjuntos *Bootstrap* para delimitar a região de confiança.

EFRON, (1979), em seu trabalho de introdução do método, demonstra que o *Jackknife* é um caso particular do *Bootstrap*, sendo uma aproximação linear. Também, BRADLEY e GONG, (1983), levantam dois pontos em seu trabalho de comparação entre os métodos: as reamostragens do *Bootstrap* são mais precisas enquanto que o cálculo do *Jackknife* é computacionalmente mais viável. Ambos os motivos são explicados pelo fato já apresentado, o *Jackknife* se trata de uma aproximação linear, desta forma, o cálculo é mais rápido devido a simplificação do método, mas não representa adequadamente todos os modelos (BRADLEY; GONG, 1983).

Neste trabalho o algoritmo *Bootstrap* é implementado utilizando uma abordagem que difere da metodologia padrão. Ao invés de perturbar os valores preditos pelo modelo, serão as variáveis dependentes experimentais que sofrerão modificações em seus valores. A ideia principal da alteração imposta ao método é que

seus resultados não sejam influenciados pelo erro contido nas predições, ou seja, a qualidade dos parâmetros gerados será função apenas do erro experimental. A perturbação gerada para garantir a reamostragem é provocada por valores aleatórios normalmente distribuídos entre o intervalo das variâncias dos dados experimentais. Assim, aproveitando o conhecimento destes valores, pode-se utilizá-los para reduzir a incerteza dos parâmetros, melhorando a qualidade da estimação.

Para as técnicas de construção da região de confiança dos parâmetros é possível ressaltar alguns aspectos. Os estudos das regiões surgiram nos anos 60 com a abordagem de desenvolver técnicas que melhorassem a confiabilidade de pesquisas e processos. A ideia inicial destes trabalhos eram determinar a solução analítica dessas regiões, ou seja, determinar uma solução geral que fosse possível aplicar em quaisquer tipos de modelos (BEALE, 1960; HARTLEY, 1964; WILLIAMS, 1962). Contudo, como foi mencionado, a solução algébrica considera o cálculo da matriz Hessiana dos parâmetros, que pode possuir muitas dimensões e, dessa forma, seu cálculo se tornava trabalhoso para os recursos deste período e havia a possibilidade de não possuir inversa, o que inviabiliza seu cálculo. Alguns métodos alternativos, de proposta semelhante à dos métodos estocásticos surgiram nessa época, mas a necessidade de cálculos repetitivos em grande escala era inviável para o processamento computacional que estava disponível.

A melhora do processamento computacional conforme o passar dos anos possibilitou a exploração mais ampla das técnicas estocásticas, visto que o tempo gasto com cálculos tenderia a diminuir. É possível perceber em trabalhos da década de 90, que embora se tenha o incremento na exploração destas técnicas, o número de avaliações de determinado cálculo era baixo ou o tempo gasto era significativo. Pesquisando o termo “*confidence regions*” nas publicações hospedadas pela CAPES, (2020), são encontrados os seguintes resultados da Tabela 1 segregados por década.

Tabela 1: Números de publicações encontradas envolvendo o uso das regiões de confiança para cada década desde 1960.

<b>Período</b>	<b>Número de publicações</b>
1960 - 1969	52
1970 - 1979	118
1980 - 1989	257
1990 - 1999	827
2000 - 2009	2.003
2010 - 2019	4.003

O melhoramento contínuo do desempenho dos computadores permitiu, que alguns dos métodos custosos do século passado fossem executados rapidamente. Assim, foi possível o aumento nos estudos envolvendo técnicas numéricas ou que envolvam que determinado cálculo seja executado inúmeras vezes e, assim, permitindo que modelos com soluções que exijam o uso do cálculo numérico fossem estudados sem impedimentos relacionados à sua solução analítica.

## Capítulo 3: Metodologia

---

Este capítulo é estruturado por três partes: a primeira consiste em detalhar os modelos utilizados na estimação de parâmetros, bem como os valores experimentais utilizados; a segunda parte é constituída pelas metodologias utilizadas nos procedimentos de minimização da função objetivo e da estimação de parâmetros; a terceira e última parte consiste em explicar as configurações empregadas nos cálculos das regiões de confiança para os métodos empregados.

### 3.1. Modelos e dados experimentais

Os modelos escolhidos para o estudo proposto neste trabalho possuem regiões de confiança distintas, sendo compostos por modelos com dois e três parâmetros.

Os modelos com dois parâmetros consistem em um modelo linear e três modelos não-lineares, enquanto, para os casos com três parâmetros, há um modelo linear e dois não-lineares. O primeiro modelo apresentado é o linear com dois parâmetros, sendo representado pela equação de reta a seguir, definida pela Equação (3.1), que para simplificar a nomenclatura será chamado de Modelo 1.

$$y = \theta_1 + \theta_2 x \quad (3.1)$$

Os valores das variáveis independentes e dependentes utilizados na estimação são apresentados na Tabela 2: Valores das variáveis independentes e dependentes para o modelo linear da Equação (3.1), Modelo 1.. Estes dados foram obtidos com a Equação (3.1) igualando os valores dos parâmetros  $\theta_1$  e  $\theta_2$  a 5,0. Os valores das variáveis dependentes obtidos foram perturbados com erros aleatórios distribuídos normalmente com média 0,0 e desvio padrão igual a 2,0.

Tabela 2: Valores das variáveis independentes e dependentes para o modelo linear da Equação (3.1), Modelo 1.

<b>x</b>	<b>y</b>
1,0	9,92
2,0	16,89
3,0	17,12
4,0	26,03
5,0	30,71
6,0	33,28
7,0	39,83
8,0	42,44
9,0	50,44
10,0	53,63

Para representar o modelo não-linear com dois parâmetros, foi utilizada a Equação (3.2), apresentada por BATES e WATTS, (1988), no Anexo A1.3 de sua obra, correspondendo ao modelo de Michaelis-Menten para a cinética de reações enzimáticas. Essa equação representa os Modelos 2.A e 2.B quando utilizadas com os dados experimentais disponíveis nas Tabela 3 e Tabela 4 respectivamente, os quais são disponibilizados nas Tabela A1.3 da mesma obra.

$$y = \frac{\theta_1 x}{\theta_2 + x} \quad (3.2)$$

Os valores de variáveis dependentes e independentes utilizados para a confecção do Modelo 2.A estão descritos na Tabela 3.

Tabela 3: Valores das variáveis independentes e dependentes para o modelo não-linear da Equação (3.2), Modelo 2.A.

<b>x</b>	<b>y</b>
0,02	76
0,02	47
0,06	97
0,06	107
0,11	123
0,11	139
0,22	159
0,22	152
0,56	191
0,56	201
1,10	207
1,10	200

Já os valores de variáveis dependentes e independentes utilizados para a confecção do Modelo 2.B são descritos na Tabela 4.

Tabela 4: Valores das variáveis independentes e dependentes para o modelo não-linear da Equação (3.2), Modelo 2.B.

<b>x</b>	<b>y</b>
0,02	67
0,02	51
0,06	84
0,06	86
0,11	98
0,11	115
0,22	131
0,22	124
0,56	144
0,56	158
1,10	160

Para representar o último modelo não-linear com dois parâmetros, a Equação (3.3) é definida. Essa equação é encontrada em diversas situações, como, por

exemplo, para definir a cinética de primeira ordem de uma reação em proporções estequiométricas entre dois reagentes.

$$y = \theta_1(1 - \exp(-\theta_2 x)) \quad (3.3)$$

Este modelo também pode caracterizar a demanda bioquímica de oxigênio presente em amostras de água como o exemplo disponibilizado por BATES e WATTS, (1988), com o conjunto de dados apresentados na Tabela 5 no Anexo A1.4 de sua obra, a qual caracteriza o Modelo 3.

Tabela 5: Valores das variáveis independentes e dependentes para o modelo não-linear da Equação (3.3), Modelo 3.

x	y
1,0	8,3
2,0	10,3
3,0	19,0
4,0	16,0
5,0	15,6
7,0	19,8

Os modelos com três parâmetros são os próximos a serem apresentados. O primeiro modelo dessa classe se trata de um linear em relação aos parâmetros, caracterizando uma parábola. O modelo linear escolhido é representado pela Equação (3.4) a seguir e caracteriza o Modelo 4.

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 \quad (3.4)$$

Os valores de suas variáveis independentes e dependentes são apresentados na Tabela 6. Estes dados foram obtidos com a Equação (3.4) igualando os valores dos parâmetros  $\theta_1$ ;  $\theta_2$  e  $\theta_3$  a 10, 2,0 e 3,0 respectivamente. Os valores das variáveis dependentes obtidos foram perturbados com erros aleatórios distribuídos normalmente com média 0,0 e variância igual a 2,5.

Tabela 6: Valores das variáveis independentes e dependentes para o modelo linear com três parâmetros da Equação (3.4), Modelo 4.

x	y
1,0	17,29
2,0	24,91
3,0	40,05
4,0	61,32
5,0	88,17
6,0	121,77
7,0	159,32
8,0	202,47
9,0	248,99
10,0	303,36
11,0	361,36

Os modelos não-lineares com três parâmetros tendem a ser mais complexos do que os casos com dois parâmetros, devido à existência de um número maior de interações entre os parâmetros. Para o estudo deste trabalho o modelo não-linear escolhido é definido pela Equação (3.5) e representa a isoterma de adsorção proposta por SIPS, (1948).

$$y = \frac{\theta_1 (\theta_2 x)^{\theta_3}}{1 + (\theta_2 x)^{\theta_3}} \quad (3.5)$$

Os valores das variáveis dependentes e das variáveis independentes experimentais são apresentados na Tabela 7 e, junto com o modelo apresentado pela Equação (3.5) passam a ser chamados de Modelo 5.A. Os dados foram obtidos igualando os valores dos parâmetros  $\theta_1$ ,  $\theta_2$  e  $\theta_3$  a 100, 0,1 e 0,9 respectivamente. Os valores das variáveis dependentes obtidos foram perturbados com erros aleatórios distribuídos normalmente com média 0,0 e variância igual a 2,0.

Tabela 7: Valores das variáveis independentes e dependentes para o modelo não-linear com três parâmetros da Equação (3.5), Modelo 5.A.

<b>x</b>	<b>y</b>
1,0	11,90
5,0	36,44
15,0	57,75
25,0	68,20
35,0	72,89
50,0	79,63
75,0	83,35
100,0	85,79
150,0	89,78
200,0	88,67

O último modelo a ser estudado neste trabalho também utiliza a isoterma de SIPS, (1948), Equação (3.5), e o conjunto de variáveis dependentes e independentes apresentados na Tabela 8, chamado de Modelo 5.B. Os valores dessa tabela foram gerados igualando-se os valores dos parâmetros  $\theta_1$ ;  $\theta_2$  e  $\theta_3$  a 100, 0,1 e 0,9 respectivamente, sendo que, estes resultados das variáveis dependentes foram perturbados com erros aleatórios distribuídos normalmente com média 0,0 e variância igual a 4,0. A principal diferença entre este modelo e o Modelo 5.A está na variância utilizada para gerar as perturbações no conjunto de dados, sendo maior para este.

Tabela 8: Valores das variáveis independentes e dependentes para o modelo não-linear com três parâmetros da Equação (3.5), Modelo 5.B.

<b>x</b>	<b>y</b>
5,0	118,76
10,0	130,80
20,0	137,77
40,0	137,75
60,0	140,22
80,0	139,56
90,0	143,18
110,0	138,45
130,0	139,26

Na Tabela 9, são demonstrados o número de dados experimentais utilizados para cada modelo e o seu número de parâmetros. Ambos os valores são utilizados no cálculo dos valores das distribuições estatísticas utilizadas neste trabalho.

Tabela 9: Valores do número de experimentos e de parâmetros de cada modelo utilizado neste trabalho.

<b>Modelo</b>	<b>NE</b>	<b>NP</b>
1	10	
2.A	12	2
2.B	11	
3	6	
4	11	
5.A	10	3
5.B	9	

### 3.2. *Estimação de parâmetros*

Os procedimentos de estimação de parâmetros, como foi discutido no Capítulo 2, consistem na minimização do valor da função objetivo através de sucessivas iterações, na qual, a cada uma, um novo valor para o conjunto de parâmetros é utilizado até que os seus melhores valores, correspondente ao ponto ótimo, sejam determinados.

A função objetivo escolhida para este estudo se trata da função de mínimos quadrados. Sua escolha se deve pelo fato de seu cálculo ser simples e não ser computacionalmente extensivo. Outro motivo para o uso desta função se dá pelo fato de só haver uma variável dependente, não ocorrendo correlação com as demais e seu erro pode ser considerado constante. A exceção para o emprego desta função se deu nos casos em que o cálculo é realizado para as regiões *Bootstrap*, que por considerarem a variância dos dados experimentais, o uso da função do mínimos quadrados ponderados se torna mais indicado.

Como descrito anteriormente, existem várias metodologias que podem ser utilizadas para minimizar o valor da função objetivo, entre elas há: os métodos baseados nos conceitos do cálculo diferencial, que asseguram a rápida convergência do algoritmo ao ponto ótimo, contudo, dependendo do modelo, este pode ser um ótimo local, resultando em uma estimação errônea; os métodos estocásticos, que realizam

inúmeras iterações em toda a região de busca e garantem a obtenção do ótimo global, porém estes não conseguem convergir ao ponto exato de ótimo, concluindo as iterações em sua proximidade.

Posto isso, a abordagem utilizada neste trabalho consiste em um método híbrido envolvendo as duas formas de minimização da função objetivo apresentadas. A primeira parte consiste em otimizar o valor da função objetivo através do método estocástico do Enxame de Partícula e a segunda se baseia em concluir a otimização através do método de Newton utilizando a aproximação de Gauss-Newton (SCHWAAB *et al.*, 2008; SCHWAAB; ALBERTON; PINTO, 2010; SCHWAAB; PINTO, 2007a)

As vantagens de inicializar a minimização com o método estocástico incluem: não necessitar de uma estimativa para o valor inicial dos parâmetros próxima do ponto ótimo, já que todas as partículas do enxame estão aleatoriamente distribuídas no intervalo de busca, exploração total do espaço de busca através das partículas distribuídas no espaço, assegurando a convergência ao ótimo global; robustez para otimizar diferentes tipos de modelos, visto que os métodos que utilizam derivadas possuem dificuldade em minimizar funções cujos modelos apresentam não-linearidade acentuada.

A escolha do Enxame de Partículas se deu devido à sua fácil implementação, já que não são necessárias as definições de muitos parâmetros ou ajustes para o método; velocidade de convergência, visto que sua metodologia garante o avanço rápido das partículas em direção ao ótimo e não há a necessidade de inúmeros cálculos a cada iteração, quando comparado com outros métodos estocásticos e robustez, sendo capaz de solucionar modelos lineares e não-lineares sem dificuldades.

A partir das posições iniciais das partículas geradas aleatoriamente, a velocidade de cada uma é determinada pela Equação (2.18) que considera o valor do peso de inércia proposto por SHI e EBERHART (1998). O uso desta adaptação garante o controle e balanceamento entre a exploração global e local da região de busca. Para que seja possível conciliar ambas as situações, o valor do peso de inércia é atualizado a cada iteração conforme a reta decrescente caracterizada pela Equação (2.19). A escolha de um valor mais elevado inicialmente resulta na melhor exploração do espaço de busca dos parâmetros e a sua regressão a um valor mais baixo garante

a rápida convergência do algoritmo após determinado número de iterações. Após, a posição da partícula é atualizada com o valor da velocidade através da Equação (2.17) e o algoritmo é repetido até que o valor da função objetivo convirja.

Os valores definidos para minimização através do Enxame de Partículas utilizado em todos os modelos foram:  $c_1$  e  $c_2$  iguais a 1,5, assegurando pesos iguais para a atração ao melhor ponto encontrado pela partícula e para o melhor ponto encontrado pelo enxame;  $W_o$  e  $W_f$  iguais a 0,95 e 0,75 respectivamente, permitindo a inicialização com característica exploratória e, conforme avançam as iterações, a exploração local para que o método convirja. Para cada modelo, o número de partículas do enxame foi definido igual a 100 e o número iterações para cada uma foi estipulado igual a 1.000, totalizando 100.000 avaliações da função objetivo. Dessa forma, visa-se assegurar que o espaço seja bem explorado e o mínimo encontrado.

A minimização da função objetivo utilizado o Enxame de Partículas não necessita de uma estimativa para o valor inicial das partículas, contudo é necessário delimitar uma faixa de valores para restringir o espaço de busca das partículas e garantir que elas se distribuam de maneira adequando neste espaço, assegurando que não haja problemas para a convergência. Os limites dos parâmetros para cada modelo estudado neste trabalho são apresentados na Tabela 10.

Tabela 10: Valores dos limites inferior e superior utilizados para a distribuição das partículas do enxame no espaço dos parâmetros.

Modelos	Parâmetro	Limite Inferior	Limite Superior
1	$\theta_1$	0	50
	$\theta_2$	0	50
2.A	$\theta_1$	100	300
	$\theta_2$	0	1
2.B	$\theta_1$	50	250
	$\theta_2$	0	1
3	$\theta_1$	0	100
	$\theta_2$	0	10
4	$\theta_1$	0	100
	$\theta_2$	0	100
	$\theta_3$	0	10
5.A	$\theta_1$	90	105
	$\theta_2$	0,05	0,2
	$\theta_3$	0,3	12
5.B	$\theta_1$	132	147
	$\theta_2$	0	2,5
	$\theta_3$	0,3	2,7

Os limites para os Modelos 1, 2.A, 2.B, 3, 4 poderiam ser menores, mas, como a convergência observadas para eles não apresentou problemas, optou-se por mantê-los. Já para os Modelos 5.A e 5.B, devido à maior complexidade deste modelo não-linear foi necessário restringir melhor o intervalo de busca do algoritmo.

Após a determinação do valor dos parâmetros ótimos através do Enxame de Partículas, estes são utilizados como estimativa inicial do método de Gauss-Newton. O cálculo dos valores dos parâmetros através do método de Newton é realizado através da Equação (2.13) sendo que o valor da matriz Hessiana é substituído pela aproximação de Gauss-Newton, Equação (2.15), dessa forma, os problemas relacionados ao cálculo da inversa da matriz Hessiana são contornados e a aproximação é suficientemente boa para encontrar o valor do ótimo, visto que a

estimativa inicial é próxima de seu valor. A convergência do algoritmo se dá quando a variação do valor da função objetivo é suficientemente pequena em relação a uma tolerância pré-determinada (geralmente igual a  $10^{-6}$ ). Assim, a obtenção do melhor valor de parâmetros que satisfaz o mínimo do modelo é garantida.

### 3.3. *Regiões de confiança*

O principal escopo deste trabalho é determinar, a partir do valor ótimo, a região de confiança dos parâmetros para os modelos apresentados anteriormente. Esta ideia é concebida em duas análises distintas: a primeira parte desse trabalho consiste em comparar a qualidade das regiões obtidas utilizando a matriz Hessiana em relação aos parâmetros e a aproximação de Gauss-Newton para os modelos com dois parâmetros. A segunda parte consiste no estudo de algumas técnicas para gerar as regiões de confiança não elípticas para as estimativas dos parâmetros, em particular para modelos não-lineares.

Inicialmente tem-se a análise da região de confiança dos parâmetros obtida através das Equações (2.22) e (2.34). Como foi visto no Capítulo 2, esta equação descreve uma elipse centrada nos valores ótimos dos parâmetros. Essa equação foi obtida pela aproximação da função objetivo por uma série de Taylor truncada no termo da derivada de segunda ordem e foram consideradas hipóteses para que fosse possível estabelecer a distribuição  $F$  de Fisher para ela. A hipótese assumida para a construção da equação da elipse se dá para o termo da matriz Hessiana, na qual usualmente ele é substituído pela matriz de covariância dos parâmetros quando se considera a aproximação de Gauss-Newton. Para os modelos lineares, essa aproximação é exata devido ao fato que suas derivadas de segunda ordem são nulas. Já para os modelos não-lineares, essa aproximação pode não ser uma boa escolha, já que mesmo quando os desvios entre os valores experimentais e os preditos são pequenos, se as derivadas segundas do modelo em relação aos parâmetros são grandes, desconsiderar este termo pode levar a uma região elíptica de má qualidade.

Primeiramente, é preciso definir como será realizado o cálculo das derivadas de segunda ordem da função objetivo; podendo-se escolher entre os métodos analíticos ou numéricos. A escolha do método analítico resulta na obtenção dos valores exatos das derivadas e, conseqüentemente, maior confiabilidade nos resultados. Contudo, a determinação algébrica de cada uma das derivadas pode ser

difícil, pois depende da complexidade do modelo, e seria necessário refazer esta derivação sempre que um novo modelo fosse estudado.

O cálculo das derivadas de segunda ordem pelo método numérico se torna vantajoso já que é necessário o uso de somente uma rotina genérica para qualquer modelo, não sendo preciso determinar a sua expressão para cada novo modelo testado. Por se tratar de um método numérico, haverá um erro quando comparado como o seu valor exato, dessa forma também existem ressalvas ao seu uso. Uma comparação envolvendo a qualidade da aproximação numérica da matriz Hessiana é demonstrada no Apêndice A.4 deste trabalho.

Pelo método numérico, o cálculo da matriz Hessiana é realizado utilizando duas expressões obtidas de NOCEDAL e WRIGHT, (2006), e detalhadas no Apêndice A.1 deste trabalho: a primeira corresponde ao valor da derivada de segunda ordem em função à um parâmetro;

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{f(x + \Delta x_i e_i) - 2f(x) + f(x - \Delta x_i e_i)}{\Delta x_i^2} + O(\Delta x_i) \quad (3.6)$$

a segunda correspondo ao valor da derivada de segunda ordem em respeito à mais de um parâmetro.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{f(x + \Delta x_i e_i + \Delta x_j e_j) - f(x + \Delta x_i e_i) - f(x + \Delta x_j e_j) + f(x)}{\Delta x_i \Delta x_j} + O(\sqrt{\Delta x_i \Delta x_j}) \quad (3.7)$$

Onde  $f(x)$  é o valor da função para o ponto  $x$ ,  $\Delta x_i$  e  $\Delta x_j$  são perturbações pequenas, escalares e positivas nas direções do parâmetro  $i$  e  $j$ , respectivamente;  $e_i$  e  $e_j$  são vetores unitários nas direções dos parâmetros  $i$  e  $j$ , respectivamente e  $O(x)$  representa o erro na ordem de  $x$ .

O cálculo da equação elíptica é realizado através da Equação (2.34), o procedimento algébrico para a sua resolução é detalhado na Apêndice A.2 deste trabalho. A partir do valor ótimo dos parâmetros, é possível explicitar um deles como uma variável independente e, dessa forma, o utilizar para calcular o valor dos demais, assim, é possível estabelecer um conjunto solução de parâmetros capaz de desenhar a elipse característica da região de confiança.

O valor da matriz de covariâncias dos parâmetros presente na Equação (2.34) é substituído pela sua aproximação com a matriz Hessiana. Quando esta é substituída pela Equação (2.14), tem-se a matriz Hessiana da função objetivo em relação aos parâmetros, sendo que os valores das derivadas de segunda ordem do modelo são calculados através das Equações (3.6) e (3.7). Já quando este termo é substituído pela Equação (2.15), tem-se a aproximação de Gauss-Newton, que desconsidera as derivadas de segunda ordem pois assume que os desvios entre as medições experimentais e as preditas são pequenos, dessa forma, só as derivadas de primeira ordem impactariam no cálculo da região de confiança.

Como as regiões elípticas possuem seu traçado explícito por uma equação, não é necessário um valor elevado de pontos para a traçar, desta forma, todas as regiões elípticas apresentadas foram construídas com 100 pontos. Seus contornos são limitados pelo valor da distribuição  $F$  de Fisher como  $NP$  e  $NE - NP$  graus de liberdade e nível de confiança de  $1 - \alpha$ . O valor de significância,  $\alpha$ , é igual a 0,05, resultando em um nível de confiança equivalente a 95%. Para os casos analisados na primeira parte do Capítulo 4, são demonstrados os valores das matrizes Hessianas da função objetivo para as duas abordagens.

Para a segunda parte deste estudo, foram utilizados alguns métodos capazes de gerar as regiões de confiança, sendo que as vantagens e desvantagens foram abordadas no Capítulo 2. O foco deste trabalho foi o estudo das regiões de confiança resultantes do uso de três diferentes metodologias: o método do Contorno, o método do Perfil e o método *Bootstrap*.

Uma metodologia proposta neste trabalho se trata da região do Contorno. A principal ideia na concepção deste método está em fazer o algoritmo “caminhar” sobre os limites da região de confiança delimitando o seu traçado. A ideia principal é, a partir do ponto ótimo, alterar o valor de um dos parâmetros até que o algoritmo encontre um ponto sobre a curva que delimita a região de confiança; após o mesmo avança ao longo do seu traçado até retornar à posição inicial. Após, tendo como valor inicial o ponto extremo encontrado, o avanço é feito através de um passo com tamanho fixo em cada etapa e ajustando o ângulo no plano de forma que o ponto final do passo esteja em cima da curva. Este ajuste do ângulo do passo em um problema com 2 parâmetros é feito resolvendo-se um problema de equação algébrica não-linear com o método de Newton-Raphson. A principal característica deste método se dá pela

diminuição automática do tamanho do passo, nas regiões com maior excentricidade para a curva limite da região de confiança, necessárias para delimitar bem as regiões onde ocorre uma rápida mudança de direção.

Para modelos com mais de dois parâmetros, há uma mudança na abordagem do cálculo da região pelo método do Contorno. Para determinado par de parâmetros, é necessário maximizar a distância dele em relação ao ponto ótimo, contudo, este valor deve estar sujeito às seguintes restrições: o valor da função objetivo deve ser igual ao valor crítico que delimita a região de confiança, de forma que os demais parâmetros também podem ser modificados. Assim, para estes casos este método se torna mais complexo e sujeito a erros. Para modelos com dois parâmetros, não existem muitos problemas, pois o valor da função objetivo dependerá, unicamente, dos parâmetros que estão sendo modificados para encontrar o contorno. Já, para modelos com mais de dois parâmetros, a qualidade do funcionamento do método pode ser condicionada pelo tipo de região, visto que o valor da restrição dependerá dos outros parâmetros do modelo, dificultando a ação da otimização. Destacam-se alguns dos problemas: como o esgotamento do número de iterações ou o retorno do algoritmo pelo caminho já executado, contudo, espera-se que o comportamento a ser observado seja similar ao dos casos com dois parâmetros. Os detalhes do funcionamento deste método são demonstrados no Algoritmo 1 do Apêndice A.5.

O segundo método estudado para determinar as regiões de confiança dos parâmetros se trata do Perfil proposto por BATES; WATTS, (1988), para descrever a linearidade de um modelo. Para determinar as regiões de confiança através deste método é preciso, primeiramente, determinar o perfil para cada parâmetro. Seus valores são calculados pela Equação (2.36) na qual se considera o resultado da função objetivo do modelo quando o parâmetro é desconsiderado da matriz de derivadas de primeira ordem durante a etapa de minimização. O perfil é calculado para o valor ótimo, correspondendo a  $\tau$  igual a 0 e para os demais pontos até que se atinja o valor de  $\tau_{MAX}$ , que corresponde aos valores da distribuição *t-Student* com  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$ , sendo o valor de significância,  $\alpha$ , igual a 0,05.

Relacionando um par de perfis é possível formar os traços, que consistem na interação entre o resultado minimizado,  $\tilde{\theta}_i$ , do parâmetro do perfil,  $\theta_i$ , com outro parâmetro do par,  $\theta_j$ ; assim, os traços são descritos com uma sobreposição das

curvas  $\tilde{\theta}_i(\theta_j)$  e  $\tilde{\theta}_j(\theta_i)$  (BATES; WATTS, 1991). Estes são utilizados para a confecção do esboço, o qual corresponde a representação da região de confiança de BATES e WATTS, (1988). A partir da localização dos pontos onde a região intercepta os traços, têm-se quatro pontos pertencentes ao contorno da curva. Somado a isso, há o fato que a tangente da curva é vertical nestes pontos, pois estes independem da direção do parâmetro  $\theta_i$ . Assim, um método é descrito nos Apêndices 3 e 6 da obra dos autores. Como o valor dessas interpolações é submetido a funções trigonométricas, e, após o devido tratamento matemático, os valores onde a região intercepta os traços são convertidos à escala de ângulos e fases, a escolha de um método adequado de interpolação se faz essencial. Dessa forma, assim como sugerido pelos autores, o método do *spline* cúbico foi selecionado para interpolar os dados em todas as etapas necessárias. Mais detalhes desta técnica são discutidos no Apêndice A.3 deste trabalho. Os detalhes do funcionamento deste método estão disponíveis no Algoritmo 2 do Apêndice A.5.

O terceiro método para delimitar as regiões de confiança dos parâmetros a ser explorado neste trabalho é o *Bootstrap*. Como visto no Capítulo 2, este é um método de reamostragem, o qual evita a necessidade de realizar diversos experimentos para obter uma amostra de tamanho significativo. O método *Bootstrap* clássico consiste em estimar o valor ótimo dos parâmetros e gerar as variáveis dependentes preditas, sendo estas que serão perturbadas para resultar em um novo valor para os parâmetros que corresponderão aos valores “ótimos” perturbados.

Aqui é sugerida uma adaptação, os valores perturbados foram as variáveis dependentes experimentais, as quais serão minimizadas para encontrar novos valores “ótimos” para os parâmetros. As perturbações provocadas são aleatórias e possuem distribuição normal. Seus valores foram limitados ao intervalo de 95% de confiança utilizando a Equação (3.8).

$$RDL = 0,975 - 0,950.(1 - RDN) \quad (3.8)$$

Onde RDL representa um número aleatório limitado e RDN representa um número aleatório entre 0 e 1. O valor de RDL é convertido em um valor aleatório com distribuição normal, média igual a zero e variância igual a 1. Em seguida, este valor é

multiplicado pela incerteza experimental, e somado ao valor da variável experimental, de forma a simular desvios para cada nova medição.

Para os Modelos 1, 4, 5.A, 5.B, nos quais os valores experimentais foram gerados para este estudo, a obtenção da variância é simples, visto que a mesma foi utilizada para a obtenção dos dados experimentais. Já para os demais modelos, 2.A, 2.B e 3, cujos dados foram retirados da literatura, o valor da variância foi aproximado de acordo com a Equação (3.9).

$$s^2 = \frac{S(\theta)}{NE - NP} \quad (3.9)$$

Para todas as regiões obtidas utilizando o método *Bootstrap* foram realizadas 10.000 perturbações aleatórias nos valores das variáveis dependentes experimentais. Os detalhes do funcionamento deste método são demonstrados no Algoritmo 3 do Apêndice A.5.

Outro aspecto importante do método *Bootstrap* é que como a perturbação utilizada nas variáveis dependentes usa a variância experimental, foi considerado o uso da função dos mínimos quadrados ponderados, Equação (2.9), que se torna mais adequado, visto que o valor da incerteza será conhecido; assim, para as regiões de confiança obtidas através deste método, o valor da função objetivo diferirá das demais.

Todos os algoritmos utilizados para os processos de estimação de parâmetros e para os cálculos das regiões de confiança deste trabalho foram implementados na linguagem computacional Fortran 90 utilizando como interface principal o pacote computacional desenvolvido por SCHWAAB, ALBERTON e PINTO, (2010), para estimação de parâmetros e planejamento de experimentos. As figuras foram geradas com o uso da linguagem R. As rotinas computacionais deste trabalho foram executadas através do *Microsoft Visual Studio* em um computador com processador Intel Core i3 e memória RAM de 4 GB.

## Capítulo 4: Resultados

---

Neste capítulo são apresentados os resultados obtidos para o estudo das regiões de confiança dos parâmetros dos modelos apresentados no Capítulo 3. Primeiramente serão demonstrados e discutidos os resultados obtidos para o cálculo da região de confiança elíptica comparando o uso da matriz Hessiana completa dos parâmetros e da aproximação de Gauss-Newton. Após, são demonstrados os resultados para as regiões de confiança geradas com os métodos do Contorno, Perfil e *Bootstrap* e suas análises.

Aproveitando alguns conceitos e a metodologia das regiões de confiança, foi elaborado um estudo na forma de artigo que visa observar a incerteza da estimação de parâmetros na previsão do período final de contágio de COVID-19 para diferentes porções do mesmo conjunto de dados. Este estudo é apresentado no Apêndice A.6.

### 4.1. *Análise da região de confiança elíptica dos parâmetros*

Como já foi discutido, as Equações (2.22) e (2.34) descrevem uma elipse centrada nos valores ótimos dos parâmetros, a qual caracteriza uma região no espaço dos parâmetros cujos valores, estatisticamente, são iguais ao conjunto ótimo. Sua obtenção se dá pela aproximação da função objetivo em uma série de Taylor truncada no termo da derivada de segunda ordem. Após, utilizando as abordagens algébricas e estatísticas demonstradas no Capítulo 2, é possível determinar que ela segue a distribuição  $F$  de Fisher.

O principal dificultador deste cálculo é o termo da matriz Hessiana com a derivada segunda do modelo em relação aos parâmetros, que dependendo do tipo de modelo, pode ser de difícil obtenção. Assim a maior hipótese assumida para a construção da equação da região elíptica é a substituição dessa matriz pela matriz de covariância dos parâmetros, ou seja, quando se considera a aproximação de Gauss-Newton. Como foi explorado no Capítulo 2, essa aproximação considera que os desvios entre os valores das variáveis dependentes experimentais e as previstas pelo modelo são próximas a zero, desta forma é possível ignorar este termo mesmo para modelos não lineares.

Deve-se ter em mente que, para os modelos lineares, essa aproximação é exata devido ao fato que suas derivadas de segunda ordem são nulas, ou seja, a

matriz Hessiana do modelo em relação aos parâmetros é nula, tornando a Equação (2.24) verdadeira para esses modelos. Para modelos não-lineares, essa aproximação pode não ser uma boa escolha, já que mesmo quando os desvios entre os valores experimentais e os preditos são pequenos, se as derivadas segundas do modelo em relação aos parâmetros são altas, desconsiderar este termo pode levar a uma região elíptica de má qualidade.

Nesta seção será proposto o cálculo da região de confiança dos parâmetros utilizando a Equação (2.14), na qual será considerado o valor da matriz Hessiana da função objetivo em relação aos parâmetros do problema e, os mesmos serão comparados com a região de confiança elíptica descrita pela Equação (2.34), a qual considera a aproximação de Gauss-Newton.

Todas as minimizações da função objetivo para a estimação de parâmetros foram calculadas através do método híbrido do Enxame de Partículas com Gauss-Newton utilizando as configurações descritas no Capítulo 3. A obtenção da equação elíptica da região de confiança se deu através do procedimento indicado naquele capítulo utilizando os cálculos que estão detalhados no Apêndice A.2.

Para comparar a região elíptica considerando o cálculo da matriz Hessiana com a resultante da aproximação de Gauss-Newton e atestar se a região se comportará como esperada, foram utilizados modelos distintos, um linear e três não-lineares.

#### 4.1.1. Regiões de confiança elípticas de modelos lineares

A primeira análise a ser demonstrada é para o Modelo 1, o qual é linear. A escolha deste tipo de modelo é importante para validar o estudo, pois, como já argumentado, a aproximação de Gauss-Newton é exata para estes casos, dado que o valor de sua matriz Hessiana é nulo. O valor da função objetivo encontrado após a minimização deste modelo é igual a 20,75 e corresponde aos valores ótimos dos parâmetros  $\hat{\theta}_1$  e  $\hat{\theta}_2$  iguais a 5,40 e 4,84 respectivamente. A correlação obtida para este par de parâmetros é igual a -0,89 e o valor das covariâncias entre os parâmetros é demonstrada na matriz da Equação (4.1).

$$V_{\theta} = \begin{pmatrix} 1,21 & -0,173 \\ -0,173 & 3,14 \cdot 10^{-2} \end{pmatrix} \quad (4.1)$$

Inicialmente é interessante compararmos a qualidade da aproximação dos dados experimentais com os gerados pelo modelo, desta maneira é possível comparar graficamente os valores das variáveis independentes da Tabela 2 com os valores gerados pelo modelo quando substituimos na Equação (3.1) os valores das variáveis independentes da Tabela 2 e os valores ótimos dos parâmetros.

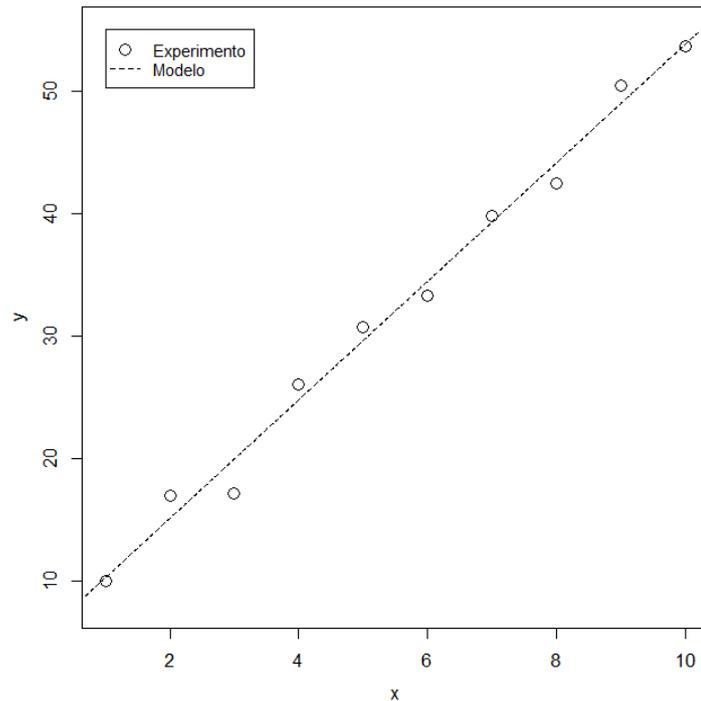


Figura 1: Comparação entre os valores de saída experimentais e os preditos para o Modelo1.

Como é possível observar na Figura 1, os valores obtidos experimentalmente, representados pelos pontos, estão próximos a reta pontilhada, que representa os valores gerados pelo modelo ajustado. É interessante apontar que a reta resultante do ajuste intercepta a região central dos pontos experimentais, sendo que estes estão distribuídos ora acima, ora abaixo da curva. Assim, tem-se a ideia de que o ajuste obtido é adequado, posto que ele tende a representar todos os pontos do conjunto, evitando seguir a algum comportamento específico.

Calculando os valores da matriz Hessiana pela Equação (2.14) e pela aproximação, Equação (2.15), obtemos o mesmo conjunto solução, visto que, como o termo das derivadas de segunda ordem da função objetivo em relação aos parâmetros é nulo para modelos lineares, as duas equações são idênticas para esta situação.

$$V_{\theta}^{comp} = \begin{pmatrix} 1,21 & -0,173 \\ -0,173 & 3,14 \cdot 10^{-2} \end{pmatrix} \quad (4.2)$$

Substituindo os valores da matriz apresentada na Equação (4.1) na Equação (2.27) é possível representar no plano cartesiano a região de confiança elíptica centrada no valor ótimo dos parâmetros do modelo linear, Figura 2. Como mencionado anteriormente, para modelos lineares, a aproximação de Gauss-Newton é exata, sendo assim, o caso estudado nessa seção se trata de uma peculiaridade e, para quaisquer modelos que sejam lineares, serão observadas situações idênticas a apresentada nessa figura, na qual o gráfico das regiões de confiança elípticas para ambos os casos estarão sobrepostos.

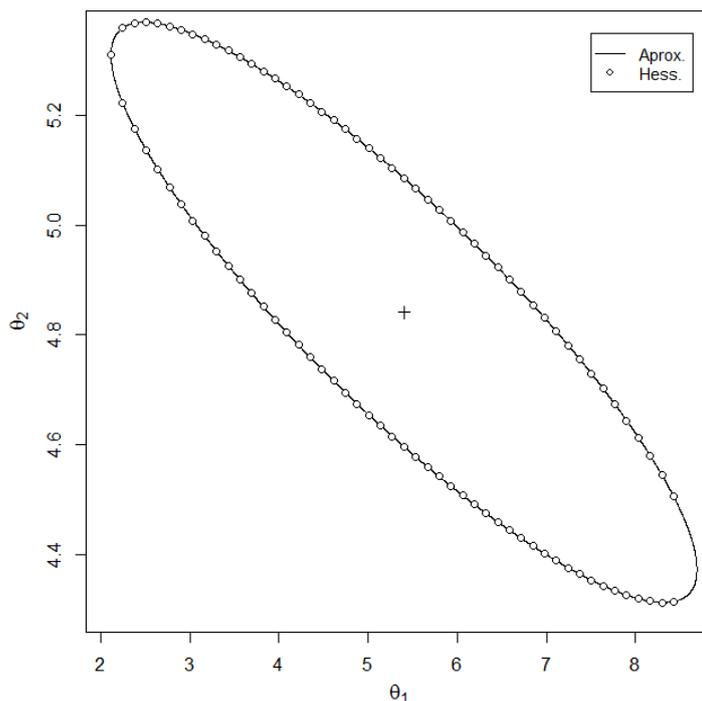


Figura 2: Regiões de confiança do Modelo 1 com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia).

O emprego desta situação permite que seja possível validar se as implementações propostas são consistentes e não apresentam nenhum erro grosseiro. Como as duas situações são representadas pela matriz da Equação (4.1), pode-se afirmar que o cálculo numérico da matriz Hessiana está correto, já que

quando os termos dessa matriz são nulos, a Equação (2.15) representa fielmente a região de confiança dos parâmetros. Assim, é possível utilizar este cálculo para representar as regiões de modelo não-lineares.

#### 4.1.2. Regiões de confiança elíptica de modelos não-lineares

Para validar o uso da aproximação de Gauss-Newton, torna-se essencial observar como a simplificação da Equação (2.14) influencia na região de confiança elíptica de modelos quando o valor da matriz Hessiana não é nulo. A qualidade é condicionada por dois fatores envolvendo essa equação: se a diferença entre os valores experimentais e os preditos é pequena, sendo próxima a zero; ou os valores das derivadas de segunda ordem do modelo são próximas a zero, dessa maneira, o modelo é quase-linear. Ambas as situações seriam condições ideais para justificar a validade da Equação (2.15).

Para observar se existe desvio entre o cálculo da matriz Hessiana pelas Equações (2.14) e (2.15), foi utilizado o Modelo 2.A, que se trata de um modelo não-linear. O valor da função objetivo após a minimização é igual a 1195,45, correspondendo aos valores ótimos dos parâmetros  $\hat{\theta}_1$  e  $\hat{\theta}_2$  iguais a 212,68 e 0,0641 respectivamente. A correlação observada entre os dois parâmetros do modelo é igual a 0,76 e a Equação (4.3) representa a matriz de covariâncias do modelo obtida através do cálculo da Aproximação de Gauss-Newton, Equação (2.15), e da Equação (2.24)

$$V_{\theta} = \begin{pmatrix} 48,26 & 4,40 \cdot 10^{-2} \\ 4,40 \cdot 10^{-2} & 6,86 \cdot 10^{-5} \end{pmatrix} \quad (4.3)$$

Os valores experimentais da Tabela 3 e os preditos com o Modelo 2.A após a obtenção dos valores ótimos dos parâmetros são apresentados e comparados na Figura 3. Como observado no ajuste dos dados do Modelo 1, Figura 1, nota-se que a curva predita tende a passar em um ponto médio entre as medições, ou seja, representando todo o conjunto de dados.

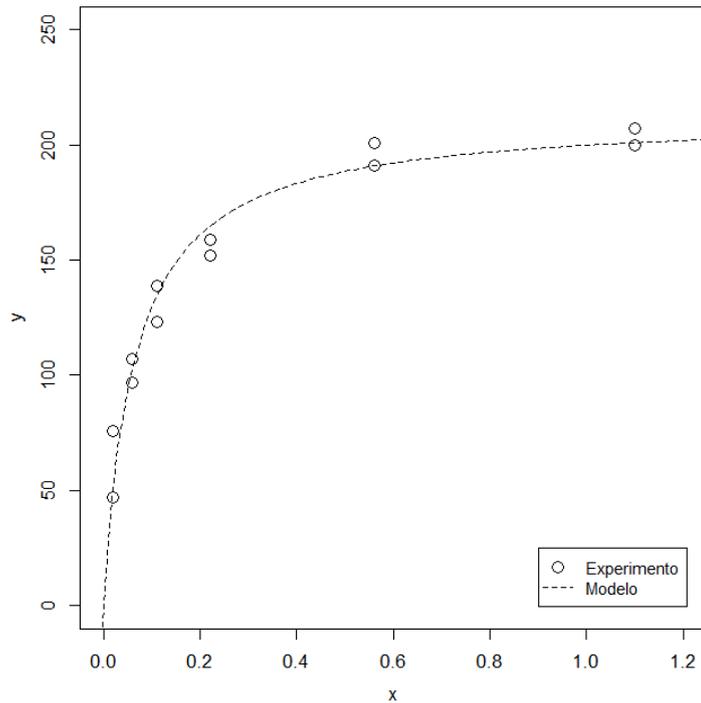


Figura 3: Comparação entre os valores de saída experimentais e os preditos para o Modelo 2.A.

O cálculo do valor da matriz Hessiana utilizando a Equação (2.14), ou seja, considerando as derivadas de segunda ordem do modelo, resulta nos valores da matriz de covariâncias abaixo.

$$V_{\theta}^{comp} = \begin{pmatrix} 51,27 & -4,97 \cdot 10^{-2} \\ -4,97 \cdot 10^{-2} & 7,59 \cdot 10^{-5} \end{pmatrix} \quad (4.4)$$

As matrizes apresentadas nas Equações (4.3) e (4.4) quando substituídas na equação da elipse, de forma análoga ao Modelo 1, representam duas elipses concêntricas, porém ambas diferem uma da outra em seus contornos devido ao desvio assumido pelo emprego da aproximação de Gauss-Newton, conforme demonstram as curvas da Figura 4. Nota-se que o tamanho da elipse resultante da equação que considera a matriz Hessiana é maior do que resultante da aproximação de Gauss-Newton, ou seja, a inclusão da matriz no cálculo provoca diferenças no contorno da região de confiança, principalmente em seus extremos.

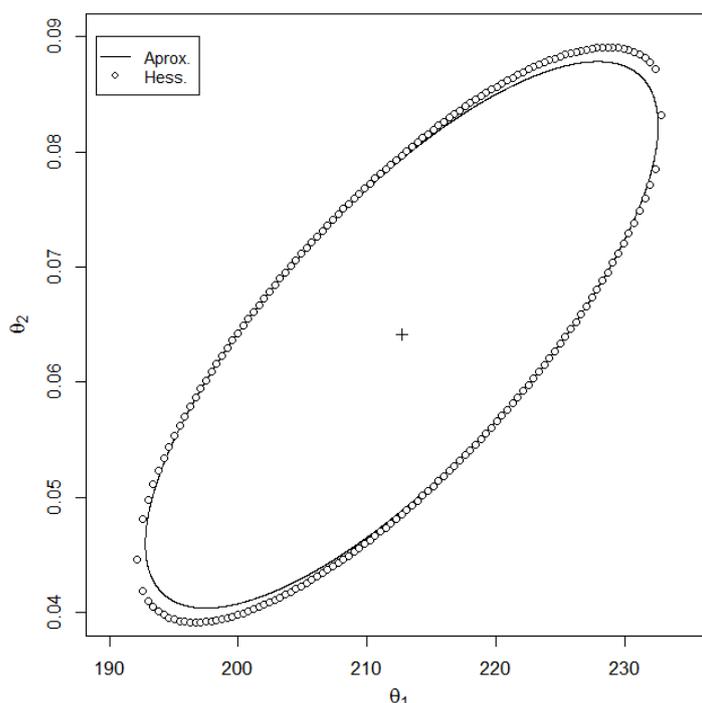


Figura 4: Regiões de confiança do Modelo 2.A com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia).

Para verificar se a diferença entre as elipses é mais acentuada do que a obtida no caso acima ou se este é um comportamento padrão, é interessante submeter o método a mais alguns modelos, como o Modelo 2.B, cuja a equação característica é mesma do Modelo 2.A, mas possui um conjunto diferente de dados experimentais. Assim, um novo conjunto de valores, Tabela 4, é utilizado com a Equação (3.2). O valor da função objetivo após a minimização é igual a 859,60, correspondendo aos valores ótimos dos parâmetros  $\hat{\theta}_1$  e  $\hat{\theta}_2$  iguais a 160,28 e 0,0477 respectivamente. O valor obtido para a correlação entre os parâmetros foi igual a 0,78 e as covariâncias dos parâmetros deste modelo, calculada pela Aproximação de Gauss-Newton são demonstrados na matriz da Equação (4.5).

$$V_{\theta} = \begin{pmatrix} 41,99 & -3,92 \cdot 10^{-2} \\ -3,92 \cdot 10^{-2} & 6,05 \cdot 10^{-5} \end{pmatrix} \quad (4.5)$$

Com este novo conjunto de valores, foram obtidos novos valores para a função objetivo e para os parâmetros ótimos. A comparação entre os valores obtidos experimentalmente, Tabela 4, e os preditos pelo modelo é demonstrada na Figura 5

e, é possível perceber que o ajuste está bem estruturado, visto que tende a compreender todos os pontos obtidos.

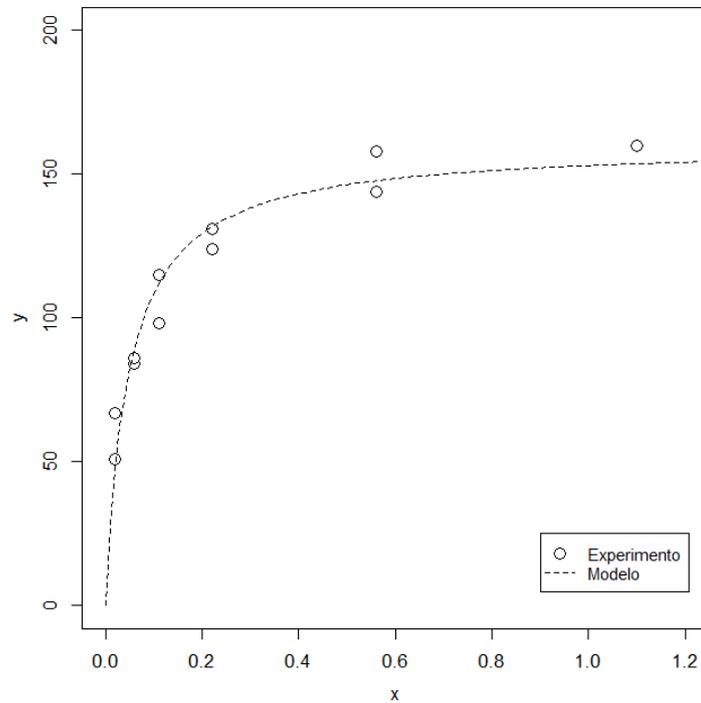


Figura 5: Comparação entre os valores de saída experimentais e os preditos para o Modelo 2.B.

O valor da matriz Hessiana dos parâmetros através da equação (2.14), que considera as derivadas de segunda ordem da função objetivo é definida pela Equação (4.6).

$$V_{\theta}^{comp} = \begin{pmatrix} 46,32 & -4,59 \cdot 10^{-2} \\ -4,59 \cdot 10^{-2} & 7,09 \cdot 10^{-5} \end{pmatrix} \quad (4.6)$$

Comparando-se as equações elípticas obtidas utilizando os valores das matrizes de covariâncias representadas nas Equações (4.5) e (4.6) para delimitar a região de confiança elíptica do Modelo 2.B, obtemos as curvas apresentadas na Figura 6. É possível observar que as elipses diferem de maneira semelhante à apresentada na Figura 4, onde a curva obtida considerando a aproximação de Gauss-Newton é menor do que resultante utilizando a matriz Hessiana dos parâmetros completa.

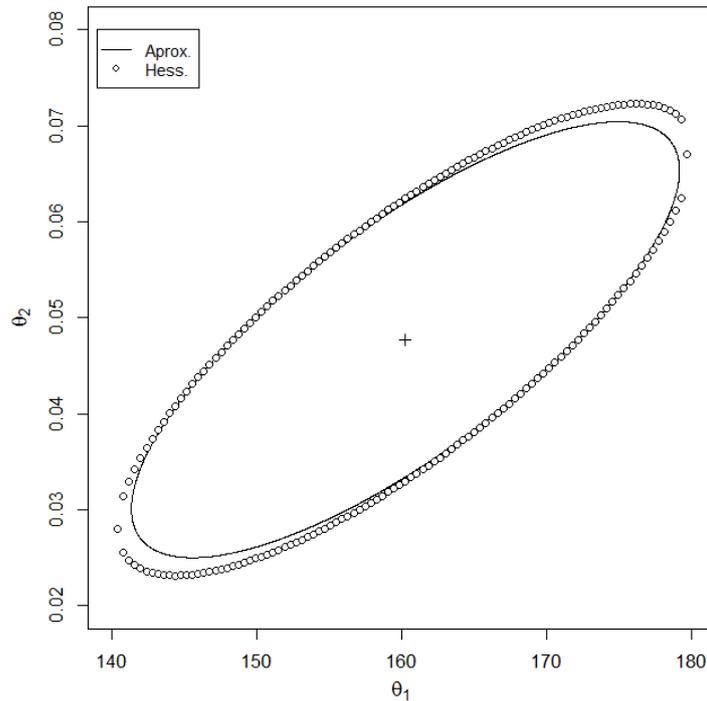


Figura 6: Regiões de confiança do Modelo 2.B com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia).

Outra importante observação é que, para os dois conjuntos de dados observados, a diferença entre o emprego da aproximação de Gauss-Newton e o da matriz Hessiana completa não foi significativa, visto que o aumento da região obtido utilizando a matriz Hessiana completa difere da aproximação de Gauss-Newton na região mais afastada do ponto ótimo dos parâmetros e o ganho obtido com o uso da matriz Hessiana é pequeno.

A observação da Figura 4 e da Figura 6 nos permite ter uma ideia da não-linearidade do conjunto modelo e dados experimentais pelo quanto a região que considera a matriz Hessiana completa difere da aproximação. Outra maneira de observar o desvio da linearidade se dá pelo emprego do método do Perfil, como visto no Capítulo 2. Os perfis permitem comparar o valor dos parâmetros reparametrizados pela Equação (2.36), resultantes da reparametrização da função objetivo quando a derivada de primeira ordem do parâmetro de estudo é anulada na minimização da função, com os valores fornecidos pela Equação (2.38).

Já os resultados da Equação (2.38) são obtidos pela parametrização dos parâmetros, descrevendo os valores lineares hipotéticos para o mesmo caso. É

possível observar a não-linearidade de um determinado modelo pelo desvio entre as curvas geradas pelos valores resultantes das duas equações.

Para comparar os dois modelos anteriores, foram observados os desvios da linearidade de cada um dos parâmetros dos modelos separadamente. A não-linearidade do parâmetro  $\theta_1$  pode ser observada na Figura 7.

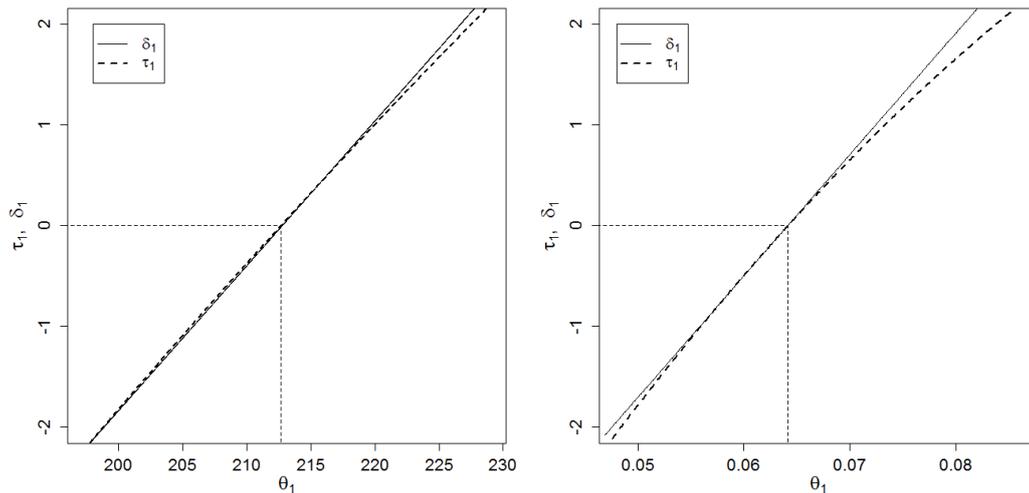


Figura 7: Comparação entre ao desvio da linearidade do parâmetro  $\theta_1$  do Modelo 2.A, Figura 7.a, e do Modelo 2.B, Figura 7.b.

Na Figura 7, a linha cheia representa o comportamento linear, enquanto a linha pontilhada representa a não-linearidade de cada uma das situações apresentadas para os Modelos 2.A e 2.B. No valor ótimo dos parâmetros, os valores de  $\tau_1$  e  $\delta_1$  são idênticos e iguais a zero, como esperado, visto que, o ponto mínimo deve ser o mesmo para ambas as situações. À medida que o valor dos parâmetros se afasta do ponto ótimo, cada um começa a demonstrar os efeitos de sua não-linearidade, onde a curva que a descreve se afasta da reta dos parâmetros parametrizados. Também é possível verificar que o caso da Figura 7.b é mais não-linear do que o apresentado pela Figura 7.a no parâmetro  $\theta_1$ ; sendo possível confirmar tal caso observado que a curva não linear do Modelo 2.B é mais acentuada do que a do Modelo 2.A, inferindo que aquele é o mais não-linear. A linearidade destes modelos está relacionada com os desvios entre o conjunto de dados experimentais e os preditos, visto que o modelo utilizado para ambos os casos é o mesmo.

A mesma análise pode ser feita para o parâmetro  $\theta_2$  dos Modelos 2.A e 2.B. Assim como para o parâmetro  $\theta_1$ , os valores das parametrizações são idênticos no ponto ótimo. Já para os parâmetros mais afastados do ponto de mínimo, é possível

notar que os desvio da linearidade do Modelo 2.B é maior do que a do Modelo 2.A, devido ao afastamento entre a curva e a reta das parametrizações.

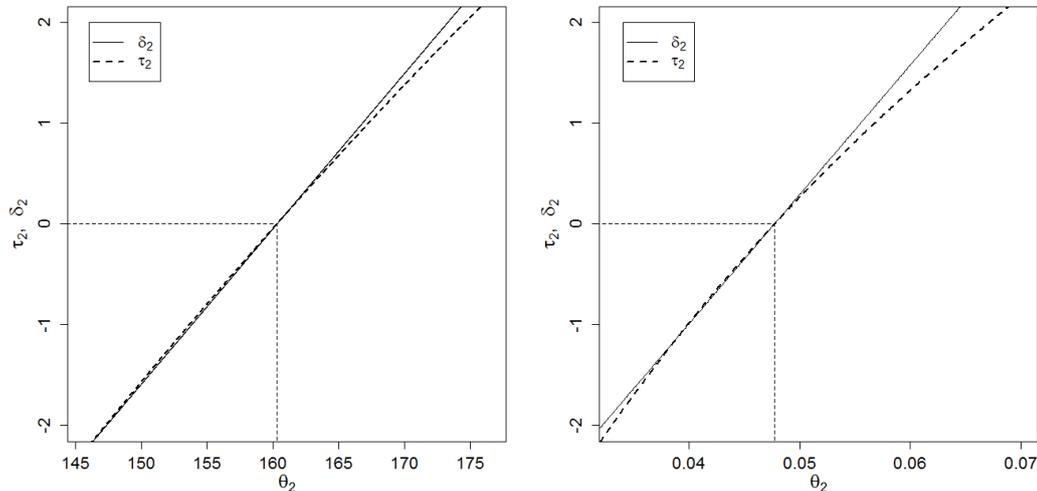


Figura 8: Comparação entre ao desvio da linearidade do parâmetro  $\theta_2$  do Modelo 2.A, Figura 8.a, e do Modelo 2.B, Figura 8.b.

O Modelo 2.B é o mais não-linear observando os perfis dos parâmetros  $\theta_1$  e  $\theta_2$  para os dois modelos. Os perfis dos dois parâmetros deste modelo desviam mais da linearidade dos que os do Modelo 2.A, e, assim, é possível relacionar este resultado com o traçado da elipse, já que o cálculo com a matriz Hessiana completa gera uma elipse com maior contorno do que a da aproximação de Gauss-Newton quando comparadas com as suas respectivas situações aproximadas. A diferença entre as elipses obtidas com as duas abordagens é mais acentuada para o Modelo 2.B do que para o Modelo 2.A, dessa forma é possível utilizar tanto a comparação entre as regiões elípticas como os perfis para indicar a não-linearidade de um modelo.

Outra ideia importante que é possível obter analisando os perfis é que o parâmetro  $\theta_1$  é o mais linear do modelo, visto que o seu desvio é menor do que o do apresentado pelo parâmetro  $\theta_2$ . Esse fato também é observável nas derivadas de segunda ordem analíticas do modelo, já que essa derivada é nula em função do parâmetro  $\theta_1$ , enquanto que, para o parâmetro  $\theta_2$ , ela depende de ambos os parâmetros.

A análise das regiões de confiança elípticas foi observada para a mesma equação do modelo não-linear, então é importante observar outro tipo para que seja possível definir se o comportamento observado nos casos acima é uma exclusividade

do tipo de modelo que foi empregado ou uma situação aplicável a quaisquer uns. Assim, a última análise das regiões de confiança elípticas será realizada para o Modelo 3. O valor encontrado para a função objetivo após a minimização é igual a 25,99, correspondendo aos valores ótimos dos parâmetros  $\hat{\theta}_1$  e  $\hat{\theta}_2$  iguais a 19,14 e 0,5311 respectivamente. A correlação observada entre os parâmetros é igual a -0,85 e as covariâncias dos parâmetros, calculadas através da Aproximação de Gauss-Newton, estão demonstradas na matriz da Equação (4.7).

$$V_{\theta} = \begin{pmatrix} 6,23 & -0,432 \\ -0,432 & 4,12 \cdot 10^{-2} \end{pmatrix} \quad (4.7)$$

Na Figura 9 são comparados os dados obtidos experimentalmente, Tabela 5, e os preditos pelo Modelo 3 utilizando os valores dos parâmetros. O ajuste segue a tendência demonstrada nos demais modelos vistos anteriormente, Figura 1, Figura 3 e Figura 5, na qual os pontos gerados pelo modelo se encontram entre os experimentais. Também se nota que, entre os modelos já apresentados, este é o que apresenta maior erro na qualidade do ajuste, devido a distância acentuada entre os valores medidos e os preditos.

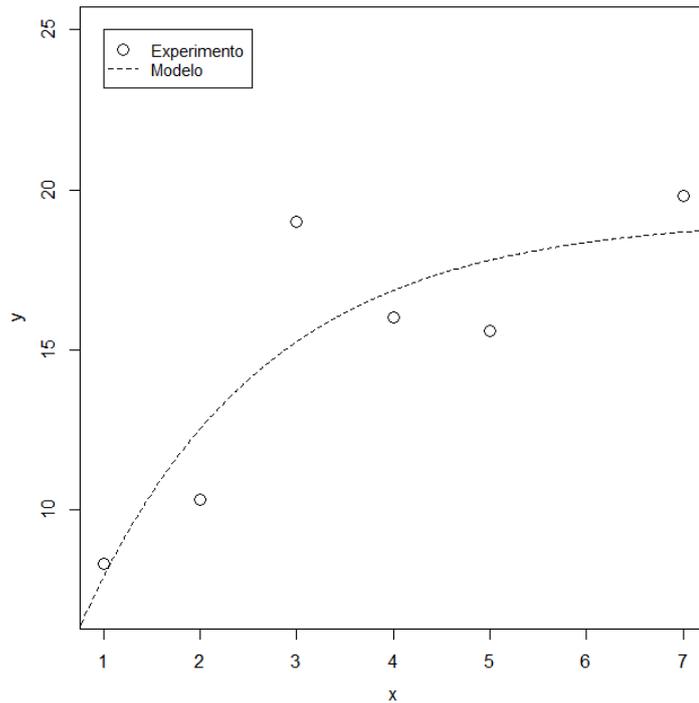


Figura 9: Comparação entre os valores de saída experimentais e os preditos para o Modelo 3.

O valor da matriz Hessiana completa, utilizando a Equação (2.14) é definido pela Equação (4.8).

$$V_{\theta}^{comp} = \begin{pmatrix} 6,30 & -0,439 \\ -0,439 & 4,19 \cdot 10^{-2} \end{pmatrix} \quad (4.8)$$

Pode ser observado que os valores das matrizes de covariâncias, calculada pelas duas abordagens de cálculo da matriz Hessiana, definidas pelas Equações (4.7) e (4.8) são bastante próximos. Utilizando-se estes valores para a geração da região de confiança elíptica dos parâmetros obtêm-se a Figura 10. Para este modelo, o emprego da matriz Hessiana da função objetivo completa não apresentou diferença significativa da aproximação de Gauss-Newton, visto que os valores das matrizes de covariâncias, Equações (4.7) e (4.8), são próximos e as elipses resultantes estão sobrepostas.

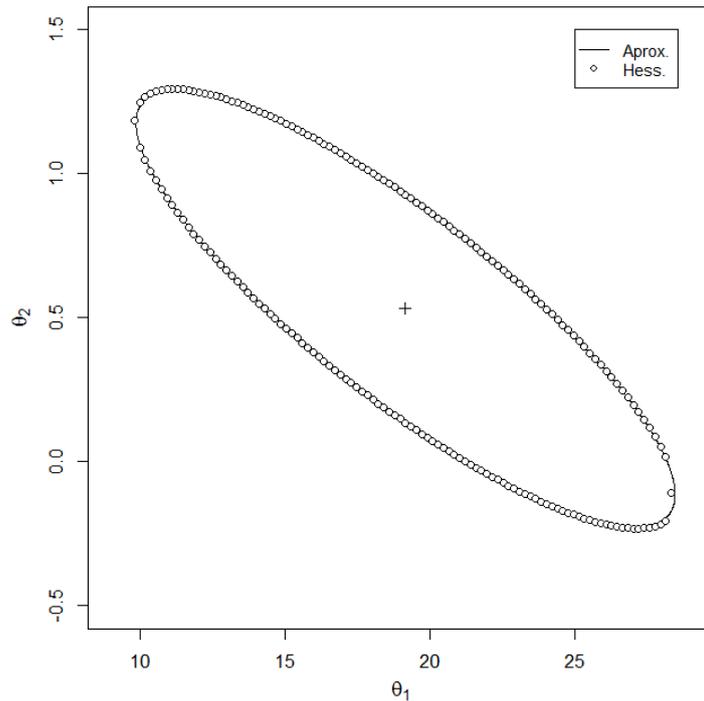


Figura 10: Regiões de confiança do Modelo 3 com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia).

Em problemas que envolvem a minimização da função objetivo, como os de estimação de parâmetros, e nos que propõem a delimitação da região de confiança do modelo é comum o emprego da aproximação de Gauss-Newton para simplificar e facilitar os resultados. Como foi observado nos modelos das Equações (3.1), (3.2) estudados, o emprego da matriz Hessiana completa difere minimamente da região gerada pela aproximação. Nos Modelos 2.A e 2.B a região de confiança obtida pela matriz Hessiana completa é maior do que a gerada pela aproximação, contudo, esta diferença é observada na área com maior excentricidade da região e o seu ganho não acrescenta muito com à região. Os resultados para o Modelo 3 diferem minimamente, tanto que as elipses geradas se sobrepõem em quase todos os pontos. Contudo, como será visto na próxima seção, as regiões de confiança de modelos não-lineares possuem o contorno completamente diferente da região elíptica, ou seja, mesmo que esta discussão aborde qual das duas elipses representa melhor a região ou se os contornos são próximos ou semelhantes, sua representação da real região de confiança está errada.

#### 4.2. Métodos para determinação de regiões de confiança não elípticas

A segunda parte deste capítulo consiste no estudo de alguns métodos para gerar as regiões de confiança não elípticas para as estimativas dos parâmetros, em particular para modelos não-lineares.

Como mencionado no Capítulo 2, as regiões de confiança são o conceito expandido dos intervalos de confiança; a diferença entre as duas abordagens é estabelecida pela existência de mais de um parâmetro no problema. A presença de dois ou mais parâmetros resulta na (provável) existência de correlações entre os valores, as quais passam a influenciar a geometria da região. Em modelos lineares, essa influência é bem nítida, podendo ser observada na inclinação do elipsoide gerado. Modelos com parâmetros altamente correlacionados apresentam elipses com contornos mais estreitos. Já os modelos com correlações pequenas entre os parâmetros possuem regiões mais amplas, por isso, suas elipses se aproximam ao traçado de circunferência.

Por sua vez, o comportamento das regiões de confiança de modelos não-lineares não é tão previsível. Esses modelos têm regiões mais difíceis de delimitar, visto que a solução analítica apresentada na Equação (2.14) não é a solução exata e, por isso, não representa corretamente as regiões desses modelos. Assim, é importante que haja uma maneira de definir as regiões destes modelos com mais exatidão, pois, como já visto, não existe uma solução analítica para esse problema. Tal situação é resolvida com o uso de métodos estocásticos, que, por utilizarem abordagens numéricas, são capazes de delimitar as regiões.

Existem algumas abordagens para definir um método estocástico que possa gerar as regiões de confiança, sendo que as vantagens e desvantagens de alguns foram abordadas no Capítulo 2. O foco deste trabalho foi o estudo das regiões de confiança resultantes do uso de três diferentes metodologias: o método do Contorno, o método do Perfil e o método *Bootstrap*.

Os métodos propostos neste trabalho tendem a fugir da ideia principal dos algoritmos heurísticos, na qual a partir de um conjunto de pontos gerados aleatoriamente, a cada iteração, a função objetivo se aproxima do valor mínimo e os pontos gerados são utilizados na construção da região. A razão para evitar esse mecanismo é a intensa quantidade de avaliações da função objetivo realizada por eles, que resulta em gastos computacionais significativos para problemas com

elevadas dimensões. Além disso, muitas vezes o contorno da região de confiança não-elíptica não fica bem definido pela nuvem de pontos que fazem parte da região de confiança.

Esta seção está estruturada de acordo com o número de parâmetros dos modelos, primeiramente serão abordados os métodos para delimitar as regiões de confiança de modelos com dois parâmetros, onde são apresentados quatro modelos; e, após, para modelos com três parâmetros, no qual três modelos foram estudados.

#### 4.2.1. Regiões de confiança de modelos com dois parâmetros

As primeiras avaliações dos métodos para determinar as regiões de confiança dos parâmetros foram realizadas para modelos com dois parâmetros. Estes são os mais simples quanto à correlação entre os parâmetros, já que só existe um par possível para ser relacionado. Contudo, as regiões dos modelos não-lineares geralmente podem apresentar geometrias difíceis de ser representadas pelos métodos, o que compromete suas eficácias.

A melhor forma de validar se o método proposto é capaz de determinar uma região de confiança corretamente é o comparando com outro já consagrado ou com modelos cujas regiões já são conhecidas, como, por exemplo, a dos modelos lineares, que possuem solução analítica e região com geometria bem definida. Assim, os primeiros resultados apresentados nesse capítulo são para um modelo linear com dois parâmetros, ou seja, sua matriz Hessiana é nula.

O modelo linear utilizado foi o Modelo 1, apresentado no Capítulo 3 e representado pela Equação (3.1) e pelos valores da Tabela 2.

O contorno gerado pelo método é constituído por pontos distribuídos ao longo da região de confiança. É possível notar que o método do Contorno foi capaz de delimitar a região de confiança dos parâmetros do Modelo 1, visto que ela se sobrepõe razoavelmente bem à região elíptica da Equação (2.34). A Figura 11 também permite notar que a região do método do Contorno concentra mais pontos nas partes excêntricas da região, necessários para delimitar bem as regiões onde ocorre uma rápida mudança de direção.

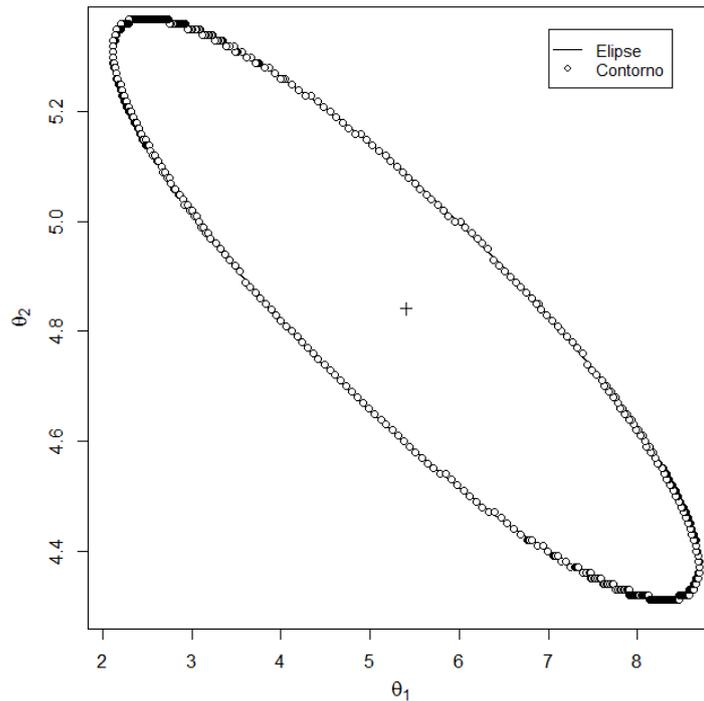


Figura 11: Comparação entre as regiões de confiança do Modelo 1 determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.1) (linha cheia).

A comparação entre as regiões geradas para o Modelo 1 permite afirmar que este método apresentou bom resultado, pois é capaz de representar uma região de confiança com resultado exato e conhecido. A próxima etapa para a validação do método do Contorno é o utilizar para descrever as regiões de modelos não-lineares, que não possuem uma solução analítica e suas curvas diferem dos traçados elípticos, podendo ser de difícil estimação, como será visto mais adiante.

O próximo método a ser testado para o Modelo 1 é o do Perfil, sendo proposto por BATES e WATTS, (1988), para descrever a linearidade de um modelo, como foi explorado na seção anterior deste capítulo. Os valores obtidos para os desvios padrão dos parâmetros  $\theta_1$  e  $\theta_2$ , de  $\tau_{MAX}$  e o valor limite da distribuição  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$  são apresentados na Tabela 11 para  $NP$  igual a 2,  $NE$  igual a 10 e  $\alpha$  igual a 0,05.

Tabela 11: Valores dos desvios padrão dos parâmetros, de  $\tau_{MAX}$  e do valor limite da distribuição  $F$  de Fisher para o Modelo 1.

$\sigma_1$	$\sigma_2$	$\tau(\text{NE} - \text{NP}, 1 - \alpha)$	$F(\text{NP}, \text{NE} - \text{NP}, 1 - \alpha)$
1,10	0,178	2,31	4,46

Os perfis do Modelo 1 são apresentados na Figura 12 (pontos) e seus valores são delimitados pela distribuição  $t$ -Student (linha tracejada). O uso do algoritmo sugerido por BATES; WATTS, (1988), para determinar a região de confiança baseando-se nos perfis resulta na elipse demonstrada na figura (linha pontilhada).

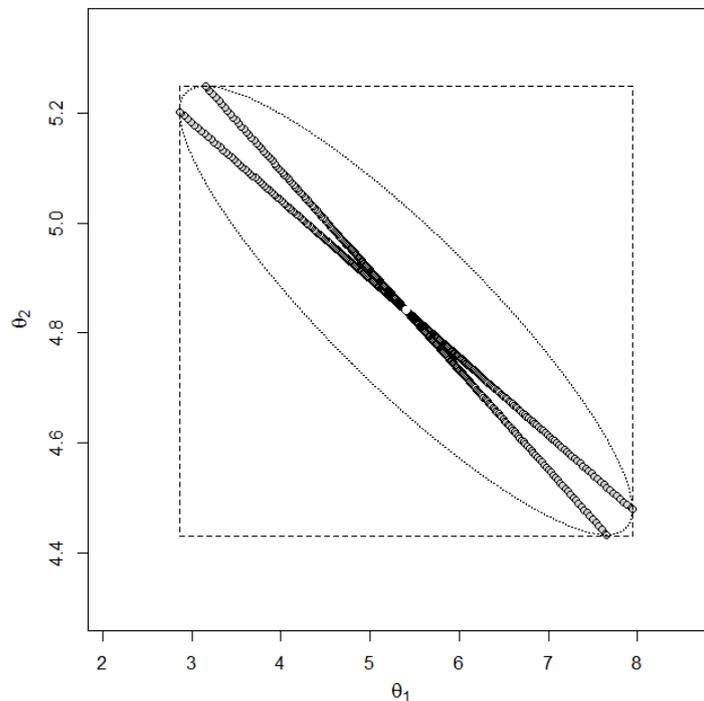


Figura 12: Representação dos perfis do Modelo 1 (pontos), dos limites dos perfis (linha tracejada) e da elipse gerada pelo método (linha pontilhada).

A comparação da região de confiança delimitada com o método do Perfil e pela Equação (2.34) é demonstrada na Figura 13. A diferença observada entre elipse analítica e a gerada pelo método do Perfil na figura pode ser explicada pelos valores apresentados na Tabela 11, visto que o perfil impõe como restrição os valores da distribuição  $t$ -Student, enquanto a região elíptica é limitada pela distribuição  $F$  de Fisher, tornando a região do Perfil menor do que a solução analítica. Contudo, também é possível reparar que o formato e a inclinação da região de confiança são similares

para os dois métodos testados, ou seja, o método é válido para representar o comportamento das regiões.

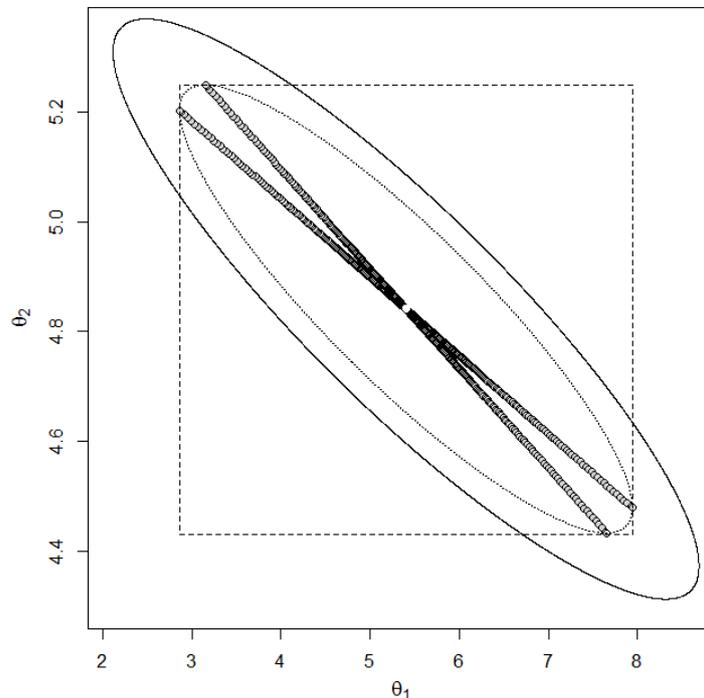


Figura 13: Comparação entre as regiões de confiança do Modelo 1 determinadas pelo método do Perfil (linha pontilhada) e pela Equação (2.34) com os valores da matriz da Equação (4.1) (linha cheia).

O terceiro método para delimitar as regiões de confiança dos parâmetros a ser explorado neste trabalho é o *Bootstrap*. Explicado nos Capítulos 2 e 3, este é um método de reamostragem, o qual evita a necessidade de realizar diversos experimentos para obter uma amostra de tamanho significativo. Como a perturbação utilizada nos parâmetros considera a variância dos dados experimentais, é possível utilizar, para fins de comparação, a função dos mínimos quadrados ponderados.

A Figura 14 exemplifica o uso do método *Bootstrap* para gerar as regiões de confiança. Para este modelo, a variância dos dados experimentais é igual a 4,0 e o valor obtido para a função objetivo utilizando a Equação (2.9) é igual 5,19. É possível observar que os pontos gerados pelo método seguem as características padrões das regiões de modelos lineares: os pontos gerados estão centrados ao redor do valor ótimo dos parâmetros e o formato é elíptico. Quando se compara a região *Bootstrap* com a da Equação (2.35), é possível observar algumas características: devido ao comportamento aleatório do método, têm-se duas situações, a primeira corresponde aos pontos cujos dados experimentais foram perturbados de maneira favorável, os

quais estão majoritariamente concentrados dentro dos limites da região elíptica; também há pontos cujas perturbações não contribuíram adequadamente para delimitar a região de confiança, sendo estes os que extrapolam seu contorno. Embora a parcela de pontos que se encontram além dos limites da região não representa adequadamente o espaço de confiança dos parâmetros, pode-se observar que o número de pontos que o representam corretamente, situados no interior da região, é significativamente maior.

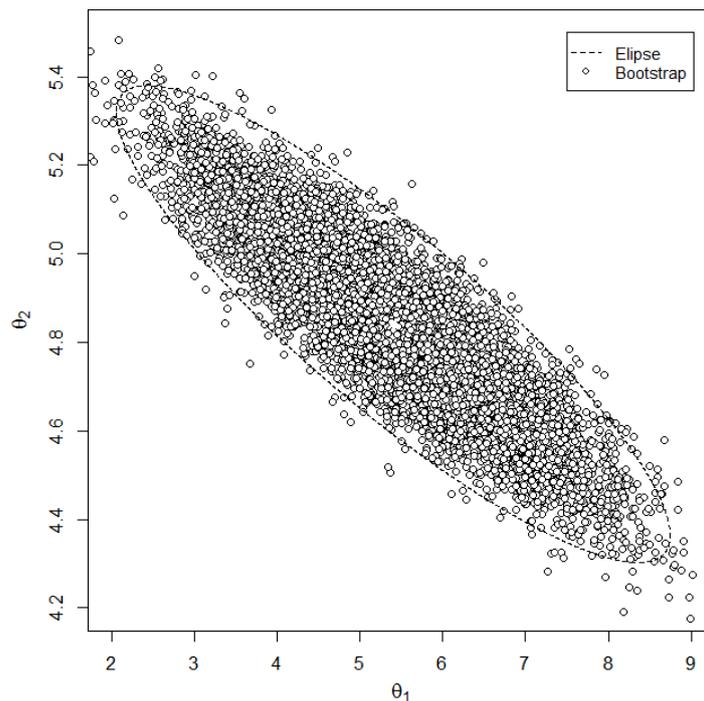


Figura 14: Comparação entre as regiões de confiança do Modelo 1 determinadas pelo método Bootstrap (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.1) (linha tracejada).

Após observar como os métodos descrevem uma região de confiança conhecida, é a vez de utilizá-los para descrever situações mais complexas, como é o caso das regiões de modelos não-lineares. Não existe um padrão para explicar o comportamento dessas regiões, sendo possível encontrar todo tipo de formato na literatura. Contudo, nesse trabalho, foram comparados dois modelos cujas regiões possuem geometrias distintas: uma região fechada, que não difere muito do contorno elíptico; e uma região aberta, em que seus limites se estendem ao infinito.

O primeiro modelo não-linear a ser trabalhado nessa seção é o Modelo 2.A, apresentado no Capítulo 3, sendo descrito pela Equação (3.2) e pelos valores da Tabela 3.

A Figura 15 permite comparar as regiões geradas pelo método do Contorno e pela elipse representada pela Equação (2.34) e com os valores da matriz de covariâncias da Equação (4.3). A primeira observação que essa figura permite inferir é a diferença entre as duas regiões, na qual a região do Contorno difere em tamanho e inclinação da elipse analítica. A razão para tal é explicada pela não-linearidade do modelo, ou seja, já que o modelo é não-linear, a Equação (2.34) não representa a solução exata da região do modelo e é condicionada pelo desvio entre os valores preditos e os obtidos experimentalmente.

A região de confiança do método do Contorno para este modelo se comporta de forma análoga a do Modelo 1, onde as áreas mais excêntricas concentram mais pontos para descrever a mudança de sentido da região. Outra observação importante que a comparação entre as regiões permite é o fato de que o Modelo 2.A não possui desvio acentuando da linearidade, pois a região mantém o traçado elíptico, o qual é a característica principal deste tipo de modelo.

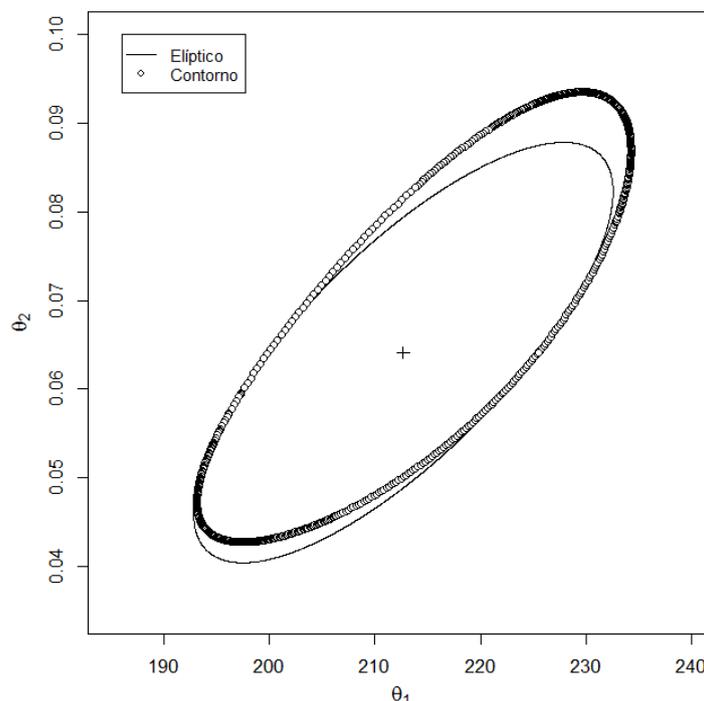


Figura 15: Comparação entre as regiões de confiança do Modelo 2.A determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.3) (linha cheia).

A escolha do Modelo 2.A para ser alvo do estudo se deu por alguns motivos, o primeiro é a característica da sua região de confiança que, como visto na curva gerada pelo método do Contorno, não difere muito dos casos lineares, mantendo o

contorno elíptico e, por isso, não se caracteriza como um modelo cuja região é de difícil estimação. O segundo motivo associado à sua escolha reside na ideia de que este é o modelo utilizado para descrever o método do perfil por BATES e WATTS, (1988). O Capítulo 6 da obra destes autores apresenta os valores intermediários do modelo utilizados para a construção da região de confiança, sendo esta uma escolha óbvia para validar o método.

Os valores dos desvios padrão dos parâmetros  $\theta_1$  e  $\theta_2$ , de  $\tau_{MAX}$  e o valor limite da distribuição  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$  são apresentados na Tabela 12 para  $NP$  igual a 2,  $NE$  igual a 12 e  $\alpha$  igual a 0,05.

Tabela 12: Valores dos desvios padrão dos parâmetros, de  $\tau_{MAX}$  e do valor limite da distribuição  $F$  de Fisher para o Modelo 2.A.

$\sigma_1$	$\sigma_2$	$\tau(\text{NE} - \text{NP}, 1 - \alpha)$	$F(\text{NP}, \text{NE} - \text{NP}, 1 - \alpha)$
6,95	0,0083	2,23	4,10

A Figura 16 permite comparar a região de confiança gerada pelo método do Perfil com a região gerada pelo método do Contorno, pois este representa melhor a região do que a elipse gerada pela Equação (2.34). É possível observar, assim como ocorrido com o Modelo 1, que o tamanho das regiões apresentadas difere; isso se deve ao fato de que a região do Perfil é menor do que a do Contorno, pois seu traçado é restrito pelos limites estabelecidos pelo valor da distribuição  $t$ -Student, apresentado na Tabela 12. Também é interessante mencionar que, embora haja diferença quanto ao tamanho, as principais características da região, seu formato e sua inclinação, foram bem reproduzidas pelo método.

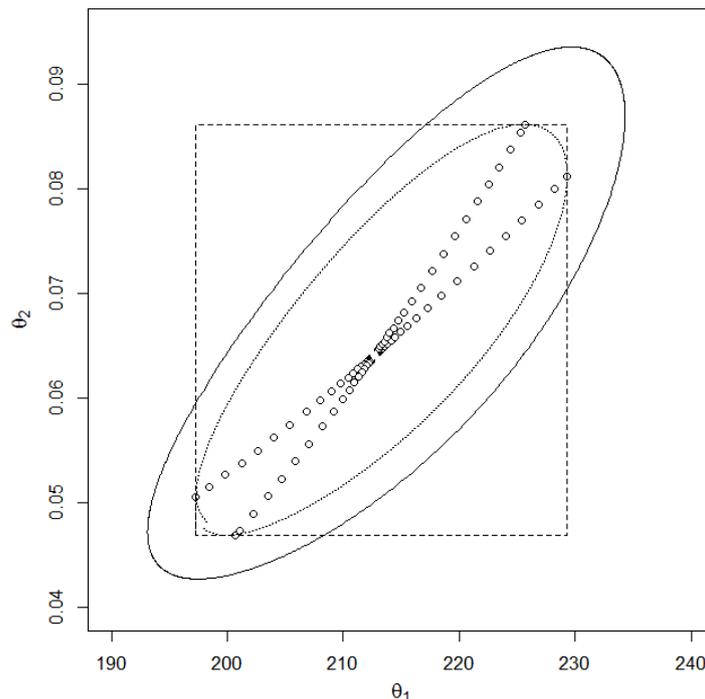


Figura 16: Comparação entre as regiões de confiança do Modelo 2.A determinadas pelo método do Perfil (linha pontilhada) e pelo método do Contorno (linha cheia).

Outra interessante comparação que pode ser feita com este modelo e método é a qualidade com que o método do Perfil reproduz a região. Como foi explicado no Capítulo 2 deste trabalho, a obra de BATES e WATTS, (1988), possui uma ótima

explicação para a geração dos perfis, utilizados no Capítulo 3, e para os traços, os quais relacionam os perfis dos dois parâmetros. Contudo, a explicação dos esboços, a região de confiança, é um tanto vaga em algumas de suas etapas, portanto a comparação dos resultados obtidos com os fornecidos pelos autores indica se a implementação do método foi adequada. Para o Modelo 2.A em especial, os autores apresentam os resultados de etapas intermediárias, as quais são utilizadas para a construção da região. Assim, é possível comparar as duas regiões de confiança na Figura 17.

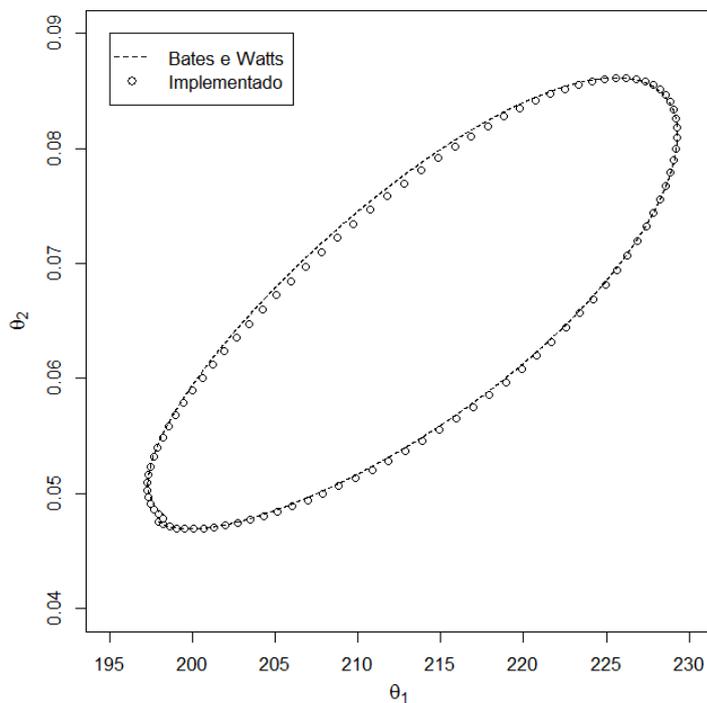


Figura 17: Comparação entre as regiões de confiança geradas pelo método do Perfil de acordo com a implementação deste trabalho (linha tracejada) e com a disponível na literatura (pontos).

A curva gerada com os pontos, apresentada na Figura 17, delimita a região de confiança dos parâmetros do Modelo 2.A obtida com os valores disponíveis na literatura. Percebe-se que a principal diferença entre as regiões obtidas acontece em sua parte menos excêntrica, na qual os valores calculados do modelo tendem a terem seu módulo mais elevado na coordenada do parâmetro  $\theta_2$ , contudo esse desvio entre os valores não afeta a qualidade da região, visto que, visualmente, essa diferença é ínfima e não compromete a sua escala e o seu formato.

A região de confiança determinada pelo método *Bootstrap* para o Modelo 2.A é apresentada e comparada com a gerada pelo método do Contorno na Figura 18. A

escolha da região do Contorno se deve ao fato de que esta representação é mais exata do que a obtida com a Equação (2.34). A variância dos dados experimentais não é disponibilizada, assim o seu valor foi estimado pela aproximação definida na Equação (3.9) sendo igual a 119,5. O valor da função objetivo dos mínimos quadrados ponderados é igual a 10,00.

Para que seja possível estabelecer uma comparação adequada para a região de confiança obtida com o método *Bootstrap*, foi gerada a região do Contorno para os modelos não-lineares, assumindo que as variâncias de cada um são conhecidas. Dessa forma, é possível determinar a região de forma análoga a da Equação (2.26), a qual resulta em um traçado menor do que o usual, visto que os valores das incertezas experimentais são conhecidos e, assim, é possível atribuir à equação a distribuição *qui-quadrado*.

Assim como observado para o Modelo 1, é possível reparar que os pontos gerados pelo método *Bootstrap* estão bem distribuídos ao redor do ponto de mínimo, sendo que estão dispostos no mesmo formato da região do Contorno, com alguns extrapolando os seus limites. Os pontos cujos dados experimentais foram perturbados de maneira favorável estão concentrados dentro dos limites da região elíptica e os pontos cujas perturbações não contribuíram adequadamente para delimitar a região de confiança extrapolam seu contorno. Novamente, é importante observar que o número de pontos que se encontram além dos limites da região é significativamente menor do que os estão dentro dos limites da região de confiança.

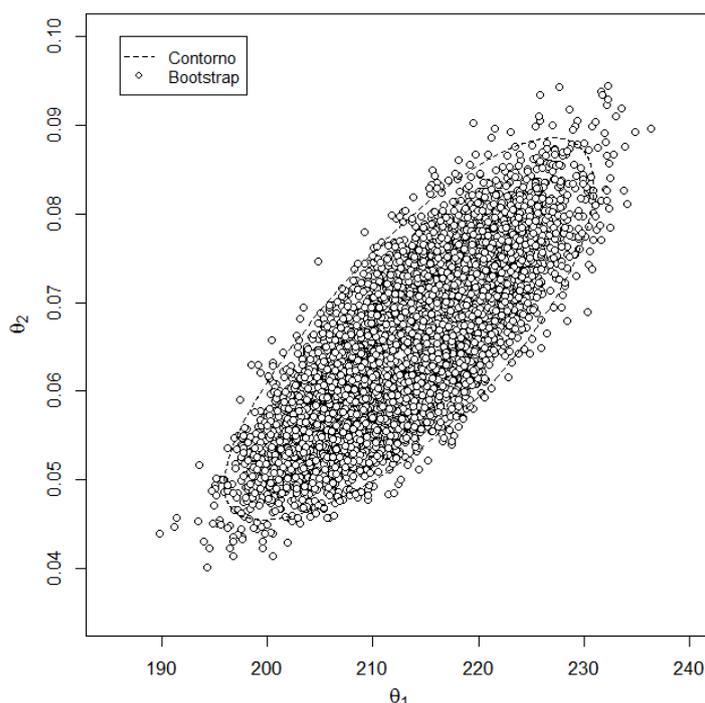


Figura 18: Comparação entre as regiões de confiança do Modelo 2.A determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada).

O próximo modelo não-linear a ser estudado é o Modelo 2.B, apresentado no Capítulo 3. Sua função é definida pela mesma equação do modelo anterior, Equação (3.2). A escolha deste modelo é devida a não-linearidade mais acentuada do que a do Modelo 2.A, como foi possível observar na comparação dos perfis dos parâmetros realizada no capítulo anterior.

A comparação da região de confiança gerada pelo método do Contorno com a gerada pela Equação (2.34) com matriz de covariâncias da Equação (4.5) é demonstrada na Figura 19. Assim como para o Método 2.A, a região elíptica fornece um traçado distorcido do real comportamento da região de confiança do modelo devido à falta de exatidão da Equação (2.34) ao ser utilizada no cálculo de regiões não-lineares. Quanto à forma, é nítida a semelhança entre as regiões dos Modelos 2.A e 2.B e do comportamento do método, que dispõe de mais pontos em regiões onde há a mudança de sentido da curva.

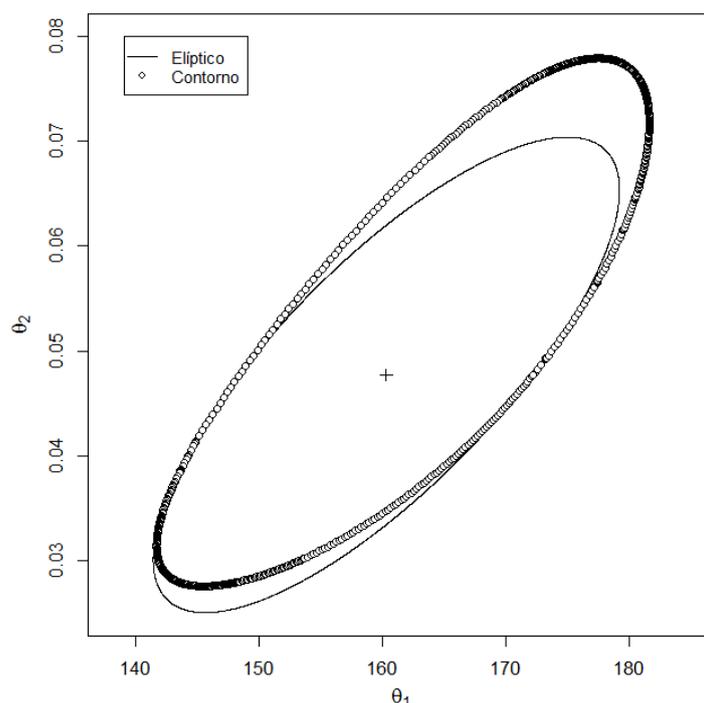


Figura 19: Comparação entre as regiões de confiança do Modelo 2.B determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.5) (linha cheia).

O uso do Modelo 2.B para determinar a região de confiança dos parâmetros também se deu pela sua similaridade com o Modelo 2.A, padrão do método do Perfil. Embora seja mais não-linear, ambas as regiões são parecidas e, por isso, não deve haver mudança brusca na qualidade de ajuste do método. Os valores obtidos para os desvios padrão dos parâmetros  $\theta_1$  e  $\theta_2$ ,  $\tau_{MAX}$  e o valor limite da distribuição  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$  são apresentados na Tabela 13 para  $NP$  igual a 2,  $NE$  igual a 11 e  $\alpha$  igual a 0,05.

Tabela 13: Valores dos desvios padrão dos parâmetros, de  $\tau_{MAX}$  e do valor limite da distribuição  $F$  de Fisher para o Modelo 2.B.

$\sigma_1$	$\sigma_2$	$\tau(NE - NP, 1 - \alpha)$	$F(NP, NE - NP, 1 - \alpha)$
6,48	0,0078	2,26	4,26

De fato, a região de confiança gerada pelo método do Perfil para este modelo é destacada na Figura 20, e é possível observar características idênticas à da Figura 16. A região gerada pelo método do Perfil é menor do que a região de confiança obtida pelo método do Contorno, visto que seu traçado é limitado pelo valor da distribuição

*t-Student*, enquanto que o método do Contorno, assim como a solução analítica é limitado pela distribuição *F* de Fisher. Outra importante observação é que, assim como para os modelos já apresentados, o método do Perfil consegue descrever o formato e a inclinação da região.

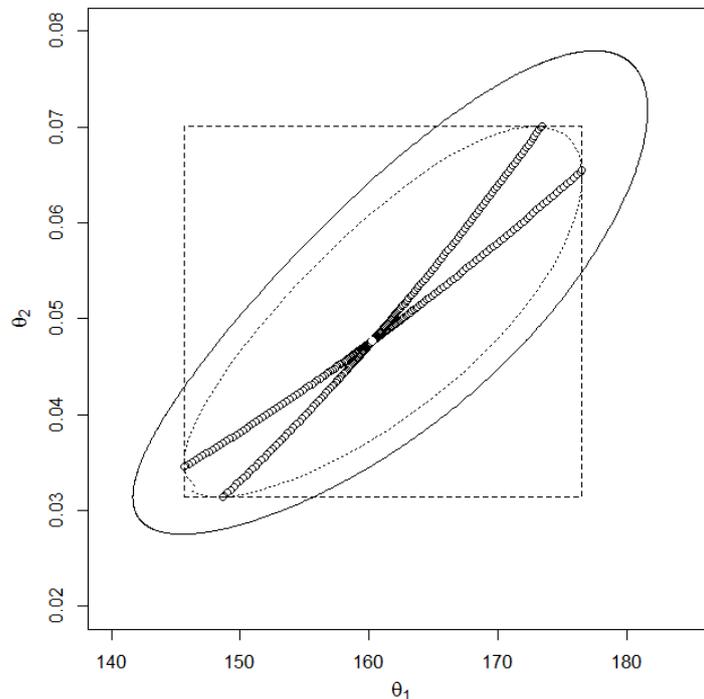


Figura 20: Comparação entre as regiões de confiança do Modelo 2.B determinadas pelo método do Perfil (linha pontilhada) e pelo método do Contorno (linha cheia).

O uso do método *Bootstrap* para delimitar a região de confiança do Modelo 2.B, resulta em uma situação análoga à dos Modelo 1 e 2.A. O valor da variância experimental utilizado, obtido com a Equação (3.9), é igual a 99,5, obtendo-se o valor igual a 8,64 para a função objetivo dos mínimos quadrados ponderados. Na Figura 21, percebe-se que a maioria dos pontos se concentram dentro do formato da região do Contorno, sendo possível inferir o seu formato. Alguns desses pontos a extrapolam, devido à aleatoriedade das perturbações utilizadas para o cálculo dos novos valores “ótimos” dos a parâmetros.

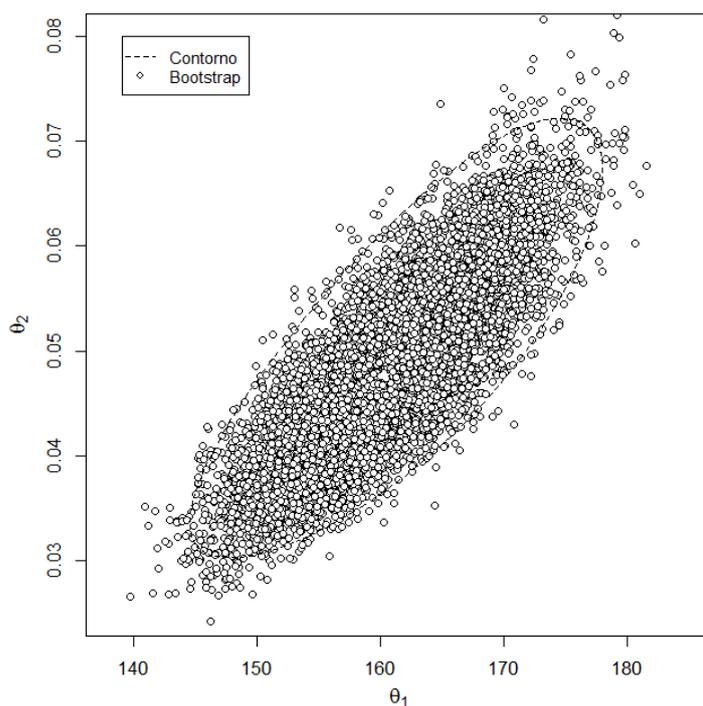


Figura 21: Comparação entre as regiões de confiança do Modelo 2.B determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada).

O último modelo não-linear com dois parâmetros utilizado foi o Modelo 3, apresentado no Capítulo 3, representado pela função da Equação (3.3) e pelos dados da Tabela 5.

A escolha deste modelo se deve à sua comparação com os demais apresentados. Observando os desvios apresentados pela Figura 9, pode-se notar que estes são mais elevados dos que os dos outros modelos não lineares, Figura 3 e Figura 5. Essa diferença entre os valores experimentais e os calculados pelo modelo está relacionada com a não-linearidade do Modelo 3, resultando em uma região de confiança atípica das demais vistas anteriormente.

A Figura 22 compara as regiões de confiança geradas pelo método do Contorno e pela Equação (2.34), sendo que seus formatos diferem completamente. A região elíptica resulta em uma representação errada da região, pois seu contorno real não segue o mesmo formato nem sua característica principal, as áreas abertas. É possível observar que o contorno traçado pelo método é capaz de delimitar a região, contudo é necessário impor uma restrição ao avanço ao longo da curva, pois as áreas abertas impedem retorno ao ponto inicial. Também é possível perceber o

comportamento idêntico aos demais casos em que foram utilizados o método do Contorno, nos lugares onde há mudança de sentido, há a concentração de pontos.

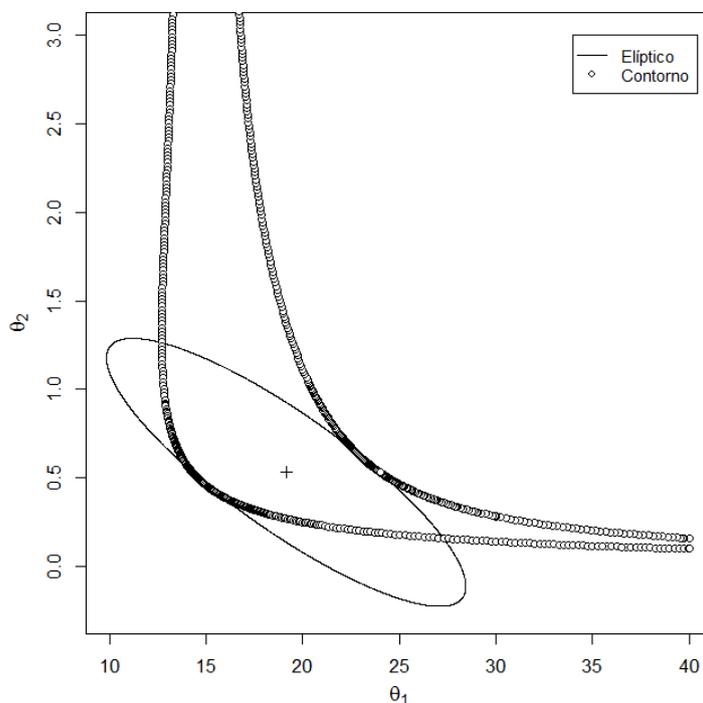


Figura 22: Comparação entre as regiões de confiança do Modelo 3 determinadas pelo método do Contorno (pontos) e pela Equação (2.34) com os valores da matriz da Equação (4.7) (linha cheia).

Devido à característica do método do Perfil em traçar a região de confiança a partir dos pontos limites da distribuição *t-Student*, não foi possível gerar essa região para o Modelo 3, já que seria necessário estipular algum valor limite para interromper o cálculo do algoritmo e, assim, só seria possível representar uma parcela da região.

A variância dos dados experimentais do modelo, determinada com a Equação (3.9), é igual a 6,5. O valor da função objetivo obtida pela equação dos mínimos quadrados ponderados é igual a 3,99. A Figura 23 permite observar que os pontos gerados pelo *Bootstrap*, os quais representam a região da razão de verossimilhança, se expandem além da curva gerada pelo método do Contorno. Também é possível verificar que houve outros comportamentos análogos ao das demais regiões geradas com o método *Bootstrap*, como a concentração da maioria dos pontos em uma área específica da curva e a representação do formato da região de confiança.

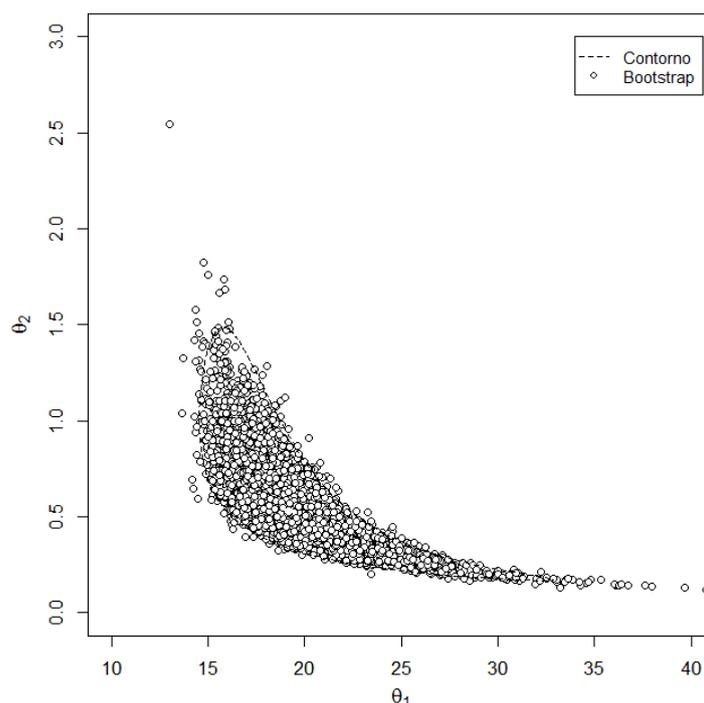


Figura 23: Comparação entre as regiões de confiança do Modelo 3 determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada).

Para os modelos com dois parâmetros é possível fazer algumas observações com as regiões de confiança obtidas para os Modelos 1, 2.A, 2.B e 3. A qualidade da região obtida com o método do Contorno foi similar para todos os modelos, visto que ele foi capaz de descrever as curvas das regiões de confiança. O principal comportamento que se pode destacar com o uso de método se dá pela necessidade de diminuir o valor do passo do algoritmo quando os pontos utilizados na construção da região estão mudando de direção. Uma desvantagem de técnicas do tipo contorno é o seu uso em regiões com as características do Modelo 3, o qual possui extremidade aberta. Para este tipo de situação, é necessário conhecer previamente o comportamento da região, para definir limites em seu avanço ou inserir medidas que garantam seja o contorno de determinada parte da região seja explorado, como foi feito neste trabalho.

A desvantagem principal das regiões de confiança obtidas utilizando o método do Perfil é que elas não representam as obtidas com a Equação (2.34), para modelos lineares, e diferem dos demais casos por não utilizar como limite o valor da distribuição  $F$  de Fisher e sim o valor da  $t$ -Student. Outra desvantagem é que o método não pode ser utilizado para o Modelo 3, devido à característica de sua região. Contudo, as

curvas obtidas com o método do Perfil conseguem reproduzir a geometria e a inclinação da região de confiança e, como maior vantagem, os perfis gerados servem como indicativo para a linearidade do modelo, o que não é observado para os demais métodos estudados neste trabalho.

As regiões de confiança do método *Bootstrap* obtidas apresentam o mesmo formato observado quando comparadas com as do método do Contorno. A distribuição dos pontos no espaço das regiões de confiança acontece, majoritariamente, nas proximidades do ponto ótimo, enquanto uma menor quantidade de pontos se espalha na região. Também é possível observar que existem pontos na região externa no contorno devido ao caráter aleatório deste método.

#### 4.2.2. Regiões de confiança de modelos a três parâmetros

Na seção acima, foi discutida a qualidade da representação da região de confiança dos parâmetros utilizando os modelos do Contorno, Perfil e *Bootstrap* para modelos lineares e não-lineares com dois parâmetros, assim como as suas vantagens e desvantagens. O foco desta seção será verificar se os métodos apresentados são capazes de descrever as regiões de confiança de modelos com mais de dois parâmetros.

A escolha de modelos com três parâmetros para estudo pode ser explicada pelo acréscimo de um novo parâmetro ao modelo. Um modelo com dois parâmetros possuirá apenas a relação entre  $\theta_1$  e  $\theta_2$ . Já um modelo com três parâmetros possuirá as relações entre  $\theta_1$  e  $\theta_2$ ,  $\theta_1$  e  $\theta_3$ ,  $\theta_2$  e  $\theta_3$ , expandindo as dimensões do problema. À medida que um modelo apresenta mais parâmetros, as relações entre os parâmetros aumentam, assim como a complexidade do problema.

Outro motivo para a escolha deste tipo de modelo se deve às dimensões da região de confiança obtida. Quando se trabalha com dois parâmetros, haverá uma área para descrever o espaço onde os parâmetros estão confinados; ao adicionar mais um parâmetro, a mesma é descrita por um espaço com três dimensões. Um problema com  $NP$  parâmetros resultaria em um espaço com  $NP$  dimensões, sendo impossível de representar graficamente quando este valor é maior do que três. Assim, a forma mais usual de representar as regiões de confiança de modelos com mais de dois parâmetros é a projeção de sua sombra em um plano, tornando o problema com  $NP$  dimensões em  $\frac{1}{2}NP(NP - 1)$  problemas bidimensionais.

Assim como na seção anterior, as primeiras regiões a serem estudadas serão as de um modelo linear, visto que, independentemente do número de dimensões, se o modelo possui matriz Hessiana nula em relação aos parâmetros, a região será elíptica.

Para o Modelo 4, o valor da função objetivo após a minimização é igual a 11,97 e corresponde aos valores ótimos dos parâmetros  $\hat{\theta}_1$ ,  $\hat{\theta}_2$  e  $\hat{\theta}_3$  iguais a 10,7; 2,05 e 2,71 respectivamente. As variáveis dependentes medidas e as previstas pelo modelo são apresentadas na Figura 24.

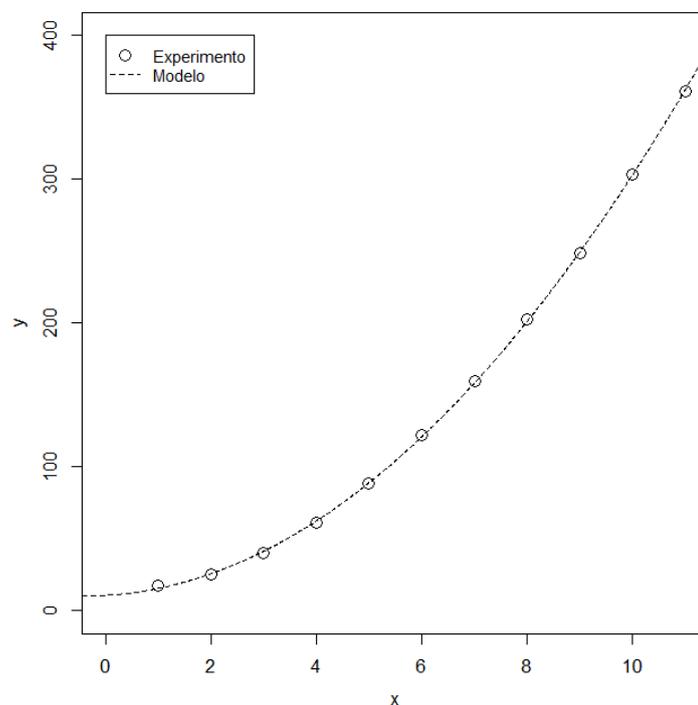


Figura 24: Comparação entre os valores de saída experimentais e os preditos para o Modelo 4.

Por se tratar de um modelo linear em relação aos parâmetros, a Equação (2.15) representa a solução analítica do problema e suas regiões de confiança possuem geometria bem definida, em formato de elipse. O esperado dessa seção é que os métodos do Contorno, Perfil e *Bootstrap* possam ser expandidos para problemas de dimensões maiores, assim, faz-se interessante os comparar como um modelo cuja região é conhecida e exata, o Modelo 4.

As correlações observadas foram iguais -0,91 para os parâmetros  $\theta_1$  e  $\theta_2$ , 0,81 para os parâmetros  $\theta_1$  e  $\theta_3$  e -0,97 para os parâmetros  $\theta_2$  e  $\theta_3$ . As covariâncias obtidas para os parâmetros são representadas na matriz definida pela Equação (4.9).

$$V_{\theta} = \begin{pmatrix} 1,81 & -0,626 & 4,53 \cdot 10^{-2} \\ -0,626 & 0,265 & -2,09 \cdot 10^{-2} \\ 4,53 \cdot 10^{-2} & -2,09 \cdot 10^{-2} & 1,74 \cdot 10^{-3} \end{pmatrix} \quad (4.9)$$

As regiões de confiança geradas pelo método do Contorno são comparadas com as elipses analíticas na Figura 25. É possível observar que contorno gerado pelo método se sobrepõe às soluções analíticas, permitindo concluir que o método consegue reproduzir corretamente as regiões de modelos com mais parâmetros. O comportamento do contorno é idêntico ao observado para o Modelo 1, Figura 13, no qual são necessários mais pontos para se descrever a parte excêntrica da região devido à mudança na direção da região.

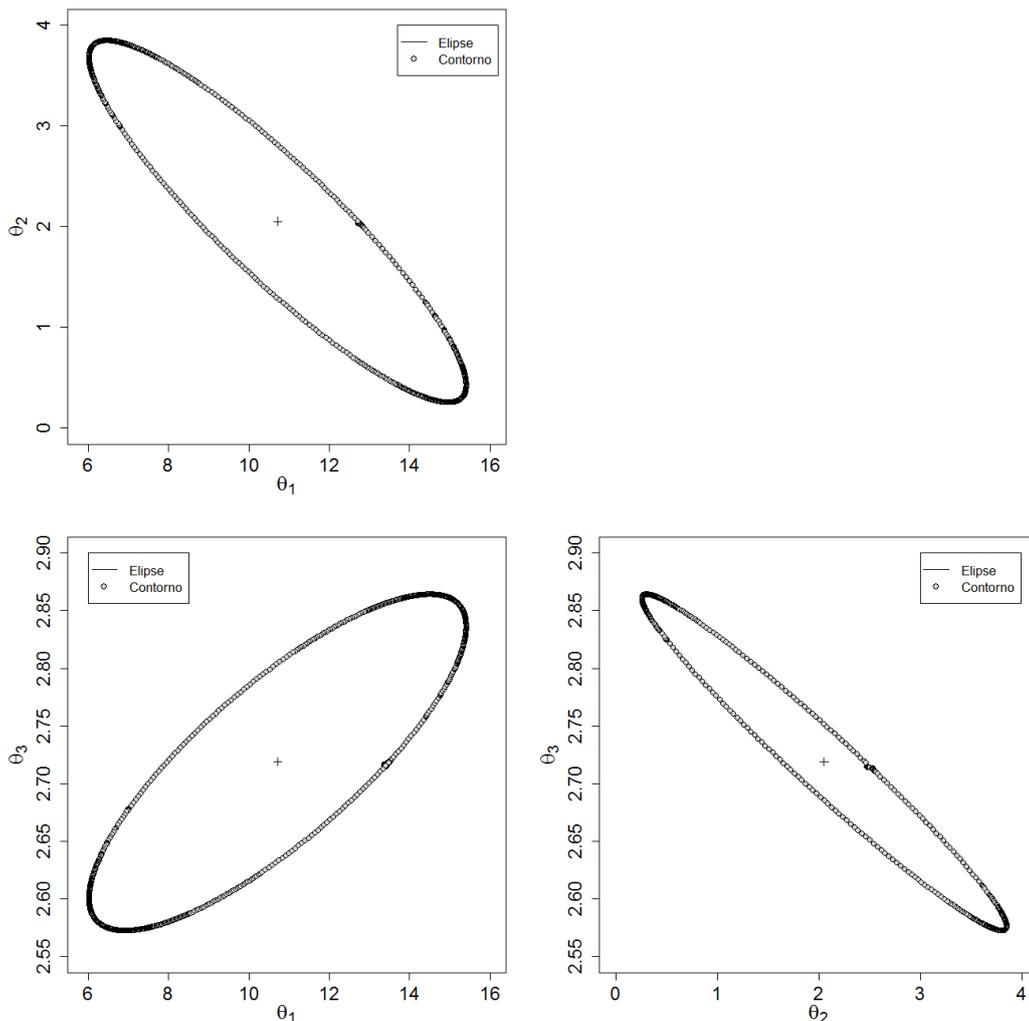


Figura 25: Comparação entre as regiões de confiança do Modelo 4 determinadas pelo método do Contorno (pontos) e pela Equação (2.34) (linha cheia).

O método do Perfil é uma técnica para descrever a não-linearidade de um modelo através de relações bidimensionais, dessa forma, supõe-se que não haverá dificuldades em representar um modelo independentemente do número de parâmetros, dependendo apenas do seu grau de não-linearidade. Os valores obtidos para os desvios padrão dos parâmetros  $\theta_1$ ,  $\theta_2$  e  $\theta_3$ ,  $\tau_{MAX}$  e o valor limite da distribuição  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual a  $1 - \alpha$  são apresentados na Tabela 14 para  $NP$  igual a 3,  $NE$  igual a 11 e  $\alpha$  igual a 0,05.

Tabela 14: Valores dos desvios padrão dos parâmetros, de  $\tau_{MAX}$  e do valor limite da distribuição  $F$  de Fisher para o Modelo 4.

$\sigma_1$	$\sigma_2$	$\sigma_3$	$\tau(NE - NP, 1 - \alpha)$	$F(NP, NE - NP, 1 - \alpha)$
1,34	0,51	0,042	2,31	4,07

Os perfis do modelo, assim como as regiões de confiança resultantes, são apresentados na Figura 26. A representação das regiões segue o comportamento da obtida para o Modelo 1, pois tanto o seu formato, quanto a sua inclinação coincidem com a solução analítica, diferindo apenas em escala. Esta observação está baseada nos limites impostos pela distribuição  $t$ -Student nos valores dos parâmetros, sendo este menor do que o da distribuição  $F$  de Fisher para os mesmos graus de liberdade e nível de confiança, conforme a Tabela 14 ilustra.

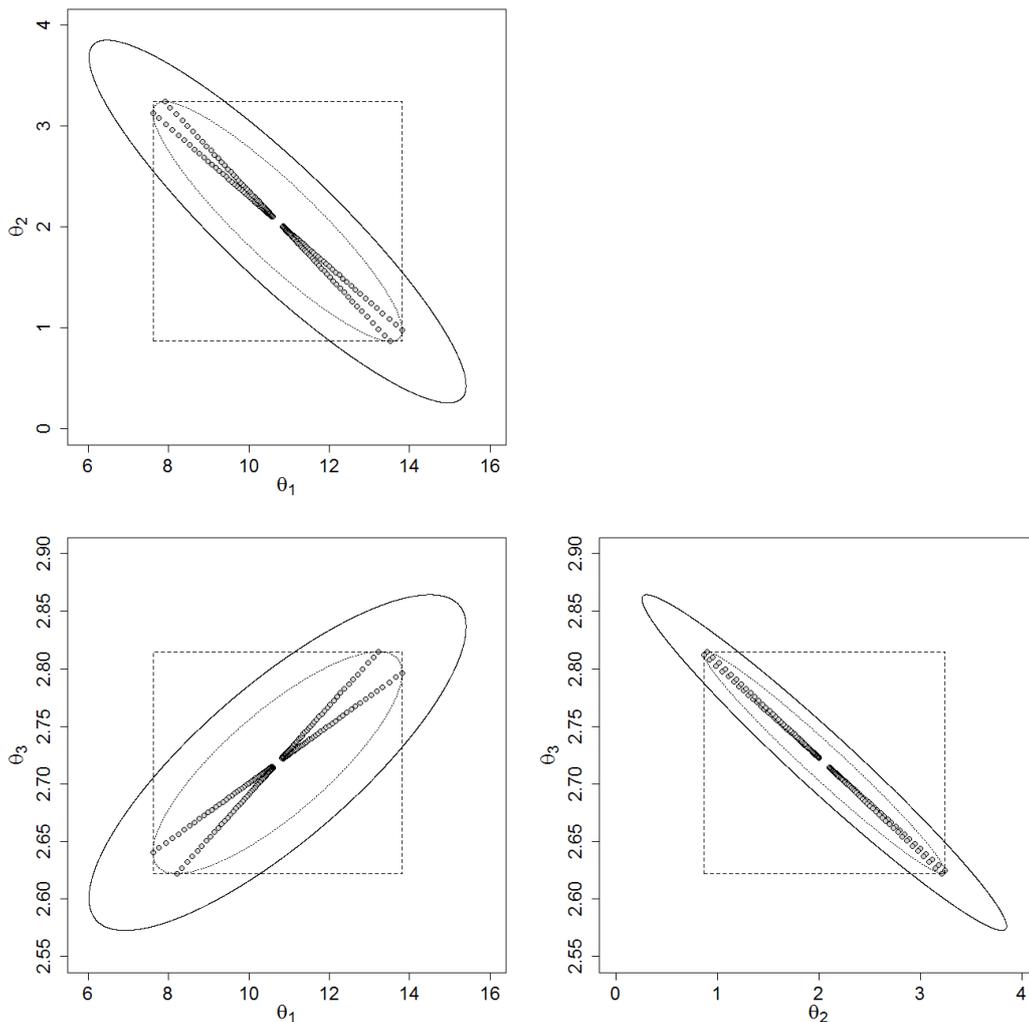


Figura 26: Comparação entre as regiões de confiança do Modelo 4 determinadas pelo método do Perfil (linha pontilhada) e pela Equação (2.34) (linha cheia).

Outra importante interpretação que a Figura 26 proporciona diz a respeito à correlação entre os parâmetros, pois quanto maior esta, maior será a excentricidade da região de confiança. A região envolvendo os parâmetros  $\theta_2$  e  $\theta_3$  da figura representa um caso de alta correlação, como demonstra o comportamento de seus perfis. É possível perceber que os mesmos estão quase sobrepostos enquanto que os demais perfis da figura são bem distintos. Uma interpretação para tal situação pode ser que, para modelos com correlações entre os parâmetros elevadas, o método do perfil possa apresentar dificuldades devido ao tamanho da região.

A região *Bootstrap* foi gerada com variância igual a 2,5 e o valor da função objetivo pelos métodos dos mínimos quadrados ponderados calculado é igual a 4,78. Como foi visto para os modelos anteriores, o método *Bootstrap* resulta em regiões similares às obtidas com o método da razão de verossimilhança; assim como foi

observado para os modelos com dois parâmetros, os parâmetros “ótimos” perturbados, em sua maioria, localizados no espaço interior da região de confiança, enquanto que alguns a extrapolam. Na Figura 27 estão representadas as regiões obtidas pelo método *Bootstrap*, podendo-se perceber que a inclinação da região, bem como o seu formato, está de acordo com os da solução analítica.

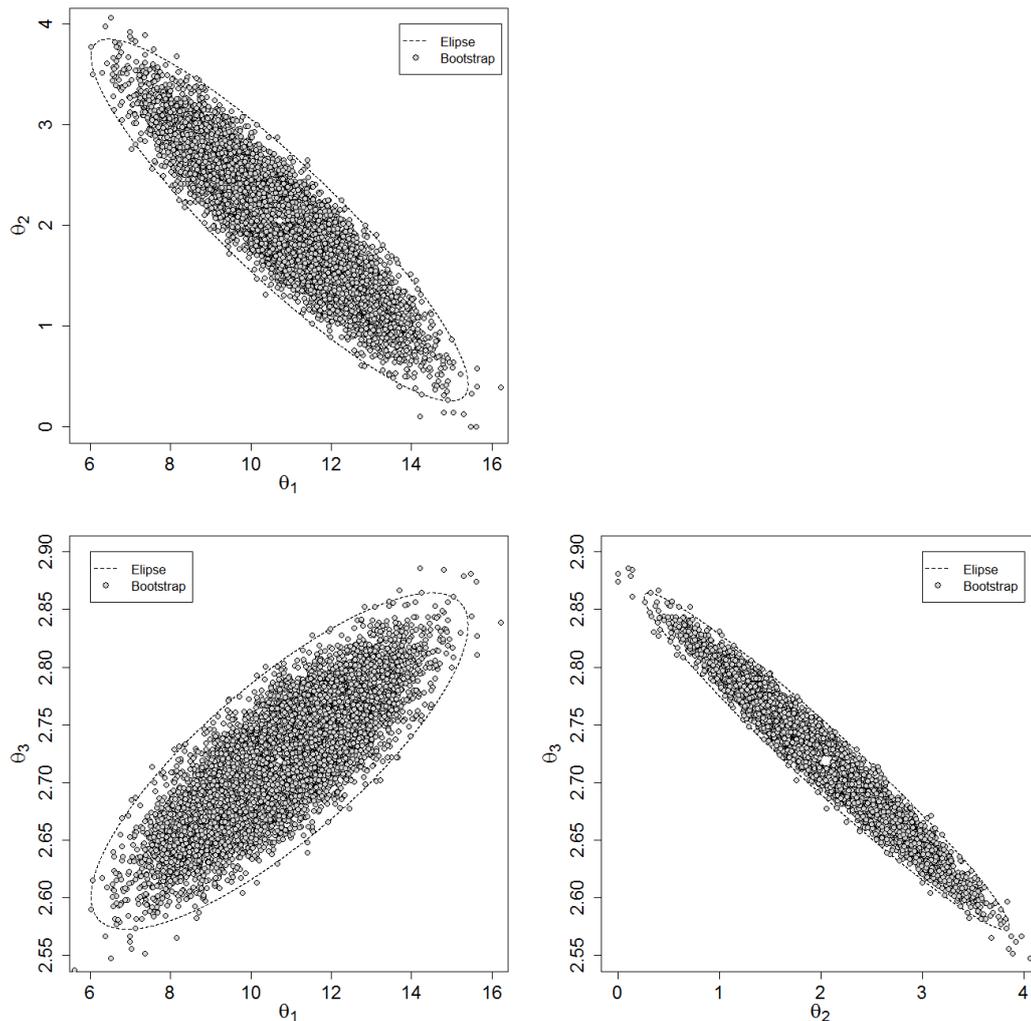


Figura 27: Comparação entre as regiões de confiança do Modelo 4 determinadas pelo método Bootstrap (pontos) e pela solução analítica (linha tracejada).

Os métodos propostos foram capazes de representar o modelo linear sem problemas e de maneira análoga à vista para os modelos bidimensionais. A próxima etapa consiste em os validar para modelos não-lineares, cujas regiões de confiança possuem comportamento mais complexo, fato este que dificulta a ação de técnicas para delimitá-las.

Para realizar o estudo das regiões de confiança dos modelos não-lineares com três parâmetros, foi utilizado o Modelo 5.A. O valor de sua função objetivo após a minimização é igual a 6,45 e corresponde aos valores ótimos dos parâmetros  $\hat{\theta}_1$ ,  $\hat{\theta}_2$  e  $\hat{\theta}_3$  iguais a 96,06; 0,1104 e 0,879 respectivamente.

A Figura 28 apresenta os valores das variáveis dependentes obtidas experimentalmente da Tabela 7 quando comparadas com as geradas pelo Modelo 5.A utilizando os valores dos parâmetros ótimos e das variáveis independentes da mesma tabela. O modelo ajustado foi capaz de representar bem os dados, posto que a curva resultante sobrepõe a maioria dos pontos e os demais se encontram distribuídos abaixo ou acima da curva, evitando comportamentos tendenciais.

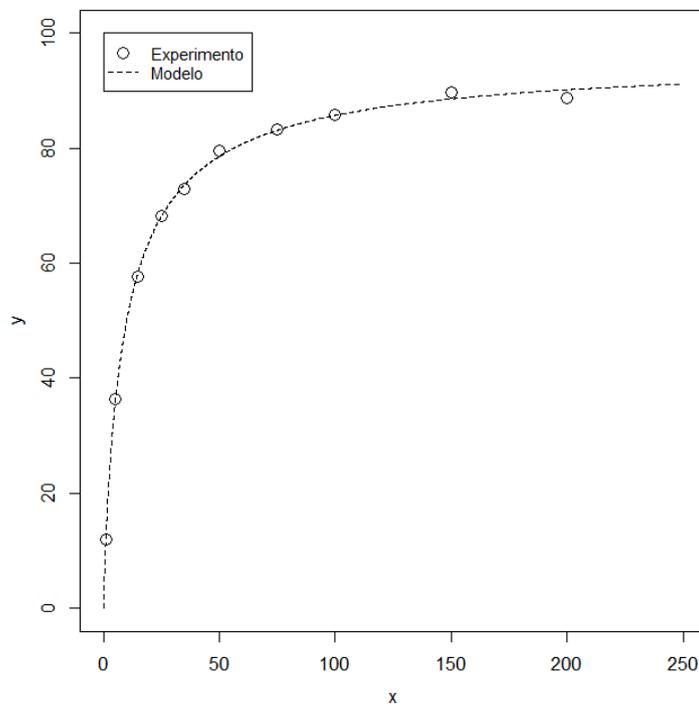


Figura 28: Comparação entre os valores de saída experimentais e os preditos para o Modelo 5.A.

As correlações observadas entre os parâmetros  $\theta_1$  e  $\theta_2$  é igual a -0,81, para os parâmetros  $\theta_1$  e  $\theta_3$  é igual a -0,80 e para parâmetros  $\theta_2$  e  $\theta_3$  é igual a 0,45. As covariâncias obtidas para os parâmetros são representadas na matriz definida pela Equação (4.10).

$$V_{\theta} = \begin{pmatrix} 1,53 & -4,34.10^{-3} & -2,99.10^{-2} \\ -4,34.10^{-3} & 1,88.10^{-5} & 5,93.10^{-5} \\ -2,99.10^{-2} & 5,93.10^{-5} & 9,01.10^{-4} \end{pmatrix} \quad (4.10)$$

As regiões de confiança do Modelo 5.A pelo método do Contorno são representadas na Figura 29. O método é comparado com as regiões elípticas descritas pelas Equações (2.14) e (2.15). Como esperado para um modelo não-linear, a elipse é uma aproximação da região, não sendo capaz de representar fielmente o seu comportamento em nenhuma das relações entre os parâmetros. O comportamento do método é similar ao já observado nos demais modelos vistos, sendo capaz de descrever a região de confiança sem problemas, apenas necessita de um número maior de pontos para delimitar a mudança de direção da região.

A análise dos gráficos apresentados nesta figura permite algumas observações: a região correspondente aos parâmetros  $\theta_1$  e  $\theta_3$  é a que mais se aproxima da elipse, pois suas duas regiões com maior excentricidade possuem formatos semelhantes e há vários pontos sobrepostos pelo Contorno e pela elipse; a região que correlaciona os parâmetros  $\theta_2$  e  $\theta_3$  é a que possui o comportamento mais não-linear, devido ao deslocamento da curva gerada pelo método do Contorno em relação à região elíptica; nesta também é possível verificar que os parâmetros  $\theta_2$  e  $\theta_3$  não são tão bem correlacionados quanto os outros, pois seu traçado tende a distorcer a inclinação da região.

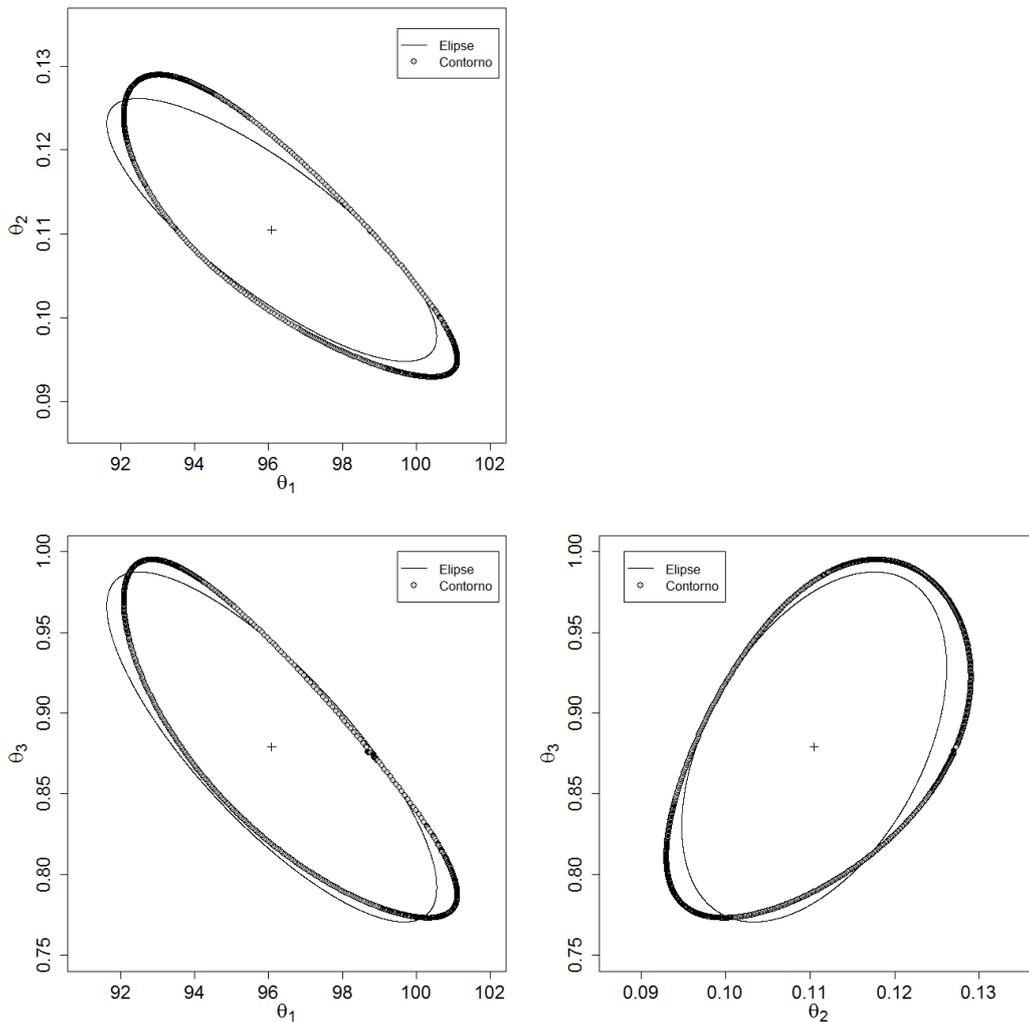


Figura 29: Comparação entre as regiões de confiança do Modelo 5.A determinadas pelo método do Contorno (pontos) e pela Equação (2.34) (linha cheia).

Os valores obtidos para os desvios padrão dos parâmetros  $\theta_1$ ,  $\theta_2$  e  $\theta_3$ ,  $\tau_{MAX}$  e o valor limite da distribuição  $F$  de Fisher com  $NP$  e  $NE - NP$  graus de liberdade e confiança igual à  $1 - \alpha$  são apresentados na Tabela 15 para  $NP$  igual a 3,  $NE$  igual a 10 e  $\alpha$  igual a 0,05.

Tabela 15: Valores dos desvios padrão dos parâmetros, de  $\tau_{MAX}$  e do valor limite da distribuição  $F$  de Fisher para o Modelo 5.A.

$\sigma_1$	$\sigma_2$	$\sigma_3$	$\tau(NE - NP, 1 - \alpha)$	$F(NP, NE - NP, 1 - \alpha)$
1,24	0,0043	0,030	2,36	4,35

As regiões geradas pelo método do perfil podem ser observadas na Figura 30 e são comparadas com as curvas obtidas com o método do Contorno. Quanto aos traços das regiões, nota-se que os mesmos representam a correlação entre os parâmetros de acordo com sua inclinação. Como abordado no parágrafo anterior, os parâmetros  $\theta_2$  e  $\theta_3$  são os menos correlacionados, visto que seus ângulos são mais próximos de  $90^\circ$  do que o dos demais. Outra observação diz a respeito às regiões obtidas com o método, as quais seguem o comportamento característico identificados nos modelos anteriores. Embora siga a mesma geometria da região de confiança gerada pelo do método do Contorno, a distância de seus pontos do centro é menor, visto que o Perfil é limitado pelo valor da distribuição *t-Student*.

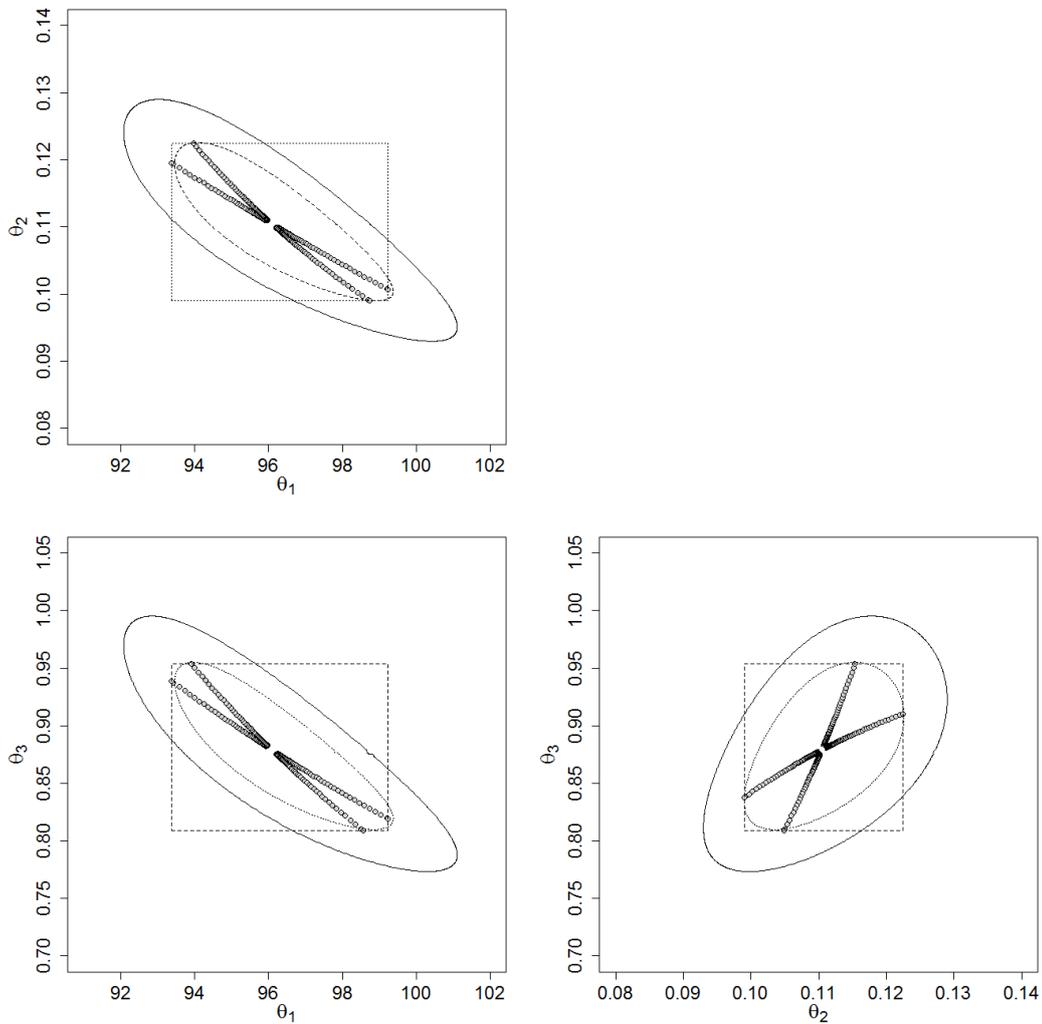


Figura 30: Comparação entre as regiões de confiança do Modelo 5.A determinadas pelo método do Perfil (linha pontilhada) e pelo método do Contorno (linha cheia).

A Figura 31 apresenta as regiões de confiança obtidas com o uso do método *Bootstrap* com valor da função objetivo pelo método dos mínimos quadrados ponderados igual a 3,23 e comparadas com as geradas pelo método do Contorno. A tendência observada nos demais modelos também está presente neste, posto que os pontos obtidos se concentram majoritariamente dentro do traçado da região do Contorno, enquanto que alguns a extrapolam, devido ao caráter aleatório das perturbações utilizadas para o cálculo da região *Bootstrap*. O emprego de mais pontos ao método possibilitaria que a região fosse definida com melhor resolução, permitindo que o algoritmo se distribuisse na área interna do traçado, contudo, como valor de amostras geradas foi suficiente para observar as regiões e as características de suas geometrias.

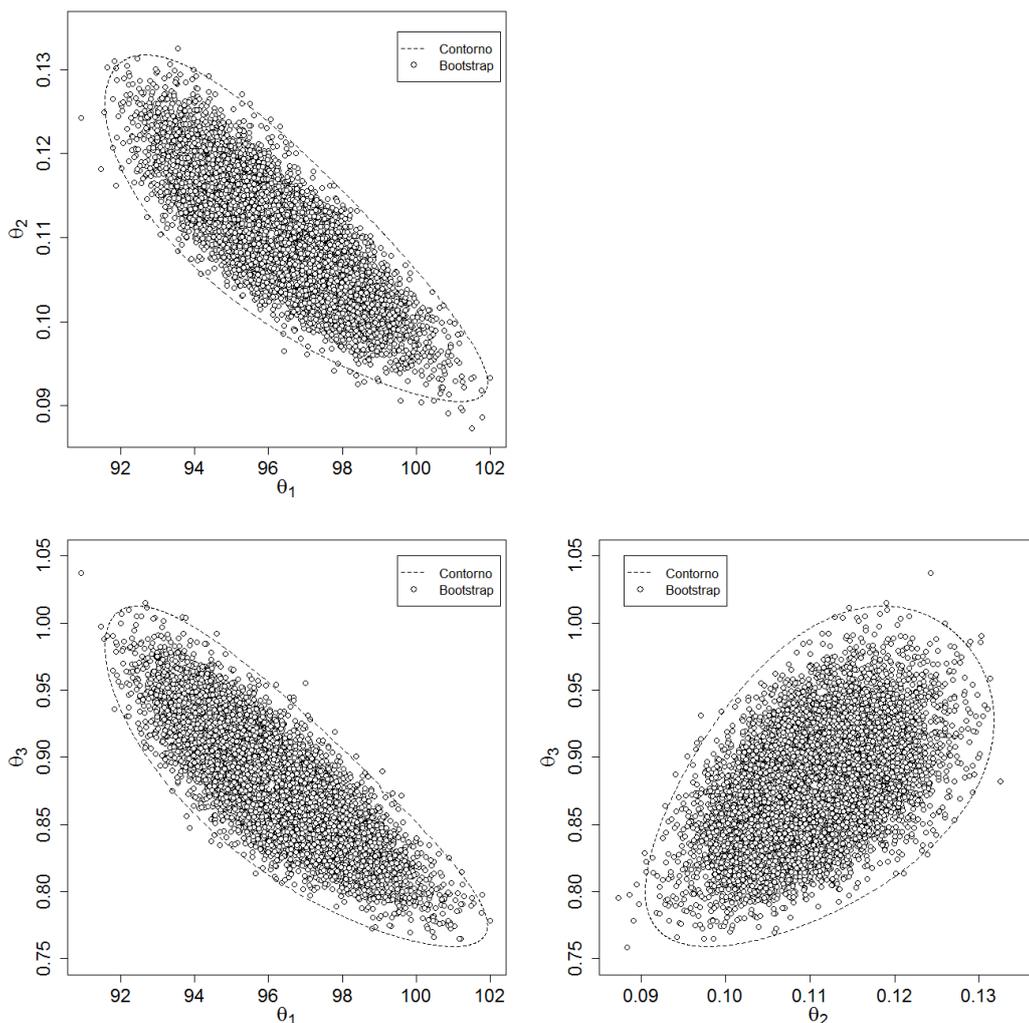


Figura 31: Comparação entre as regiões de confiança do Modelo 5.A determinadas pelo método Bootstrap (pontos) e pelo método do Contorno (linha tracejada).

O último modelo a ser estudado neste trabalho também utiliza a isoterma de SIPS, (1948), Equação (3.5), e o conjunto de variáveis dependentes e independentes apresentados na Tabela 8, chamados de Modelo 5.B. O valor da função objetivo após a minimização é igual a 17,01 e corresponde aos valores ótimos dos parâmetros  $\hat{\theta}_1$ ,  $\hat{\theta}_2$  e  $\hat{\theta}_3$  iguais a 140,46; 0,6734 e 1,39 respectivamente.

A Figura 32 apresenta as comparações entre os valores das variáveis dependentes da Tabela 8 e os obtidos com Modelo 5.B utilizando os valores dos parâmetros ótimos e das variáveis independentes da mesma tabela. Comparando esta figura com a Figura 28, é possível perceber que os dados experimentais estão concentrados na região horizontal da isoterma, enquanto, os valores do Modelo 5.A se distribuem melhor ao longo da curva. Tal comportamento serve para ilustrar experimentos mal planejados para a estimação dos parâmetros do Modelos 5.A, visto que uma amostragem ideal priorizaria os resultados em locais estratégicos da isoterma, como a região onde há a curvatura.

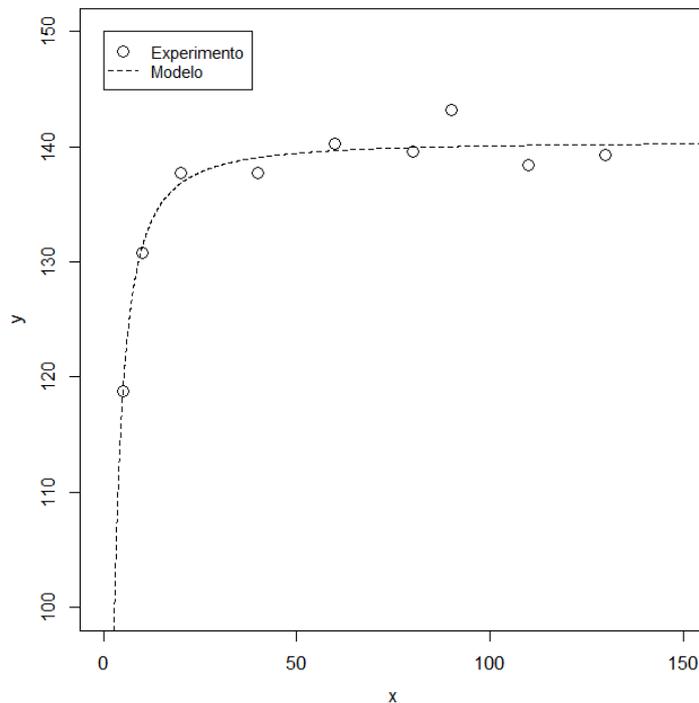


Figura 32: Comparação entre os valores de saída experimentais e os preditos para o Modelo 5.B.

As correlações observadas entre os parâmetros  $\theta_1$  e  $\theta_2$  é igual a 0,61, para os parâmetros  $\theta_1$  e  $\theta_3$  é igual a -0,72 e para parâmetros  $\theta_2$  e  $\theta_3$  é igual a -0,97. As

covariâncias obtidas para os parâmetros são representadas na matriz definida pela Equação (4.11).

$$V_{\theta} = \begin{pmatrix} 0,933 & 0,121 & -0,206 \\ 0,121 & 4,30 \cdot 10^{-2} & -5,95 \cdot 10^{-2} \\ -0,206 & -5,95 \cdot 10^{-2} & 8,81 \cdot 10^{-2} \end{pmatrix} \quad (4.11)$$

As regiões de confiança dos parâmetros geradas com o método do Contorno para o Modelo 5.B podem ser observadas na Figura 33. O primeiro fato observável nas curvas é que o Contorno não conseguiu delimitar completamente as regiões, diferentemente do que foi visto para o demais modelos. As regiões de confiança dos parâmetros  $\theta_1$ ,  $\theta_2$  e  $\theta_1$ ,  $\theta_3$  possuem mais mudanças de direções do que as apresentadas no Modelo 5.A para os mesmos conjuntos de parâmetros, sendo que a necessidade de gerar mais pontos para definir as mudanças de direção da região, compromete o seu desenho completo. Já para a região envolvendo os parâmetros  $\theta_2$  e  $\theta_3$ , a dificuldade em completar o seu traçado pode estar relacionada com a correlação alta entre os parâmetros, posto que a região se torna mais estreita e o método apresenta dificuldade em realizar a mudança de direção.

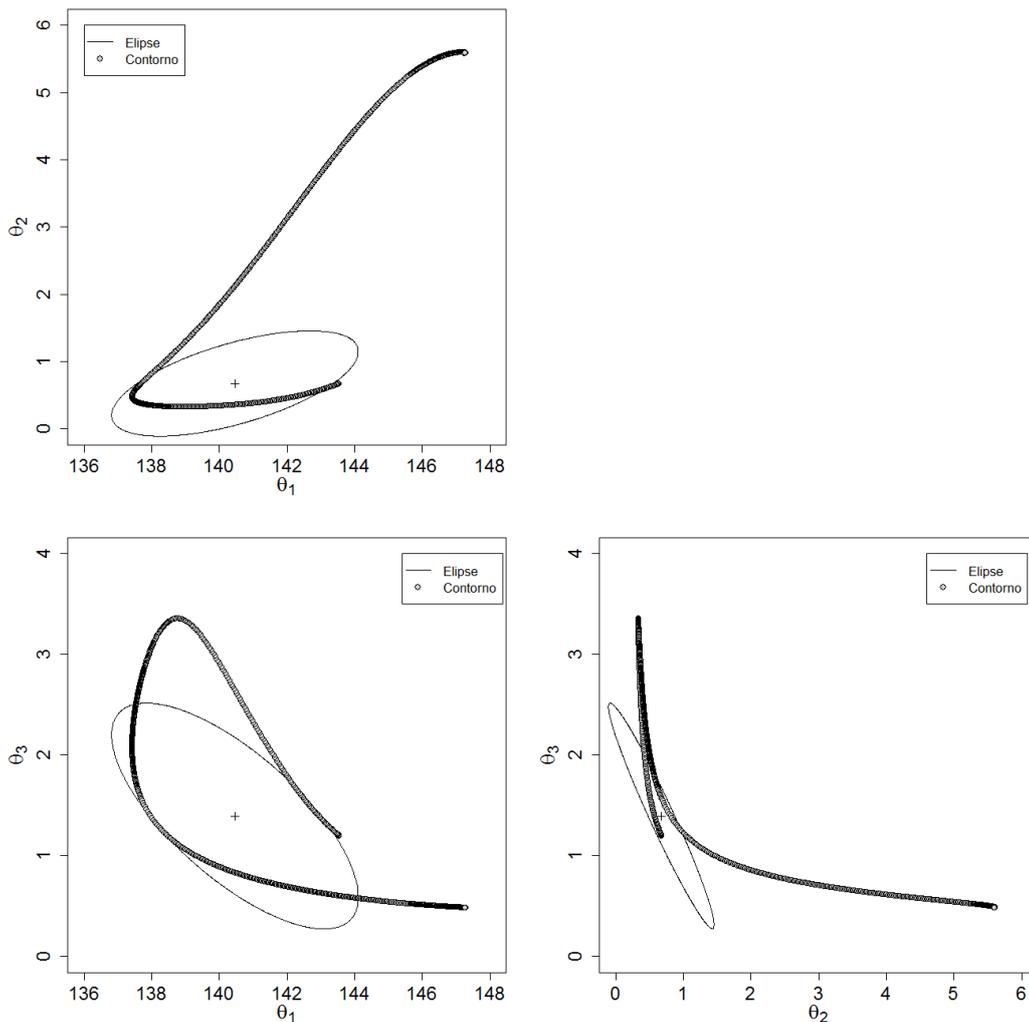


Figura 33: Comparação entre as regiões de confiança do Modelo 5.B determinadas pelo método do Contorno (pontos) e pela Equação (2.34) (linha cheia).

Embora a Equação (3.5) represente um modelo não-linear, as regiões obtidas para o Modelo 5.A mantêm a semelhança com o formato elíptico, diferindo na formação de uma curvatura na área próxima ao ponto ótimo. O Modelo 5.B, por sua vez, tende a se distanciar desse comportamento, sendo mais curvado nas proximidades do ponto ótimo e a mudança de direção nos extremos da região é mais abrupta.

A característica principal da região de confiança deste modelo inviabiliza o uso do método do Perfil, visto que a complexidade de seu traçado dificulta a ação do método. Assim, não foi possível representar esta região para o Modelo 5.B.

O método *Bootstrap* para delimitar as regiões de confiança dos parâmetros do Modelo 5.B é demonstrado na Figura 34. O valor de sua variância é igual a 4,0, e valor da função objetivo pelo método dos mínimos quadrados ponderados é igual a 4,25.

Os pontos gerados pelo método *Bootstrap* se concentram nas proximidades do ponto ótimo, assim como foi observado para os demais modelos vistos nesse trabalho. Este fato está bem nítido para os parâmetros  $\theta_1$  e  $\theta_2$ , já que a região de confiança não está centrada no ponto mínimo e, assim, o método não preenche completamente o espaço delimitado pela linha pontilhada. A região de confiança dos parâmetros  $\theta_1$  e  $\theta_3$  segue um comportamento similar ao descrito acima, estando o ponto ótimo deslocado em relação ao centro da região e, por isso, não a preenche inteiramente. Já para a região relacionando os parâmetros  $\theta_2$  e  $\theta_3$ , o método *Bootstrap* descreve melhor o seu espaço interno, porém também é possível notar que os pontos não ocupam completamente os extremos. Caso fossem utilizados mais pontos para a confecção da região de confiança, estes espaços seriam melhor preenchidos e seria possível a visualizar por completo.

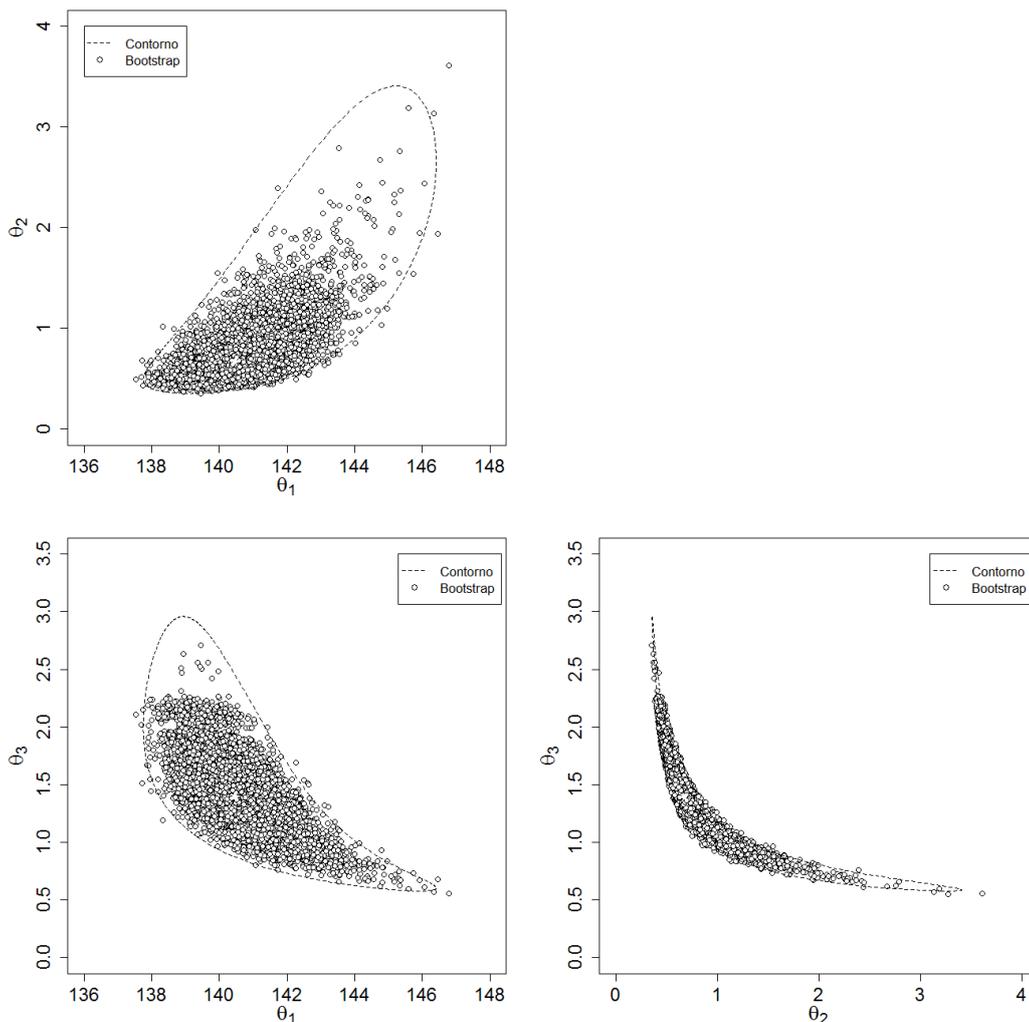


Figura 34: Comparação entre as regiões de confiança do Modelo 5.B determinadas pelo método *Bootstrap* (pontos) e pelo método do Contorno (linha tracejada).

As regiões descritas pelo método Contorno foram representadas para os Modelos 4 e 5.A sem problemas, sendo que ambas apresentaram o mesmo comportamento: necessitar de mais iterações em lugares onde a região muda de direção, como nas suas extremidades. Já para o Modelo 5.B, o método do Contorno tem dificuldade em completar a região, pois devido as curvaturas da região, bem como as extremidades afuniladas, o número de iterações utilizadas para descrever a região é ultrapassado e o algoritmo falha em concluí-la. O algoritmo tende a retornar pelo sentido já traçado, pois o seu sentido é influenciado pela mudança abrupta de direção da região.

O método do Perfil descreveu as regiões de confiança dos Modelos 4 e 5.A sem problemas e com o mesmo comportamento observado para os demais modelos no qual ele foi utilizado: utilizando como limite o valor da distribuição *t-Student* ao invés do valor da *F* de Fisher; e reproduzir a mesma geometria e inclinação da região de confiança obtida com outros métodos. A região do Perfil não conseguiu descrever o Modelo 5.B, devido ao comportamento peculiar de suas regiões de confiança, como a presença de várias curvaturas envolvendo os parâmetros  $\theta_1$ ,  $\theta_2$ , e  $\theta_1$ ,  $\theta_3$ , e a alta correlação entre os parâmetros  $\theta_2$  e  $\theta_3$ .

O método *Bootstrap* consegue reproduzir as regiões de confiança de razão de verossimilhança de todos os modelos testados, mantendo o formato da região. O principal problema visto se dá quando a região não está centrada no valor ótimo dos parâmetros, pois o método tende a concentrar mais pontos em sua proximidade. Isso ocorre porque a região de Bootstrap reproduz a distribuição de probabilidades das estimativas dos parâmetros, concentrando os valores próximos do centro onde a probabilidade é maior. Apesar de ser um resultado importante, a limitação da região de confiança fica prejudicada. Assim, faz-se necessário utilizar mais pontos iniciais para delimitar o espaço completo da região, implicando em um gasto computacional mais elevado.

## Capítulo 5: Conclusão

---

Os estudos realizados neste trabalho consideram diferentes metodologias para delimitar a região de confiança dos estimadores de parâmetros de diferentes modelos. A primeira parte deste trabalho consistiu em verificar se o uso da matriz Hessiana, utilizada no cálculo da equação da região de confiança elíptica, impacta na qualidade da região gerada, atestando se a aproximação de Gauss-Newton pode ser utilizada sem prejudicar a qualidade do ajuste. Para validar se a implementação da matriz Hessiana estava correta, as regiões de confiança elípticas de um modelo linear com dois parâmetros, Modelo 1, foram comparadas, uma utilizando a matriz Hessiana completa e a outra, a aproximação de Gauss-Newton. Neste caso particular, ambas as regiões estavam sobrepostas, validando que o cálculo das derivadas de segunda ordem está correto, já que, para modelos lineares, o seu valor é nulo.

O cálculo das regiões de confiança elípticas para os modelos não-lineares foi realizado após a validação da matriz Hessiana. Foram escolhidos três modelos, sendo que dois deles possuem a mesma equação. Para os Modelo 2.A e 2.B, a região de confiança elíptica completa possui os extremos mais alongados quando comparadas com suas respectivas aproximações, mas a zona menos excêntrica de ambas coincide; já para o Modelo 3, as duas regiões elípticas estão sobrepostas, ou seja, não há diferença entre os métodos. A comparação visual entre as duas regiões geradas demonstra que a aproximação de Gauss-Newton é uma abordagem válida, visto que, para os modelos analisados há pouca ou nenhuma diferença entre as regiões obtidas. Outro aspecto que se deve considerar é o comportamento da real região de confiança do modelo que, para os não-lineares, difere de uma elipse. No Capítulo 4, também foi possível delimitar as regiões de confiança destes modelos e verificar que ela difere da aproximação elíptica, ou seja, não vale a pena deixar a equação da elipse mais exata se o seu resultado não condiz com o verdadeiro comportamento da região de confiança.

Na segunda parte deste trabalho, abordada no Capítulo 4, foram utilizados três diferentes métodos para delimitar a região de confiança dos parâmetros de sete modelos, dois lineares e cinco não-lineares, com diferentes números de parâmetros, quatro modelos com dois parâmetros e os outros três com três parâmetros. Para a

confeção das regiões de confiança foram implementados os métodos do Contorno, Perfil e *Bootstrap*.

O método do Contorno conseguiu descrever sem problemas todas as regiões de confiança dos modelos com dois parâmetros. Já para os modelos com três parâmetros, o método falhou ao tentar representar a região do Modelo 5.B. Este modelo, em especial, possui regiões de confiança com mudanças abruptas de direção, fato este que contribuiu para que o algoritmo não conseguisse completar o traçado da figura. Também para este modelo, é possível verificar que há um par de parâmetros com alta correlação, o que contribuiu para que o método não conseguisse delimitar a região.

O método do Perfil foi capaz de delimitar as regiões de confiança de cinco modelos, não tendo sucesso para os Modelos 3 e 5.B. A principal ideia deste método é sumarizar os parâmetros em perfis para indicar a linearidade do modelo e, através destes, delimitar a região de confiança. A principal característica deste método é que as regiões geradas são menores do que as usuais, visto que seu valor é limitado pela distribuição *t-Student* e não pela *F* de Fisher. Para modelos com parâmetros muito correlacionados, como o Modelo 5.B, ou com regiões de confiança abertas, como o Modelo 4, o funcionamento do método é comprometido, visto que não é possível gerar os perfis. Esta falha pode estar associada com a implementação do método, a qual é duvidosa para calcular o traçado da região.

O método do *Bootstrap* teve sucesso em descrever as regiões de confiança de todos os modelos utilizados. Diferentemente dos demais métodos utilizados, o *Bootstrap* considera a variância dos dados experimentais para o cálculo das perturbações, desta forma é possível utilizar a função dos mínimos quadrados ponderados. As regiões resultantes, quando comparadas como as regiões elípticas ou com as do método do Contorno, extrapolam os limites, devido ao comportamento aleatório empregado no cálculo das perturbações das variáveis experimentais. O melhor preenchimento da região de confiança seria possível aumentando o número de pontos gerados pelo algoritmo, contudo, a quantidade utilizada, 10.000 pontos, foi suficiente para descrever o comportamento da região.

Ao comparar os quatro métodos abordados neste trabalho, Elipse, Contorno, Perfil e *Bootstrap*, é possível definir qual deles representa da melhor forma a região de confiança de um modelo. As regiões elípticas são exatas apenas para modelos

lineares, logo o seu uso acarreta erros para os demais modelos. O método do Perfil foi capaz de traçar as regiões de confiança dos modelos limitadas pela distribuição *t-Student*, mas apresentou problemas para casos específicos, não conseguindo as representar. A grande vantagem deste método está em utilizá-lo para explicar a linearidade dos modelos. O método do Contorno delimita quase todas as regiões de confiança, porém encontra dificuldades nas que possuem mudanças abruptas de direção. O método *Bootstrap* foi capaz de representar as regiões de todos os modelos vistos neste trabalho, sendo o mais versátil deles. Embora este método resulte na construção de outro tipo de região de confiança, ele pode ser empregado sem problemas em todos os tipos de modelo, sendo necessário apenas conhecer ou estimar a variância dos dados experimentais.

Como sugestão para trabalhos futuros, ressaltam-se a melhora do método do Contorno para que ele seja capaz de descrever as regiões de confiança com mudanças abruptas na direção do seu traçado e o aprofundamento no estudo da região do Perfil, para que seja possível o utilizar para representar as regiões com correlação elevada entre os parâmetros e as regiões abertas. Também seria interessante aplicar estes métodos a outros modelos, bem como, testar os que possuem mais parâmetros.

## REFERÊNCIAS

---

ALMEIDA, A. M.; CASTEL-BRANCO, M. M.; FALCÃO, A. C. Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods. **Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences**, [S. l.], v. 774, n. 2, p. 215–222, 2002. Disponível em: [https://doi.org/10.1016/S1570-0232\(02\)00244-1](https://doi.org/10.1016/S1570-0232(02)00244-1)

ANDREI, N. An acceleration of gradient descent algorithm with backtracking for unconstrained optimization. **Numerical Algorithms**, [S. l.], v. 42, n. 1, p. 63–73, 2006. Disponível em: <https://doi.org/10.1007/s11075-006-9023-9>

BAEYENS, E.; HERREROS, A.; PERÁN, J. R. A direct search algorithm for global optimization. **Algorithms**, [S. l.], v. 9, n. 2, p. 1–22, 2016. Disponível em: <https://doi.org/10.3390/a9020040>

BARD, Y. **Nonlinear Parameter Estimation**. San Diego: Academic Press, 1974. *E-book*.

BATES, D. M.; WATTS, D. G. **Nonlinear Regression Analysis and Its Applications**. 1. ed. New York: John Wiley & Sons, 1988. *E-book*. Disponível em: <https://doi.org/10.1002/9780470316757>

BATES, D. M.; WATTS, D. G. Model building in chemistry using profile t and trace plots. **Chemometrics and Intelligent Laboratory Systems**, [S. l.], v. 10, n. 1–2, p. 107–116, 1991. Disponível em: [https://doi.org/10.1016/0169-7439\(91\)80039-S](https://doi.org/10.1016/0169-7439(91)80039-S)

BEALE, E. M. L. Confidence Regions in Non-Linear Estimation. **Journal of the Royal Statistical Society**, [S. l.], v. 22, n. 1, p. 41–88, 1960. Disponível em: <https://doi.org/https://doi.org/10.1111/j.2517-6161.1960.tb00353.x>

BENYAHIA, B. *et al.* Emulsion copolymerization of styrene and butyl acrylate in the presence of a chain transfer agent. Part 2: Parameters estimability and confidence regions. **Chemical Engineering Science**, [S. l.], v. 90, p. 110–118, 2013. Disponível em: <https://doi.org/10.1016/j.ces.2012.12.013>

BRADLEY, E.; GONG, G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. **The American Statistician**, [S. l.], v. 37, n. 1, p. 36–48, 1983. Disponível em: <https://doi.org/https://doi.org/10.2307/2685844>

CAPES/MEC. **Portal de periódicos**. [s. l.], 2020. Disponível em: <http://www.periodicos.capes.gov.br/>.

CHECCHI, N.; GIUSTI, E.; MARSILI-LIBELLI, S. PEAS: A toolbox to assess the accuracy of estimated parameters in environmental models. **Environmental Modelling and Software**, [S. l.], v. 22, n. 6, p. 899–913, 2007. Disponível em: <https://doi.org/10.1016/j.envsoft.2006.05.019>

CHEN, P.-H. **Particle swarm optimization for power dispatch with pumped hydro**. Taiwan: [s. n.], 2009. Disponível em: <https://doi.org/10.5772/109>.

CUI, H.; YAN, G.; SONG, H. A novel curvelet thresholding denoising method based on chi-squared distribution. **Signal, Image and Video Processing**, [S. l.], v. 9, n. 2, p. 491–498, 2015. Disponível em: <https://doi.org/10.1007/s11760-014-0685-9>

DA ROS, S. *et al.* A comparison among stochastic optimization algorithms for parameter estimation of biochemical kinetic models. **Applied Soft Computing Journal**, [S. l.], v. 13, n. 5, p. 2205–2214, 2013. Disponível em: <https://doi.org/10.1016/j.asoc.2013.01.019>

DONALDSON, J. R.; SCHNABEL, R. B. Computational Experience With Confidence Regions and Confidence Intervals for Nonlinear Least Squares. **Technometrics**, [S. l.], v. 29, n. 1, p. 67–82, 1987. Disponível em: <https://doi.org/https://doi.org/10.2307/1269884>

DORNELES, N. A. **Avaliação de algoritmos para determinação de regiões de confiança de parâmetros**. 2018. 29 f. Trabalho de Conclusão de Curso - Universidade Federal do Rio Grande do Sul, [s. l.], 2018. Disponível em: <http://hdl.handle.net/10183/182405>

EFRON, B. BOOTSTRAP METHODS: ANOTHER LOOK AT THE JACKKNIFE. **The Annals of Statistics**, [S. l.], v. 7, n. 1, p. 1–26, 1979. Disponível em: <https://doi.org/10.1214/aoms/1177692541>

FENG, X.; BAI, J.; ZHENG, X. On the use of graphical method to determine the targets of single-contaminant regeneration recycling water systems. **Chemical Engineering Science**, [S. l.], v. 62, n. 8, p. 2127–2138, 2007. Disponível em: <https://doi.org/10.1016/j.ces.2006.12.081>

GUELY, F.; SIARRY, P. Gradient descent method for optimizing various fuzzy rule bases. **1993 IEEE International Conference on Fuzzy Systems**, [S. l.], n. 1, p. 1241–1246, 1993. Disponível em: <https://doi.org/10.1109/fuzzy.1993.327570>

HARIK, G. R.; LOBO, F. G.; GOLDBERG, D. E. Brief Papers. **IEEE Transactions on Evolutionary Computation**, [S. l.], v. 3, n. 4, p. 287–297, 1999. Disponível em: <https://doi.org/10.1109/4235.797971>

HARTLEY, H. O. Exact Confidence Regions for the Parameters in Non-Linear Regression Laws. **Biometrika**, [S. l.], v. 51, n. 3, p. 347–353, 1964. Disponível em: <https://doi.org/https://doi.org/10.2307/2334140>

HIMMELBLAU, D. M. **Applied Nonlinear Programming**. 1972. - McGraw-Hill, New York, 1972.

HOLLAND, J. **Adaptation In Natural and Artificial Systems**. Ann Arbor: University of Michigan Press, 1975. *E-book*.

HOOKE, R.; JEEVES, T. A. "Direct Search" Solution of Numerical and Statistical Problems. **Journal of the ACM**, [S. l.], v. 8, n. 2, p. 212–229, 1961. Disponível em: <https://doi.org/10.1145/321062.321069>

JOSHI, M.; SEIDEL-MORGENSTERN, A.; KREMLING, A. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. **Metabolic Engineering**, [S. l.], v. 8, n. 5, p. 447–455, 2006. Disponível em: <https://doi.org/10.1016/j.ymben.2006.04.003>

KANG, G.; BATES, D. M. Approximate Inferences in Multiresponse Regression Analysis. **Biometrika**, [S. l.], 1990. Disponível em: <https://doi.org/10.2307/2336810>

KENNEDY, J.; EBERHART, R. Particle swarm optimization. *In*: 1995, Perth, Australia. **Proceedings of ICNN'95 - International Conference on Neural Networks**. Perth, Australia: [s. n.], 1995. p. 1942–1948.

KESSENTINI, S.; BARCHIESI, D. Particle Swarm Optimization with Adaptive Inertia Weight. **International Journal of Machine Learning and Computing**, [S. l.], v. 5, n. 5, p. 368–373, 2015. Disponível em: <https://doi.org/10.7763/IJMLC.2015.V5.535>

KLEPPER, O.; HENDRIX, E. M. T. A COMPARISON OF ALGORITHMS FOR GLOBAL CHARACTERIZATION OF CONFIDENCE REGIONS FOR NONLINEAR MODELS. **Environmental Toxicology and Chemistry**, [S. l.], v. 13, n. 12, p. 1887–1899, 1994. Disponível em: <https://doi.org/https://doi.org/10.1002/etc.5620131204>

KODURU, P.; WELCH, S. M.; DAS, S. A particle swarm optimization approach for estimating parameter confidence regions. **Proceedings of the 9th annual conference on Genetic and evolutionary computation - GECCO '07**, [S. l.], p. 70, 2007. Disponível em: <https://doi.org/10.1145/1276958.1276969>

LEWIS, R. M. *et al.* Direct search methods : then and now. **Journal of Computational and Applied Mathematics**, [S. l.], v. 124, n. 1–2, p. 191–207, 2000. Disponível em: [https://doi.org/https://doi.org/10.1016/S0377-0427\(00\)00423-4](https://doi.org/https://doi.org/10.1016/S0377-0427(00)00423-4)

LOBRY, J.; ROSSO, L.; FLANDROIS, J. A fortran subroutine for the determination of parameter confidence limits in non-linear models. **Binary**, [S. l.], v. 3, p. 86–93, 1991. Disponível em: <https://hal.archives-ouvertes.fr/hal-00698031>

MARSILI-LIBELLI, S.; CHECCHI, N. Identification of dynamic models for horizontal subsurface constructed wetlands. **Ecological Modelling**, [S. l.], v. 187, n. 2–3, p. 201–218, 2005. Disponível em: <https://doi.org/10.1016/j.ecolmodel.2005.01.043>

MARSILI-LIBELLI, S.; GUERRIZIO, S.; CHECCHI, N. Confidence regions of estimated parameters for ecological systems. **Ecological Modelling**, [S. l.], v. 165, n. 2–3, p. 127–146, 2003. Disponível em: [https://doi.org/10.1016/S0304-3800\(03\)00068-1](https://doi.org/10.1016/S0304-3800(03)00068-1)

MONTEIRO, A. J. T. **Otimização não linear de mínimos quadrados**. 2013. - Universidade de Aveiro, Aveiro, 2013. Disponível em: <https://doi.org/http://hdl.handle.net/10961/3228>

NHAMAGE, I. A. *et al.* AN HYBRID STOCHASTIC-DETERMINISTIC OPTIMIZATION ALGORITHM FOR STRUCTURAL DAMAGE IDENTIFICATION. **Revista Interdisciplinar de Pesquisa em Engenharia**, [S. l.], v. 2, n. 17 SE-, p. 182–197, 2017. Disponível em: <https://doi.org/10.26512/ripe.v2i17.21658>

NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization**. 2. ed. New York: Springer, 2006. *E-book*.

ÖZSOY, V. S.; ÖRKÇÜ, H. H. Estimating the parameters of nonlinear regression models through particle swarm optimization. **Gazi University Journal of Science**, [S. l.], v. 29, n. 1, p. 187–199, 2016. Disponível em: <https://dergipark.org.tr/tr/download/article-file/230923>

PINTO, J. C.; LAGE, P. L. C. **Métodos Numéricos em Problemas de Engenharia Química**. 1. ed. Rio de Janeiro: E-papers, 2001. *E-book*.

PRESS, W. H. *et al.* **Numerical Recipes in Fortran 90**. 2. ed. [S. l.]: Cambridge University Press, 1996. v. 2 *E-book*.

QUENOUILLE, M. H. Approximate tests of correlation in time-series 3. **Mathematical Proceedings of the Cambridge Philosophical Society**, [S. l.], v. 45, n. 3, p. 483–484, 1949. Disponível em: <https://doi.org/10.1017/S0305004100025123>

RUDER, S. An overview of gradient descent optimization algorithms. [S. l.], 2016. Disponível em: <http://arxiv.org/abs/1609.04747>. Acesso em: 20 fev. 2020.

SCHWAAB, M. **Desenvolvimento e implementação de novas técnicas de estimação de parâmetros e planejamento sequencial de experimentos**. 2007. 171 f. Tese (Doutorado em Engenharia Química) - Universidade Federal do Rio de Janeiro, [s. l.], 2007. Disponível em: <https://doi.org/616.89-008.47:616->

SCHWAAB, M. *et al.* Nonlinear parameter estimation through particle swarm optimization. **Chemical Engineering Science**, [S. l.], v. 63, n. 6, p. 1542–1552, 2008. Disponível em: <https://doi.org/10.1016/j.ces.2007.11.024>

SCHWAAB, M.; ALBERTON, A. L.; PINTO, J. C. **ESTIMA & PLANEJA: Pacote computacional para estimação de parâmetros e de planejamento de experimentos**. Rio de Janeiro: [s. n.], 2010.

SCHWAAB, M.; PINTO, J. C. **Análise de Dados Experimentais I**. 1. ed. Rio de Janeiro: E-papers, 2007 a. *E-book*.

SCHWAAB, M.; PINTO, J. C. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant. **Chemical Engineering Science**, [S. l.], v. 62, n. 10, p. 2750–2764, 2007 b. Disponível em: <https://doi.org/10.1016/j.ces.2007.02.020>

SHI, Y.; EBERHART, R. A modified particle swarm optimizer. **1998 IEEE International Conference on Evolutionary Computation Proceedings.**, [S. l.], p. 69–73, 1998. Disponível em: <https://doi.org/10.1109/ICEC.1998.699146>

SILVA, A. J. M. **Implementação de um algoritmo genético utilizando o modelo de ilhas**. 2006. - Universidade Federal do Rio de Janeiro, [S. l.], 2006.

SILVA FILHO, A. S. da. Inferência Em Amostras Pequenas : Métodos Bootstrap. **Revista de Ciências Exatas e Tecnologia**, [S. l.], v. 5, n. 5, p. 115–126, 2010. Disponível em: <https://doi.org/https://doi.org/10.17921/1890-1793.2010v5n5p115-126>

SIMONOFF, J. S.; TSAI, C. L. Jackknife-based estimators and confidence regions in nonlinear regression. **Technometrics**, [S. l.], v. 28, n. 2, p. 103–112, 1986. Disponível em: <https://doi.org/10.1080/00401706.1986.10488111>

SIPS, R. On the structure of a catalyst surface. **The Journal of Chemical Physics**, [S. l.], v. 16, n. 5, p. 490–495, 1948. Disponível em: <https://doi.org/10.1063/1.1746922>

SULIEMAN, H.; MCLELLAN, P. J.; BACON, D. W. A Profile-Based Approach to Parametric Sensitivity Analysis of Nonlinear Regression Models. **TECHNOMETRICS**, [S. l.], v. 43, n. 4, p. 37–41, 2001. Disponível em: <https://doi.org/https://doi.org/10.1198/00401700152672519>

TARANTOLA, A. **Inverse Problem Theory and Methods for Model Parameter Estimation**. 2. ed. [S. l.]: Society for Industrial and Applied Mathematics., 2005. *E-book*.

VAN DEN BOS, A. **Parameter Estimation for Scientists and Engineers**. 1. ed. New York: Wiley-Blackwell, 2007. *E-book*. Disponível em: <https://doi.org/10.1002/9780470173862>

WANG, T.; YANG, J. A heuristic method for learning Bayesian networks using discrete particle swarm optimization. **Knowledge and Information Systems**, [S. l.], v. 24, n. 2, p. 269–281, 2010. Disponível em: <https://doi.org/10.1007/s10115-009-0239-6>

WATTS, D. G. Estimating parameters in nonlinear rate equations. **The Canadian Journal of Chemical Engineering**, [S. l.], v. 72, n. 4, p. 701–710, 1994. Disponível em: <https://doi.org/10.1002/cjce.5450720420>

WHITLEY, D. A genetic algorithm tutorial. **Statistics and Computing**, [S. l.], v. 4, n. 2, p. 65–85, 1994. Disponível em: <https://doi.org/10.1007/BF00175354>

WILLIAMS, E. J. Exact Fiducial Limits in Non-Linear Estimation. **Journal of the Royal Statistical Society**, [S. l.], v. 24, n. 1, p. 125–139, 1962. Disponível em: <https://doi.org/https://doi.org/10.1111/j.2517-6161.1962.tb00444.x>

YOO, S. K. *et al.* The Fisher-Snedecor F Distribution: A Simple and Accurate Composite Fading Model. **IEEE Communications Letters**, [S. l.], v. 21, n. 7, p. 1661–1664, 2017. Disponível em: <https://doi.org/10.1109/LCOMM.2017.2687438>

## APÊNDICES

---

### A.1 Aproximação Numérica da Matriz Hessiana

No Capítulo 4 deste trabalho foi abordado o uso da matriz Hessiana completa para calcular as regiões de confiança elípticas. Também, no Capítulo 3 foi argumentada a escolha de um método para calcular os valores dessa matriz. Embora seja possível calcular as derivadas de segunda ordem analiticamente, o recálculo a cada novo modelo testado e a necessidade de cálculos exaustivos para alguns modelos proporciona um ambiente favorável para o uso de métodos numéricos, os quais só precisam ser implementados uma única vez.

NOCEDAL e WRIGHT, (2006), abordam no Capítulo 8 de sua obra as expressões para o cálculo de duas expressões utilizadas para calcular a matriz Hessiana numérica. A primeira determina o valor de uma função diferenciada duas vezes em relação ao mesmo parâmetro, Equação (A.1-1), e a segunda determina o valor da função diferenciada em relação a dois diferentes parâmetros, Equação (A.1-2).

#### A.1.1 Derivadas de Segunda Ordem em Relação a um Parâmetro

Para determinar a expressão numérica para o cálculo da função diferenciada duas vezes em relação ao mesmo parâmetro, ou seja, os termos da diagonal principal da matriz Hessiana, é estabelecida a expressão que define uma função expandida em série de Taylor e com uma perturbação positiva no valor da variável a ser diferenciada.

$$f(x + \Delta x_i e_i) = f(x) + \Delta x_i \frac{\partial f}{\partial x_i} + \Delta x_i^2 \frac{1}{2} \frac{\partial^2 f}{\partial x_i^2} \quad (\text{A.1-1})$$

Onde  $f(x)$  é o valor da função para o ponto  $x$ ,  $\Delta x_i$  é uma perturbação pequena, escalar e positiva na direção do parâmetro  $i$  e  $e_i$  é um vetor unitário na direção do parâmetro  $i$ .

Analogamente a Equação (A.1-1), é definida a função com perturbação negativa na variável a ser diferenciada.

$$f(x - \Delta x_i e_i) = f(x) - \Delta x_i \frac{\partial f}{\partial x_i} + \Delta x_i^2 \frac{1}{2} \frac{\partial^2 f}{\partial x_i^2} \quad (\text{A.1-2})$$

Somando as Equações (A.1-1) e (A.1-2) e isolando o termo da derivada de segunda ordem em um dos lados, obtêm-se a Equação (A.1-3), que define a aproximação numérica dos termos da diagonal principal da matriz Hessiana.

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{f(x + \Delta x_i e_i) - 2f(x) + f(x - \Delta x_i e_i)}{\Delta x_i^2} + O(\Delta x_i) \quad (\text{A.1-3})$$

#### A.1.2 Derivadas de Segunda Ordem em Relação a dois Parâmetros

A expressão para determinar os valores das derivadas de segunda ordem de uma função em relação a dois diferentes parâmetros, ou seja, os termos diagonal superior e inferior da matriz Hessiana, são obtidas pela perturbação positiva no valor dos dois parâmetros. Considerando a Equação (A.1-1) definida na seção anterior e na Equação (A.1-4).

$$f(x + \Delta x_j e_j) = f(x) + \Delta x_j \frac{\partial f}{\partial x_j} + \Delta x_j^2 \frac{1}{2} \frac{\partial^2 f}{\partial x_j^2} \quad (\text{A.1-4})$$

Onde  $\Delta x_j$  é uma perturbação pequena, escalar e positiva na direção do parâmetro  $j$  e  $e_j$  é um vetor unitário na direção do parâmetro  $j$ .

E considerando uma perturbação positiva nas duas direções, tem-se a Equação (A.1-5).

$$f(x + \Delta x_i e_i + \Delta x_j e_j) = f(x) + \Delta x_i \frac{\partial f}{\partial x_i} + \Delta x_j \frac{\partial f}{\partial x_j} + \frac{1}{2} \Delta x_i^2 \frac{\partial^2 f}{\partial x_i^2} + \Delta x_i \Delta x_j \frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{1}{2} \Delta x_j^2 \frac{\partial^2 f}{\partial x_j^2} \quad (\text{A.1-5})$$

Subtraindo as Equações (A.1-1) e (A.1-4) da Equação (A.1-5), obtêm-se a Equação (A.1-6). Por fim, a derivada de segunda ordem em relação aos dois parâmetros é isolada e a expressão para o cálculo dessa derivada é definida pela Equação (A.1-7).

$$f(x + \Delta x_i e_i + \Delta x_j e_j) - f(x + \Delta x_i e_i) - f(x + \Delta x_j e_j) = -f(x) + \Delta x_i \Delta x_j \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (\text{A.1-6})$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{f(x + \Delta x_i e_i + \Delta x_j e_j) - f(x + \Delta x_i e_i) - f(x + \Delta x_j e_j) + f(x)}{\Delta x_i \Delta x_j} + o(\sqrt{\Delta x_i \Delta x_j}) \quad (\text{A.1-7})$$

## A.2 Equação da Região de Confiança Elíptica

Como foi visto no Capítulo 2, é possível definir algebricamente a função da elipse que representa uma região de confiança através da Equação (2.34). Considerando um modelo com dois parâmetros, pode-se expandir essa equação na seguinte maneira.

$$\left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \right)^T \begin{bmatrix} s_{\theta_1}^2 & s_{\theta_2\theta_1} \\ s_{\theta_1\theta_2} & s_{\theta_2}^2 \end{bmatrix}^{-1} \left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \right) \leq 2F_{2,N-2}^{1-\alpha} \quad (\text{A.2-1})$$

Reescrevendo os valores da matriz inversa presente na Equação (A.2-1).

$$\frac{\left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \right)^T \begin{bmatrix} s_{\theta_2}^2 & -s_{\theta_2\theta_1} \\ -s_{\theta_1\theta_2} & s_{\theta_1}^2 \end{bmatrix} \left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \right)}{\det(V_{\hat{\theta}})} \leq 2F_{2,N-2}^{1-\alpha} \quad (\text{A.2-2})$$

Resolvendo os produtos matriciais, pode-se representar a Equação (A.2-2) da seguinte forma.

$$\theta_1^2 s_{\theta_2}^2 + \theta_2^2 s_{\theta_1}^2 - 2\theta_1\theta_2 s_{\theta_1\theta_2} + 2\theta_1(\hat{\theta}_2 s_{\theta_1\theta_2} - \hat{\theta}_1 s_{\theta_2}^2) + 2\theta_2(\hat{\theta}_1 s_{\theta_1\theta_2} - \hat{\theta}_2 s_{\theta_1}^2) + \hat{\theta}_1^2 s_{\theta_2}^2 + \hat{\theta}_2^2 s_{\theta_1}^2 - 2\hat{\theta}_1\hat{\theta}_2 s_{\theta_1\theta_2} \leq 2 \det(V_{\hat{\theta}}) F_{2,N-2}^{1-\alpha} \quad (\text{A.2-3})$$

A Equação (A.2-3) representa uma função de segundo grau com duas variáveis. Sua solução é obtida de forma análoga a uma equação de segundo grau com apenas uma variável, pelo uso da fórmula de Baskará. Primeiramente, os termos da equação serão reescritos para facilitar a visualização. Definindo os termos como pelas seguintes representações.

$$A = s_{\theta_2}^2 \quad (\text{A.2-4})$$

$$B = s_{\theta_1}^2 \quad (\text{A.2-5})$$

$$C = -2s_{\theta_1\theta_2} \quad (\text{A.2-6})$$

$$D = 2(\hat{\theta}_2 s_{\theta_1\theta_2} - \hat{\theta}_1 s_{\theta_2}^2) \quad (\text{A.2-7})$$

$$E = 2(\hat{\theta}_1 s_{\theta_1\theta_2} - \hat{\theta}_2 s_{\theta_1}^2) \quad (\text{A.2-8})$$

$$F = \hat{\theta}_1^2 s_{\theta_2}^2 + \hat{\theta}_2^2 s_{\theta_1}^2 - 2\hat{\theta}_1\hat{\theta}_2 s_{\theta_1\theta_2} - 2 \det(V_{\hat{\theta}}) F_{2,N-2}^{1-\alpha} \quad (\text{A.2-9})$$

Esse procedimento permite reescrever a Equação (A.2-3) da seguinte forma.

$$\theta_1^2 A + \theta_2^2 B + \theta_1 \theta_2 C + \theta_1 D + \theta_2 E + F = 0 \quad (\text{A.2-10})$$

Explicitando os termos em relação a um dos parâmetros, no caso  $\theta_2$ .

$$\theta_2^2 B + \theta_2(\theta_1 C + E) + \theta_1^2 A + \theta_1 D + F = 0 \quad (\text{A.2-11})$$

Aplicando a fórmula de Baskará à Equação (A.2-11), obtêm-se as expressões que definem as regiões superior e inferior da elipse.

$$\theta_2 = \frac{-(\theta_1 C + E) + \sqrt{(\theta_1 C + E)^2 - 4B(\theta_1^2 A + \theta_1 D + F)}}{2B} \quad (\text{A.2-12})$$

$$\theta_2 = \frac{-(\theta_1 C + E) - \sqrt{(\theta_1 C + E)^2 - 4B(\theta_1^2 A + \theta_1 D + F)}}{2B} \quad (\text{A.2-13})$$

### A.3 Splines

No Capítulo 2, construção das regiões de confiança dos parâmetros pelo método do perfil foi abordada. Nela há um passo fundamental para a confecção dos esboços que se trata do cálculo das fases e ângulos das regiões, para isso, é proposto o cálculo dos *splines*.

Considere um conjunto discreto de tamanho  $N$  de variáveis independentes,  $x$ , e os seus valores correspondentes de variáveis dependentes,  $y$ .

Para se obter um valor específico dentro do intervalo desse conjunto, o método mais indicado é através da aproximação polinomial dos dados, ou seja, utilizar os valores de variáveis dependentes e independentes disponíveis para gerar um polinômio que descreve o comportamento de uma função hipotética no intervalo dos dados. Para isso, se utiliza os *splines*, os quais são um conjunto de  $N - 1$  polinômios que descrevem o comportamento de cada intervalo do conjunto de dados.

A qualidade da aproximação é predita pela ordem do polinômio, sendo que, aqui serão comparados os polinômios de primeira ordem, *splines* lineares, e os polinômios de terceira ordem, *splines* cúbicos.

#### A.3.1 Splines Lineares

Considerando um valor,  $x$ , dentro do intervalo  $[x_i, x_{i+1}]$ , e seus respectivos valores de  $[y_i, y_{i+1}]$ , pode-se estimar um valor para  $y$  correspondente na escala das variáveis dependentes através do polinômio de primeira ordem definido na Equação (A.3-1).

$$y(x) = \frac{x - x_i}{x_{i+1} - x_i} y_i + \frac{x - x_{i+1}}{x_i - x_{i+1}} y_{i+1} \quad (\text{A.3-1})$$

#### A.3.2 Splines Cúbico

Considerando um valor,  $x$ , dentro do intervalo  $[x_i, x_{i+1}]$ , seus respectivos valores de  $[y_i, y_{i+1}]$ , e que todos os  $N$  pontos do intervalo possuem derivadas de primeira e segunda ordem, pode-se representar a derivada de segunda ordem do ponto desejado com a Equação (A.3-2).

$$y'(x) = y'_i + (x - x_i) y''[x_i, x_{i+1}] \quad (\text{A.3-2})$$

Integrando a Equação (A.3-2) duas vezes, obtêm-se as expressões que definem a derivada de primeira ordem da função, Equação (A.3-3), e a própria função, Equação (A.3-4).

$$y'(x_{i+1}) = y'_i + (x_{i+1} - x_i) y''_i + \frac{1}{2} (x_{i+1} - x_i)^2 y''[x_i, x_{i+1}] \quad (\text{A.3-3})$$

$$y(x) = y_i + (x - x_i) y'_i + \frac{1}{2} (x - x_i)^2 y''_i + \frac{1}{6} (x - x_i)^3 y''[x_i, x_{i+1}] \quad (\text{A.3-4})$$

Pode-se reescrever a diferença entre o intervalo do polinômio através da seguinte forma.

$$h_i = x_{i+1} - x_i \quad (\text{A.3-5})$$

Calculando o valor da derivada de primeira ordem e da função para o ponto  $i + 1$ .

$$y'(x_{i+1}) = y'_i + h_i y''_i + \frac{1}{2} h_i^2 y''[x_i, x_{i+1}] \quad (\text{A.3-6})$$

$$y(x_{i+1}) = y_i + h_i y'_i + \frac{1}{2} h_i^2 y''_i + \frac{1}{6} h_i^3 y''[x_i, x_{i+1}] \quad (\text{A.3-7})$$

Reescrevendo o termo  $y''[x_i, x_{i+1}]$  como a diferença das derivadas de segunda ordem ponderadas pela diferença do valor das variáveis independentes, obtêm-se a Equação (A.3-8).

$$y''[x_i, x_{i+1}] = \frac{y''_{i+1} - y''_i}{h_i} \quad (\text{A.3-8})$$

Substituindo a Equação (A.3-8) nas Equações (A.3-6) e (A.3-7) obtêm-se as expressões a seguir.

$$y'(x_{i+1}) = y'_i + h_i y''_i + \frac{1}{2} h_i (y''_{i+1} - y''_i) \quad (\text{A.3-9})$$

$$y(x_{i+1}) = y_i + h_i y'_i + \frac{1}{2} h_i^2 y''_i + \frac{1}{6} h_i^2 (y''_{i+1} - y''_i) \quad (\text{A.3-10})$$

Redefinindo a nomenclatura  $y(x_{i+1}) = y_{i+1}$  e rearranjando os termos da Equação (A.3-10) é possível obter a Equação (A.3-11).

$$\frac{y_{i+1} - y_i}{h_i} = y'_i + \frac{1}{2} h_i y''_i + \frac{1}{6} h_i (y''_{i+1} - y''_i) \quad (\text{A.3-11})$$

Analogamente à representação da Equação (A.3-8), pode-se escrever o termo do lado esquerdo da Equação (A.3-11).

$$f[x_i, x_{i+1}] = y'_i + \frac{1}{2} h_i y''_i + \frac{1}{6} h_i (y''_{i+1} - y''_i) \quad (\text{A.3-12})$$

Juntando os termos da Equação (A.3-12) e calculando a função para os intervalos  $[x_i, x_{i+1}]$  e  $[x_{i-1}, x_i]$ , têm-se as Equações (A.3-13) e (A.3-14) respectivamente.

$$f[x_i, x_{i+1}] = y'_i + \frac{1}{3} h_i y''_i + \frac{1}{6} h_i y''_{i+1} \quad (\text{A.3-13})$$

$$f[x_{i-1}, x_i] = y'_{i-1} + \frac{1}{3} h_{i-1} y''_{i-1} + \frac{1}{6} h_{i-1} y''_i \quad (\text{A.3-14})$$

Subtraindo a Equação (A.3-14) da Equação (A.3-13) é possível definir a Equação (A.3-15).

$$f[x_i, x_{i+1}] - f[x_{i-1}, x_i] = y'_i - y'_{i-1} + \frac{1}{3}h_i y''_i + \frac{1}{6}h_i y''_{i+1} - \frac{1}{3}h_{i-1} y''_{i-1} - \frac{1}{6}h_{i-1} y''_i \quad (\text{A.3-15})$$

Rearranjando a Equação (A.3-9), podem-se isolar a expressão para definir a relação  $y'_i - y'_{i-1}$  descrita pela Equação (A.3-16).

$$y'_i - y'_{i-1} = \frac{1}{2}h_{i-1} y''_{i-1} + \frac{1}{2}h_{i-1} y''_i \quad (\text{A.3-16})$$

Substituindo a Equação (A.3-16) na Equação (A.3-15), escreve-se a Equação (A.3-17).

$$f[x_i, x_{i+1}] - f[x_{i-1}, x_i] = \frac{1}{6}h_{i-1} y''_{i-1} + \frac{1}{3}h_{i-1} y''_i + \frac{1}{3}h_i y''_i + \frac{1}{6}h_i y''_{i+1} \quad (\text{A.3-17})$$

Também, da equação (A.3-13), pode-se rearranjar os seus termos para calcular as derivadas de primeira ordem para cada ponto do conjunto inicial, Equação (A.3-18).

$$y'_i = f[x_i, x_{i+1}] - \frac{1}{6}h_i (2y''_i + y''_{i+1}) \quad (\text{A.3-18})$$

Por fim, para o cálculo do valor do ponto pela interpolação polinomial, utiliza-se a Equação (A.3-4) modificada com a Equação (A.3-8)

$$y(x) = y_i + (x - x_i) y'_i + \frac{1}{2}(x - x_i)^2 y''_i + \frac{1}{6}(x - x_i)^3 \frac{(y''_{i+1} - y''_i)}{h_i} \quad (\text{A.3-19})$$

Os *splines* cúbicos são condicionados pelos valores das derivadas nos pontos extremos do conjunto de dados, como: derivadas de segunda ordem nulas, *spline* natural, e derivadas de primeira ordem conhecidas.

#### A.3.2.1 *Splines* Cúbico Natural

Os *splines* cúbicos naturais correspondem ao caso no qual as derivadas de segunda ordem dos pontos nos extremos do conjunto são nulas. Seus valores para um conjunto de tamanho  $N$  são representados pelas Equações (A.3-20) e (A.3-21).

$$y_0'' = y''(a) = 0 \quad (\text{A.3-20})$$

$$y_N'' = y''(b) = 0 \quad (\text{A.3-21})$$

Utilizando a Equação (A.3-17), pode-se escrever para o intervalo  $i = 1, N - 1$  do conjunto o seguinte sistema na forma matricial, Equação (A.3-22).

$$\begin{bmatrix} \frac{h_0}{3} + \frac{h_1}{3} & \frac{h_1}{6} & 0 & \dots & 0 & 0 \\ \frac{h_1}{6} & \frac{h_1}{3} + \frac{h_2}{3} & \frac{h_2}{6} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{h_{N-2}}{6} & \frac{h_{N-2}}{3} + \frac{h_{N-1}}{3} \end{bmatrix} \cdot \begin{bmatrix} y_1'' \\ y_2'' \\ \vdots \\ y_{N-1}'' \end{bmatrix} = \begin{bmatrix} f[x_1, x_2] - f[x_0, x_1] \\ f[x_2, x_3] - f[x_1, x_2] \\ \vdots \\ f[x_{N-1}, x_N] - f[x_{N-2}, x_{N-1}] \end{bmatrix} \quad (\text{A.3-22})$$

A resolução do sistema matricial fornece os valores das derivadas de segunda ordem para cada valor do conjunto, de posse desses valores e com os impostos nas condições iniciais, aplica-se a Equação (A.3-18) para obter os valores das derivadas de primeira ordem e a Equação (A.3-19) para o cálculo do valor da função do ponto desejado.

### A.3.2.2 Splines Cúbico com condições na derivada

Também é possível calcular casos nos quais os valores das derivadas de segunda ordem dos pontos inicial e final possuam valor diferente de zero. Esses valores para um conjunto de tamanho  $N$  são representados pelas Equações (A.3-23) e (A.3-24).

$$y'_0 = y'(a) = f'_0 \quad (\text{A.3-23})$$

$$y'_N = y'(b) = f'_N \quad (\text{A.3-24})$$

De forma análoga ao *spline* cúbico natural, é possível construir um sistema na forma matricial para a resolução das derivadas de segunda ordem utilizando a Equação (A.3-17) para o intervalo  $i = 0, N$ .

$$\begin{bmatrix} \frac{h_0}{3} & \frac{h_0}{6} & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{h_0}{6} & \frac{h_0}{3} + \frac{h_1}{3} & \frac{h_1}{6} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{h_1}{6} & \frac{h_1}{3} + \frac{h_2}{3} & \frac{h_2}{6} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{h_{N-2}}{6} & \frac{h_{N-2}}{3} + \frac{h_{N-1}}{3} & \frac{h_{N-1}}{6} \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{h_{N-1}}{6} & \frac{h_{N-1}}{3} \end{bmatrix} \begin{bmatrix} y_0'' \\ y_1'' \\ y_2'' \\ \vdots \\ y_{N-1}'' \\ y_N'' \end{bmatrix} = \begin{bmatrix} f[x_0, x_1] - f'_0 \\ f[x_1, x_2] - f[x_0, x_1] \\ f[x_2, x_3] - f[x_1, x_2] \\ \vdots \\ f[x_{N-1}, x_N] - f[x_{N-2}, x_{N-1}] \\ f'_N - f[x_{N-1}, x_N] \end{bmatrix} \quad (\text{A.3-25})$$

O cálculo das derivadas de primeira ordem, assim como o valor da função no ponto desejado, é determinado pelas Equações (A.3-18) e (A.3-19) respectivamente, assim como para o caso do *spline* natural.

### A.3.3 Comparação entre o *spline* linear e o *spline* cúbico

Uma comparação interessante diz a respeito da qualidade com que os polinômios gerados por ambos os modelos de *spline* representam um conjunto de

dados. Para o conjunto de dados da Tabela 16, deseja-se obter os polinômios que descrevem cada intervalo entre os valores.

Tabela 16: Valores das variáveis independentes e dependentes para o cálculo dos *splines*.

$x$	0	1/4	1/2	3/4	1
$y$	1	2	1	0	1

Como é possível perceber, o conjunto é constituído por 5 pontos, ou seja, para os dois modelos de *splines*, haverá 4 polinômios para o representar. Inicialmente, utilizando a equação (A.3-1) para o cálculo do *spline* linear, obtêm-se os seguintes polinômios da Equação (A.3-26) divididos em seus respectivos intervalos de abrangência.

$$y(x) = \begin{cases} 1 + 4x & [0, 1/4[ \\ 3 - 4x & [1/4, 1/2[ \\ 3 - 4x & [1/2, 3/4[ \\ -3 + 4x & [3/4, 1] \end{cases} \quad (\text{A.3-26})$$

O cálculo dos polinômios pelo método do *spline* cúbico é um pouco mais complicado, primeiramente, considerou-se o caso do *spline* natural, no qual as derivadas de segunda ordem do primeiro ponto do conjunto e do último são nulas, assim, utilizando a Equação (A.3-17) e tendo como referência o sistema apresentado na matriz definida pela Equação (A.3-22), pode-se montar o sistema com as derivadas de segunda ordem do conjunto na forma matricial.

$$\begin{bmatrix} 1/6 & 1/24 & 0 \\ 1/24 & 1/6 & 1/24 \\ 0 & 1/24 & 1/6 \end{bmatrix} \cdot \begin{bmatrix} y_1'' \\ y_2'' \\ y_3'' \end{bmatrix} = \begin{bmatrix} -8 \\ 0 \\ 8 \end{bmatrix} \quad (\text{A.3-27})$$

O conjunto solução do sistema, juntamente como as condições iniciais é representado abaixo.

$$y_0'' = 0 \quad y_1'' = -48 \quad y_2'' = 0 \quad y_3'' = 48 \quad y_4'' = 0$$

Aplicando a Equação (A.3-18) para definir as derivadas de primeira ordem do conjunto, obtêm-se:

$$y_0' = 6 \quad y_1' = 0 \quad y_2' = -6 \quad y_3' = 0$$

Por fim, através da Equação (A.3-19), os polinômios de interpolação são calculados e definido pelas Equações (A.3-28) representadas abaixo junto de seus respectivos intervalos.

$$y(x) = \begin{cases} 1 + 6x - 32x^3 & [0, 1/4[ \\ 2 - 24\left(x - \frac{1}{4}\right)^2 + 32\left(x - \frac{1}{4}\right)^3 & [1/4, 1/2[ \\ 1 - 6\left(x - \frac{1}{2}\right) + 32\left(x - \frac{1}{2}\right)^3 & [1/2, 3/4[ \\ 24\left(x - \frac{3}{4}\right)^2 - 32\left(x - \frac{3}{4}\right)^3 & [3/4, 1] \end{cases}$$

(A.3-28)

Comparando os polinômios que representam o *spline* linear, Equações (A.3-26) e os que representam o *spline* cúbico, Equações (A.3-28), é possível estabelecer graficamente cada uma das interpolações dentro de seus intervalos de validade através das curvas apresentadas na Figura 35.

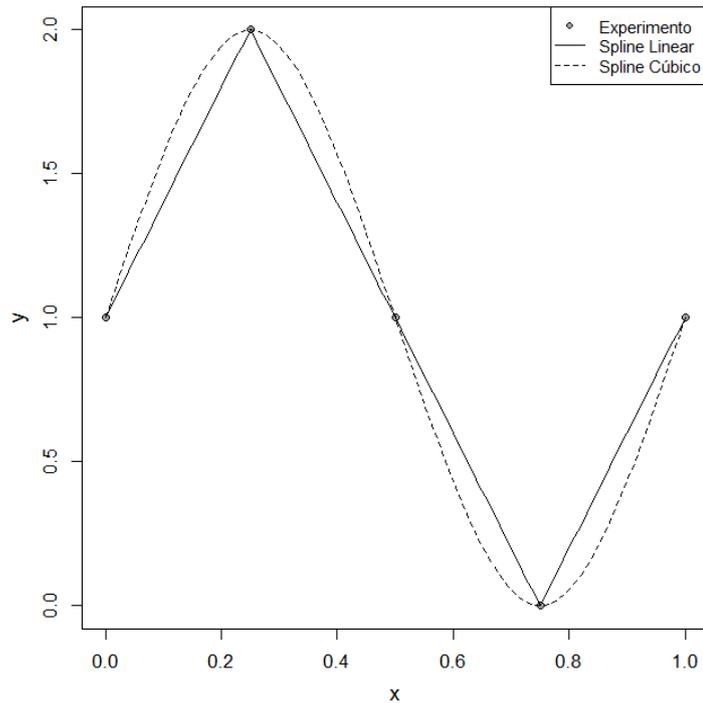


Figura 35: Comparação entre o *spline* linear e *spline* cúbico para o conjunto de dados da Tabela 16.

As curvas apresentadas na Figura 35 descrevem o conjunto de dados da Tabela 16 de diferentes formas. O cuidado que se deve ter ao utilizar os métodos, principalmente o *spline* linear, se dá pelo comportamento dos valores do conjunto. Se estes não estão distribuídos de forma linear, empregar esse *spline* pode resultar em erros na interpolação dos resultados. No cálculo das regiões de confiança pelo método do perfil, o *spline* é utilizado no cálculo dos ângulos e da fase da região a partir dos traços gerados e, como é esperado um comportamento periódico para este tipo de informação, é interessante utilizar um modelo que seja capaz de descrever as curvas e inflexões da função, aspectos que funções lineares não conseguem representar, assim, a escolha do *spline* cúbico é a mais indicada, visto que representa melhor os pontos presentes entre os valores conhecidos do conjunto.

#### *A.4 Análise da aproximação numérica da matriz Hessiana*

As vantagens da escolha do método numérico para o cálculo da matriz Hessiana da função objetivo em relação aos parâmetros foram discutidas no Capítulo 3. Uma delas está relacionada em utilizar uma única expressão para poder calcular o valor das derivadas de segunda ordem, Equações (3.6) e (3.7), para qualquer modelo, ao invés de calcular as derivadas analíticas para cada um.

O cálculo pelo método numérico necessita de uma perturbação no valor dos parâmetros para que seja possível calcular o valor da função utilizada para estimar o valor das derivadas de segunda ordem. A expressão da Equação (3.7) é proveniente da função perturbada em relação aos parâmetros expandida em uma série de Taylor truncada no termo de segunda ordem. A perturbação no valor garante que o método funcione e encontre o valor das derivadas, contudo, isso implica na dependência dos seus valores em função da perturbação, fato que pode prejudicar o ajuste da região de confiança.

Para verificar se é válido utilizar o método numérico para o cálculo da matriz Hessiana, estudou-se o comportamento das regiões de confiança elípticas quando calculadas com diferentes valores de perturbação dos parâmetros em comparação com a região gerada pela aproximação de Gauss-Newton.

As perturbações nos valores dos parâmetros para o cálculo da matriz Hessiana utilizadas foram iguais a  $10^{-7}$ ,  $10^{-8}$ , e  $10^{-9}$ , e as regiões de confiança geradas são apresentadas na Figura 36.

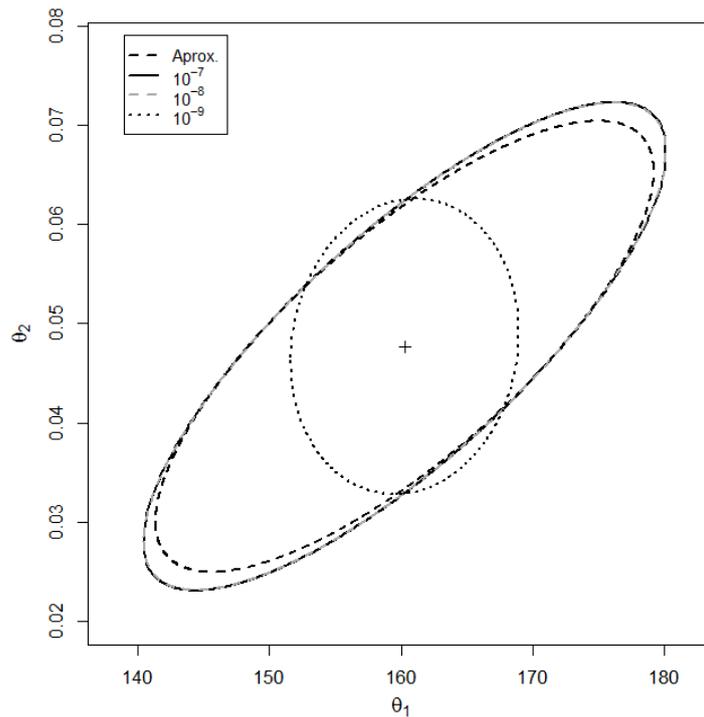


Figura 36: Regiões de confiança do Modelo 2.B com o cálculo da matriz Hessiana completa (pontos) e com o cálculo da aproximação de Gauss-Newton (linha cheia).

A região gerada com valor de perturbação igual a  $10^{-7}$  é a mesma utilizada para comparar com a aproximação de Gauss-Newton na seção anterior. Quando comparada com a perturbação igual a  $10^{-8}$ , é possível perceber que a região é idêntica a gerada com o valor  $10^{-7}$ , sendo que a mudança não alterou qualidade da estimação. Quando a perturbação é igual a  $10^{-9}$ , a região obtida se diferencia das demais, visto que ela é menor, contrariando a ideia de que o uso da matriz Hessiana expande a região e a sua geometria difere das demais, reproduzindo erroneamente a correlação entre os parâmetros e a região de confiança elíptica.

Também é possível comparar os valores numéricos da matriz Hessiana com os valores analíticos de cada uma das variáveis dependentes do modelo. A análise destes números permite observar o quanto a modificação no valor da perturbação influencia na estimação das derivadas do modelo. A matriz Hessiana analítica da Equação (A.4-1) em função das variáveis independentes e dos parâmetros é definida na Equação (A.4-1).

$$H_{\theta} = \begin{bmatrix} 0 & -\frac{x}{(\theta_2 + x)^2} \\ -\frac{x}{(\theta_2 + x)^2} & \frac{2\theta_1 x}{(\theta_2 + x)^3} \end{bmatrix} \quad (\text{A.4-1})$$

Utilizando a matriz representada pela Equação (A.4-1) para calcular o valor das derivadas de segunda ordem da Equação (3.2) com os valores de  $x$  da Tabela 3 e parâmetros ótimos  $\hat{\theta}_1$  e  $\hat{\theta}_2$  iguais a 160,28 e 0,0477 respectivamente, obtêm os resultados apresentados na Tabela 17. Estes valores são comparados com os obtidos numericamente pelas Equações (3.6) e (3.7) quando o valor da perturbação tem ordem de grandeza a  $10^{-7}$ ,  $10^{-8}$ , e  $10^{-9}$ .

Tabela 17 Comparação entre os valores da matriz Hessiana analítica e numérica do Modelo 2.B com perturbações iguais a  $10^{-7}$ ,  $10^{-8}$ , e  $10^{-9}$ .

	Analítica				$10^{-7}$			
x	H <sub>11</sub>	H <sub>12</sub>	H <sub>21</sub>	H <sub>22</sub>	H <sub>11</sub>	H <sub>12</sub>	H <sub>21</sub>	H <sub>22</sub>
0,02	0,00	-4,36	-4,36	20654,36	0,00	-4,36	-4,36	20654,53
0,02	0,00	-4,36	-4,36	20654,36	0,00	-4,36	-4,36	20654,53
0,06	0,00	-5,17	-5,17	15392,60	0,00	-5,17	-5,17	15392,68
0,06	0,00	-5,17	-5,17	15392,60	0,00	-5,17	-5,17	15392,68
0,11	0,00	-4,42	-4,42	8989,54	0,00	-4,42	-4,42	8989,58
0,11	0,00	-4,42	-4,42	8989,54	0,00	-4,42	-4,42	8989,58
0,22	0,00	-3,07	-3,07	3675,75	0,00	-3,07	-3,07	3675,76
0,22	0,00	-3,07	-3,07	3675,75	0,00	-3,07	-3,07	3675,76
0,56	0,00	-1,52	-1,52	799,86	0,00	-1,52	-1,52	799,86
0,56	0,00	-1,52	-1,52	799,86	0,00	-1,52	-1,52	799,86
1,10	0,00	-0,84	-0,84	233,24	0,00	-0,84	-0,84	233,24
	$10^{-8}$				$10^{-9}$			
x	H <sub>11</sub>	H <sub>12</sub>	H <sub>21</sub>	H <sub>22</sub>	H <sub>11</sub>	H <sub>12</sub>	H <sub>21</sub>	H <sub>22</sub>
0,02	0,00	-4,63	-4,63	20649,03	0,00	0,00	0,00	32365,29
0,02	0,00	-4,63	-4,63	20649,03	0,00	0,00	0,00	32365,29
0,06	0,01	-5,89	-5,89	15405,86	1,09	0,00	0,00	25892,23
0,06	0,01	-5,89	-5,89	15405,86	1,09	0,00	0,00	25892,23
0,11	0,01	-4,21	-4,21	8932,81	-1,09	0,00	0,00	25892,23
0,11	0,01	-4,21	-4,21	8932,81	-1,09	0,00	0,00	25892,23
0,22	0,01	-3,36	-3,36	3883,83	0,00	-168,20	-168,20	0,00
0,22	0,01	-3,36	-3,36	3883,83	0,00	-168,20	-168,20	0,00
0,56	0,00	0,00	0,00	517,84	0,00	0,00	0,00	25892,23
0,56	0,00	0,00	0,00	517,84	0,00	0,00	0,00	25892,23
1,10	0,01	0,00	0,00	517,84	0,00	0,00	0,00	25892,23

A Tabela 17 permite comparar os valores utilizados para a construção das regiões apresentadas na Figura 36. É possível reparar que os valores da matriz Hessiana analítica e numérica com perturbação igual a  $10^{-7}$  são iguais quando truncados na segunda casa decimal. Há diferença nessa casa decimal para o termo que envolvendo o parâmetro  $\theta_2$ , mas não ocorre interferência na qualidade da aproximação. Já para o cálculo numérico com aproximação igual a  $10^{-8}$ , nota-se que os valores começam a diferir mais das derivadas analíticas, principalmente para valores mais elevados da variável independente. Contudo, essa diferença não impacta na qualidade dos resultados, visto que a região de confiança obtida sobrepõe a do caso anterior. Quando a perturbação causada possui valor na ordem de  $10^{-9}$ , a

qualidade do método numérico piora, pois, as derivadas diferem significativamente dos resultados obtidos pelo método analítico, comprometendo a qualidade da região de confiança. Esta observação pode ser relacionada com a precisão do algoritmo, já que, para todos os cálculos efetuados, a precisão numérica foi mantida em  $10^{-8}$ , sendo menor que as perturbações provocadas para os casos apresentados, exceto para este último.

Ter conhecimento da precisão dos cálculos computacionais realizados é essencial para o uso de um método numérico, como o cálculo da matriz Hessiana, visto que a escolha de valores abaixo dessa precisão compromete a qualidade do método. Para o cálculo da matriz Hessiana, o método numérico pode ser substituído sem problemas pelo método analítico, visto que a maioria das funções são diferenciáveis e assim, esta situação pode ser evitada. O grande esforço no uso das derivadas analíticas se dá pela necessidade de calcular cada uma das derivadas para cada modelo utilizado. Contudo, tendo conhecimento desse aspecto dos métodos numéricos, não é necessário descartá-los, porque a qualidade da aproximação gerada representa adequadamente o valor real das derivadas.

### A.5 Pseudocódigos

Para exemplificar e facilitar o entendimento dos métodos abordados neste trabalho para delimitar as regiões de confiança de estimativas de parâmetros, foram estruturados os algoritmos utilizados para cada um na forma de pseudocódigos. O primeiro a ser demonstrado é o funcionamento do método do Contorno.

---

#### Algoritmo 1: Método do Contorno

---

Al  $\leftarrow$  passo para o cálculo do valor inicial.  
deIT  $\leftarrow$  passo do algoritmo.  
der  $\leftarrow$  derivada da função objetivo.  
DesvPad  $\leftarrow$  desvio padrão dos parâmetros.  
DT  $\leftarrow$  perturbação no ângulo.  
Fcrit  $\leftarrow$  valor da função objetivo no contorno da região de confiança.  
IC  $\leftarrow$  indicador da direção do contorno.  
Nite  $\leftarrow$  número máximo de iterações.  
Plim  $\leftarrow$  limites estabelecidos para o espaço de busca dos parâmetros.  
step  $\leftarrow$  passo para o cálculo do novo ponto.  
Theta  $\leftarrow$  ângulo da direção do contorno.  
To  $\leftarrow$  posição inicial no contorno.

#### Início

To  $\leftarrow$  0  
To(1)  $\leftarrow$  To(1) – 0.01  
Par(1)  $\leftarrow$  Param(1) + To(1)\*DesvPad(1)  
Par(1)  $\leftarrow$  **min**(Plim(1, 2), **max**(Plim(1, 1), Par(1)))  
To(1)  $\leftarrow$  - (Par(1) – Param(1))/DesvPad(1)  
**enquanto** (**abs**(Fobj/Fcrit) > 1.d-4 **ou** it < Nite) **faça** {  
    Par  $\leftarrow$  Param +To\*DesvPad  
    Fobj  $\leftarrow$  chamar o cálculo da função objetivo utilizando os valores de Par.  
    Fobj  $\leftarrow$  Fobj - Fcrit  
    it  $\leftarrow$  it + 1

```

To(1) ← To(1) +DT
Par ← Param + To*DesvPad
Fp ← chamar o cálculo da função objetivo utilizando os valores de Par.
Fp ← Fp – Fcrit
der ← (Fp – Fobj)/DT
delT ← -Fobj/der
To(1) ← To(1) – DT + delT
Al ← 1
enquanto (Par(1) > Plim(1, 1) e Par(1) < Plim(1, 2)) faça {
    Par(1) ← Param(1) + To(1)*DesvPad(1)
    Al ← Al/1.5
    To(1) ← To(1) – Al*0.5*delT
}
}
T ← To
Theta ← 90
enquanto (dist >= step – 1.d-8) então {
    Tn(1) ← T(1) + step*cos(Theta*Pi()/180)
    Tn(2) ← T(2) + step*sin(Theta*Pi()/180)
    Par ← Param + Tn*DesvPad
    IC ← 0
    para i = 1... 2 faça {
        se (Par(i) > Plim(i, 2) ou Par(i) < Plim(i, 1)) {
            Par(i) = max(Plim(i, 1), min(Plim(i, 2), Par(i)))
            T(i) ← (Par(i) – Param(i)) /Desvpad(i)
            IC ← 3 – i
        }
    }
}
se (IC /= 0) então {
    para i = 1... 2 faça {
        enquanto (abs(F/Fcrit) > 1.d-4) faça {
            Par ← Param + T*DesvPad

```

```

F ← chamar o cálculo da função objetivo utilizando os
valores de Par.
F ← F – Fcrit
T(IC) ← T(IC) + DT
Par ← Param + T*DesvPad
Fp ← chamar o cálculo da função objetivo utilizando os
valores de Par.
T(IC) ← T(IC) – DT
Fp ← Fp -Fcrit
der ← (Fp – F) /DT
passo ← - F/der
T(IC) ← T(IC) + passo
enquanto (abs(F) < abs(Fnova)) faça {
    Par ← Param + T*DesvPad
    Fnova ← chamar o cálculo da função objetivo
    utilizando os valores de Par.
    Fnova ← Fnova – Fcrit
    passo ← passo /1.5
    T(IC) ← T(IC) – 0.5*passo
}
}
TL(:, i) ← T
se (i == 1) então {
    dir ← 1.d-3
    T(IC) ← T(IC) + dir
    Par ← Param + T*DesvPad
    F ← chamar o cálculo da função objetivo utilizando os
    valores de Par.
    se (F > Fcrit) então {
        dir ← - dir
    }
    enquanto (F > Fcrit) faça {
        Fold ← F

```

```

dir ← 1.5*dir
T(IC) ← T(IC) + dir
Par ← Param + T*DesvPad
F ← chamar o cálculo da função objetivo utilizando
os valores de Par
}
T(IC) ← T(IC) – dir*((F-Fcrit)/(F – Fold))**2
}
}
Theta ← Theta +180
}
se não {
enquanto (abs(Fnova /Fcrit) > 1.d-5) faça {
Tn(1) ← T(1) + step*cos(Theta*Pi()/180)
Tn(2) ← T(2) + step*sin(Theta*Pi()/180)
Par ← Param +Tn*DesvPad
F ← chamar o cálculo da função objetivo utilizando os valores de Par.
F ← F – Fcrit
Tn(1) ← T(1) + step*cos((Theta + DT)*Pi()/180)
Tn(2) ← T(2) + step*sin((Theta + DT)*Pi()/180)
Par ← Param + Tn*DesvPad
Fp ← chamar o cálculo da função objetivo utilizado os valores de Par.
der ← Fp – Fcrit
Tn(1) ← T(1) + step*cos((Theta - DT)*Pi()/180)
Tn(2) ← T(2) + step*sin((Theta - DT)*Pi()/180)
Par ← Param + Tn*DesvPad
Fp ← chamar o cálculo da função objetivo utilizando os valores de Par.
Fp ← Fp – Fcrit
der ← (der – Fp) / (2*DT)
passo ← - F/der
Theta ← Theta + passo
enquanto (abs(F) > abs (Fnova)) faça {
Tn(1) ← T(1) + step*cos(Theta*Pi()/180)

```

```

    Tn(2) ← T(2) + step*sin(Theta*Pi()/180)
    Par ← Param + Tn*DesvPad
    Fnova ← chamar o cálculo da função objetivo utilizando os
    valores de Par.
    passo ← passo/1.5
    Theta ← Theta – 0.5*passo
  }
}
T(1) ← T(1) + step*cos(Theta*Pi()/180)
T(2) ← T(2) + step*sin(Theta*Pi()/180)
Dist ← sqrt(sum((To-T) **2))
dif ← abs(Theta – Theta0)
se (dif < 0.2) então {
  step ← step*2
  step ← min(step, 0.1)
}
se (dif > 0,5) então {
  step ← step/10
  step ← max(step, 5.d-3)
}
Theta0 ← Theta
}
}
Fim

```

O segundo pseudocódigo apresentado é o do método do perfil. Este foi elaborado interpretando os códigos disponíveis nos Apêndices 3 e 6 da obra de BATES; WATTS, (1988), para calcular o perfil dos parâmetros e definir o esboço das regiões de confiança. O cálculo da minimização da função objetivo do perfil foi estruturado adaptando o valor da derivada do parâmetro escolhido no vetor gradiente da função objetivo no método de Gauss-Newton. A rotina para o cálculo dos *splines* foi obtido nas rotinas disponíveis em PRESS *et al.*, (1996).

---

### Algoritmo 2: Método do Perfil

---

Ind ← vetor com o pares de parâmetros que serão utilizados para a confecção da região de confinaça.

DesvPad ← desvio padrão dos parâmetros.

NP ← número de parâmetros.

Nite ← número de iterações.

Param ← vetor com os valores dos parâmetros no ponto ótimo.

ParSel ← vetor com o índice de quais parâmetros são estimáveis, 0 para não estimável e 1 para estimável.

Tol ← tolerância para o término do cálculo.

Var\_antes ← salva o penúltimo ponto do perfil, utilizado para o cálculo do ponto limite pelo método da bissecção.

Var\_depois ← salva o último ponto do perfil, utilizado para o cálculo do ponto limite pelo método da bissecção.

#### Início

ParSel ← 1

Definir Nite

Definir passo

Var\_antes ← 0

Var\_depois ← 0

Fobj ← chamar a função para o cálculo da função objetivo utilizado o valor de Param.

s ← **sqrt** (Fobj/ (NE- NP))

Definir vetor Tau para salvar os valores das reparametrizações calculados

Definir matriz M para salvar o valor dos parâmetros

**para** i = 1... NP **faça** {

Par ← Param

ParSel(i) ← 0

Delta ← DesvPad(i)/ passo

M(i, 0) ← Par

Tau\_save(i, 0) ← 0

Fobj\_save(i, 0) ← 0

**para** j = 1... 2 **faça** {

Par ← Param

t ← 0

delta = -delta

invslope ← 1

**para** k = 1... Nite **faça** {

t ← t + invslope

Par(i) ← Param(i) + delta\*t

Fobj\_tau ← chamar a minimização pelo método Gauss-Newton adaptado para o perfil com os valores de Par.

sinal ← **sign**(Par(i) – Param(i))

Tau ← sinal\***sqrt**((Fobj\_tau -Fobj)/s)

Var\_antes ← Var\_depois

Var\_depois(1:NP) ← Par

Var\_depois(NP+1) ← Tau

**se** ((**abs**(Tau > Tau\_Max)) **então** {

**sair**

}

M(i,k) ← Par

Tau\_save(i, k) ← Tau

Fobj\_save(i, k) ← Fobj\_tau

invslope ← **abs**((Tau\*(desvpad(i))^2)/(2\*s))

invslope ← **min**(4, **max**(invslope, 1/16))

}

```

enquanto (abs(Tau_bis – Tau_Max) < Tol) faça {
    Par_bis ← (Var_depois(1:NP) + Var_antes(1:NP))/2
    Fobj_tau ← chamar a minimização pelo método Gauss-Newton
    para o perfil com os valores de Par.
    Tau_bis ← sinal*sqrt ((Fobj_tau -Fobj) /s)
    se (abs(Tau_bis – Tau_Max)) > 0 então {
        Var_depois(1:NP) ← Par_bis
        Var_depois(NP+1) ← Tau_bis
    }
    se não {
        Var_antes(1:NP) ← Par_bis
        Var_antes(NP+1) ← Tau_bis
    }
}
M(i, k) ← Par_bis
Tau_save(i, k) ← Tau_bis
    Fobj_save(i,k) ← Fobj_tau
}
ParSel(i) ← 1
}
para k= 1... NP-1 faça {
    para l = k+1... NP faça {
        Ind(1) ← k
        Ind(2) ← l
        para i = 1... NP faça {
            M1(:,1) ← M(i, :, Ind(1))
            M1(:,2) ← M(i, :, Ind(2))
            M2(:, 1) ← M(j, :, Ind(1))
            M2(:, 2) ← M(j, :, Ind(2))
        }
        para i = 1... 2 faça {
            para j = 1... 2 faça {
                se (i == j) então {

```

```

        pula
    }
    se (i == 1) então {
        para k = 1... size(Tau_2) faça {
            se (k > size(g12) ou k > size(Tau_1)) então
            {
                sair
            }
            g12(k) ← chamar função para o cálculo do
            spline cúbico Tau_1(M1(:, i), M2(k, i))
            g12(k) ← acos(g12(k))
        }
    }
    se não {
        para k = 1... size(Tau_1) faça {
            se (k > size(g21) ou k > size(Tau_2)) sair
            g21(k) ← chamar função para o cálculo do
            spline cúbico Tau_2(M2(:, i), M1(k, i))
            g21(k) ← acos(g21(k))
        }
    }
}
k_lim ← 1
ParN(1, 1) ← 0
ParN(2, 1) ← Pi()
ParN(3, 2) ← 0
ParN(4, 2) ← Pi()
para i = 1... 4 faça {
    selecionar caso(i) {
        caso(1){
            ParN(i, 2) ← chamar função para o cálculo do
            spline cúbico g12(Tau_1, k_lim)

```

```

    }
    caso(2){
        ParN(i ,2) ← chamar função para o cálculo do
        spline cúbico g12(Tau_1,-k_lim)
    }
    caso(3){
        ParN(i ,1) ← chamar função para o cálculo do
        spline cúbico g21(Tau_2, k_lim)
    }
    caso(4){
        ParN(i ,1) ← chamar função para o cálculo do
        spline cúbico g21(Tau_2, -k_lim)
    }
}
}
para i = 1... 4 faça {
    angulo(i) ← soma(ParN(i, :))/2
    fase(i) ← ParN(i, 1) – ParN(i, 2)
    se (fase(i) < 0) então {
        angulo(i) ← - angulo(i)
        fase(i) ← - fase(i)
    }
}
}
Definir Nite
para i =1... Nite faça {
    an ← ((i-1)*2*Pi())/(Nite-1) – Pi()
    dn ← chamar função para o cálculo do spline cúbico
    fase(angulo, an)
    T1 ← cos(an + dn/2)
    T2 ← cos(an – dn/2)
    Par_1 ← chamar função para o cálculo do spline cúbico M1(:,
    1)(Tau_1, T1*TauMax)
}

```

```
Par_2 ← chamar função para o cálculo do spline cúbico M2(:,  
2)(Tau_2, T2*TauMax)
```

```
}
```

```
}
```

```
}
```

```
Fim
```

O terceiro pseudocódigo é o do método *Bootstrap*. Para este cálculo é necessário a geração de números aleatórios e os valores das incertezas experimentais.

---

**Algoritmo 3: Método *Bootstrap***

---

EVY ← matriz de incertezas experimentais.

Niter ← número de iterações.

NP ← número de parâmetros.

NVsai ← número de variáveis dependes experimentais.

YE ← valor das variáveis dependentes experimentais.

**Início**

chamar a semente aleatória ( )

YER ← YE

**para** k = 1... Nite **faça** {

    rnd ← número aleatório

**para** i = 1... NP **faça** {

**for** j = 1...NVsai **faça** {

            erro ← inverso da distribuição cumulativa normal de rnd(i, j)

            erro ← erro\*sqrt(EVY(i, j, j))

            YE(i, j) ← YER(i, j) + erro

        }

    }

    Par ← chamar o cálculo da minimização da função objetivo utilizado os valores de YE

}

**Fim**

A.6 *Artigo*

**Statistical evaluation of prediction uncertainty of models used to describe COVID-19 behavior**

Nícolás Artifon Dorneles<sup>1</sup>, Marcio Schwaab<sup>1\*</sup>

1 - Departamento de Engenharia Química, Escola de Engenharia, Universidade Federal do Rio Grande do Sul. Rua Ramiro Barcelos, 2777 - Prédio 22202, Porto Alegre, RS, 90035-007, Brasil.

\* Corresponding author: [schwaab@ufrgs.br](mailto:schwaab@ufrgs.br)

## Statistical evaluation of prediction uncertainty of models used to describe COVID-19 behavior

### Abstract

With the outbreak of the Coronavirus Disease, it has grown the number of studies that try to estimate the end of the period of contagious. To achieve this goal, most of them uses Gaussian models, Susceptible-Infectious-Recovered models and their derivatives. Nevertheless, it is essential to choose a data set with enough information, so it will be able to describe more precisely the way that virus spreads over a country. In this way, to estimate the numbers of the days in which the curve of infected people achieves 95% of the cases, it was used the same data set with parts of the information of two countries: Italy and Austria. Obtaining only one answer value it is not an accurate procedure, and, by this way, it was established its statistical confidence interval based on the concept of the confidence region, that is more accurate than an interval. It was concluded that, to obtain an accurate estimation, it is needed at least 50% of the information or have at least the data near the inflection peak. Other important fact observed was that SIR models got a better fit in the problem scales, due its asymmetrical shape.

**Keywords:** confidence regions; COVID-19; estimation; Gaussian model; SIR model; model prediction uncertainty.

## 1. INTRODUCTION

Since the beginning of the Coronavirus Disease, COVID - 19, pandemic contagion, it has been urgently necessary a drastically change in the way of live of people from different countries. In the hope of lower the spread its contamination it has been needed close shops and industries, slowing down economies e cutting of jobs, and the it was important to us adopted news habits of live, as taking more care of our sanitation and avoiding interact physically with others people.

The first case of the virus contamination was observed in December 31<sup>st</sup> of 2019 in Wuhan, China, and, since that, travelers have been carrying it for others countries. The second one to register case of the virus was Italy and, after that, it has been spreading all over Europe, especially in The United Kingdom, in Spain and in France. Not a long time after that, the virus reached America, making a massive contamination in The United States and in Brazil, both being the two countries with the largest number of cases and deaths until this moment.

Due the fact of so many countries being infected in a little interval of time, the number of researches evolving subjects with respect to the virus has grown, as the ones that tries to estimate the end of the period of contagious. There are plenty of models used to simulate the behavior of an epidemical disease, such as Gaussian models, Susceptible-Infectious-Recovered (SIR) models, among many others.

Gaussian models are a classical method in this situation and have its mathematical proprieties well known. Some papers use these models to estimates the end of cases in a country based on the number of people infected or dead. Barmparis and Tsironis (Barmparis and Tsironis, 2020) estimated the last period of the infection for some places using this method with no problems, nevertheless it was not capable to fit all data. Schüttler et al. (Schüttler et al., 2020) predicted the number of infected and dead using some data from places where the pandemic had just began, what can lead to a large uncertainty in the prediction. Woody et al. (Woody et al., 2020) considered a logarithmical Gaussian model, that approaches to a quadratic function, what it is not advisable, due the fact that the error laying in the data will not follows the normal distribution after the procedure.

Other model, as mentioned before, is the SIR, which consists in a system of three differential equations. Some Works (Luo, 2020; Postnikov, 2020; You et al.,

2020) used this model to predict the end of the plague in some locations. Other works (Fernández-Villaverde and Jones, 2020; Yang et al., 2020), used some SIR variant models to achieve the same goal. The gain of choosing this phenological model is its capability of bending the geometry of a bell-shaped curve to one that represents with more accuracy an epidemiologically decrease, since the sine curve for a disease is usually not symmetrical as predicted by the Gaussian models.

Both models are used to predict the end of pandemics and it is always possible to find discussions to elect the best one. However, when this prediction is made with only a small amount of the data, in an attempt to determine the end of pandemic with few days of contagious, the predictions are very uncertain and must be prudently analyzed. In this way, it is fundamental to estimate the prediction uncertainty, which could be used to show how reliable are the results.

The main goal of this paper is to study the time in which the curve of infected people achieves 95% of the cases based in different sets, 10%, 30%, 50%, 70% and 100%, of the same data. This way, it will be possible to verify if the estimation of end of the pandemic are reliable when a small portion of the curve is used for predictions. Both Gaussian and SIR models were considered; its estimated parameter and predicted values (in particular the time to infect 95% of the infectable people) were evaluated also considering their respective uncertainties.

## **2. FUNDAMENTALS**

### **2.1. Gaussian Model**

A Gaussian model is a continuous symmetrical function that follows the normal distribution. This type of model is always studied when the function presents a symmetrical peak, like in chromatography estimations (Baeza-Baeza et al., 2011). Other gain on choosing this models is that its mathematical properties are well known, so it is possible to calculate analytically any desired information from the function curve (Schwaab and Pinto, 2007). A Gaussian model, used to estimate the number of infected people per day, follows Eq. (2). This equation is defined by three parameters:  $N$  is the number in which there were more infected people per day, corresponding to the maximum value of the curve;  $\mu$  is the date on which the maximum number of the

cases per day was achieved; and  $\sigma$  represents period of time that the pandemics affects the place, corresponding to the width of the curve.

$$i(t) = N \cdot \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) \quad (2)$$

As mention before, all mathematical proprieties of a Gaussian model are well known, so, having a starting set of data, it is possible, after the parameter estimation, to determinate the period of time that the curve achieve 95% of its area,  $t_{95\%}$ , indicating its end. Eq. (3) represents the value of time when 95% of the curve is achieve and  $u_{95\%}$  corresponds to the value of the inverse of the normal probabilistic distribution with a probability of 95%. This value is equal to 1,644.

$$t_{95\%} = u_{95\%} \cdot \sigma + \mu \quad (3)$$

## 2.2. SIR Model

The Susceptible–Infected–Recovered model is a system of three epidemiological models first proposed by (Kermack and McKendrick, 1927). It is the most popular model to represent an epidemic decease, (Postnikov, 2020), and is represented by the following equations, Eq. (4). The SIR system has two parameters:  $\beta$  is the transmission rate through exposure and  $\gamma$  is the rate of recovering, also it inverses represents the average infectious time. The variables  $s$ ,  $i$  and  $r$  represents, respectively, the number of susceptible, infected and recovered and dead, removed, cases.

$$\begin{aligned} \frac{ds(t)}{dt} &= -\beta \cdot s(t) \cdot i(t) \\ \frac{di(t)}{dt} &= \beta \cdot s(t) \cdot i(t) - \gamma \cdot i(t) \\ \frac{dr(t)}{dt} &= \gamma \cdot i(t) \end{aligned} \quad (4)$$

This equation is similar to a Lotka-Volterra type equations which are used to explain the interaction between two biological systems, as the relationship predators-preys, (Harko et al., 2014), also it could be associated to the kinetics of a chemical reaction, (Postnikov, 2020). This system has not a simple solution, and, for many year, only numerical procedures were able to resolve it. Recently (Harko et al., 2014) establish an analytical solution to this model, that needs the use of an auxiliary variable to resume the problem to a Bernoulli type differential equation. Nevertheless, the solution of this equation leads to an integral that has not an analytical solution, so, the numerical procedures are needed to resolve this part of the problem. In this way, it is much simpler to solve the differential equation system through numerical methods.

Other variants of this model, as the SEIR model (E stands for Exposed), that considers a part of the population that are infected, but asymptomatic, (Yang et al., 2020), and the SIRD, that distinguishes the recovered and deceased cases [9]. The main problem with the SEIR model is that the number of asymptomatic cases is hard to obtain, due the fact that is needed a massive testing of the population, which many countries are not able to perform. The SIRD model is more practical than the least one, because the number of deceased people is a common data, however, the principal equation of this system it is of the infected cases, so, in this situation, there are not much difference in choosing the SIR or the SIRD model. It is also possible to find in the available literature equations to a lot of different epidemiological situations.

### **2.3. Parameter Estimation**

The parameter estimation techniques consist of methodologies to determine these unknown values, through the approximation of the real data with that provided by the model. To determine the best set of parameters that describes a system, the initial step is to validate the model that will be used, that is, to ensure that it is capable of representing mathematically the experimental data of the problem. (Benyahia et al., 2013). For this purpose, the best values of parameters are always sought, which reduce the difference between the variables measured experimentally and those determined by the model.

The approximation between experimental values and those obtained by the model is guaranteed through the use of a specific class of functions called objective

function, (Almeida et al., 2002), which aims to reduce the existing deviation between the calculated data and those measured through successive variations in parameter values. The most common and practical objective function is the least square function, Eq. (5), that consists in the quadratic deviation between the measured,  $y_i^m$ , and predicted data,  $y_i^p$ , using the estimated parameters' values,  $\theta$ , for each experiment,  $NE$ . The basic assurances established to validate this function are assumes that the error of the measured values follows the normal distribution, there are not correlation between the experiments and the error of the dependent variables is constant. (Schwaab and Pinto, 2007).

$$S(\theta) = \sum_{i=1}^{NE} \left( y_i^m - y_i^p(x_i^p, \theta) \right)^2 \quad (5)$$

After determining the objective function, the method by which it will be minimized is chosen, (Ruder, 2016), and methods such as analytical or heuristics can be used. The basic mechanism for optimizing an objective function involves iterations that modify the values of the parameters and ensure that the algorithm scans the dimensions of the function, ensuring that values which satisfy the minimum condition are found. In this work, minimization of Eq. (4) was performed with the aid of a hybrid optimization method (Schwaab et al., 2010, 2008), where the initial minimization phase was performed with the Particle Swarm Optimization (PSO) algorithm (Kennedy and Eberhart, 1995), followed by a second estimation round where the best estimate obtained in the previous phase was used as the initial guess for the Gauss-Newton method (Bard, 1974; Schwaab and Pinto, 2007). Convergence was achieved when the relative modification of the objective function was smaller than  $10^{-6}$ .

## 2.4. Confidence Regions

The confidence region is a space that delimits all sets of parameter values that, when replaced in the model, still guarantee the quality of the adjustment to the experimental data, (Benyahia et al., 2013), allowing us to assume that the model used is being well represented by the objective function. The most common and usual

method to determine the region of confidence of the parameters is by quadratic approximation of the objective function around the minimum point, Eq. (5), in this way, it is possible to rewrite the objective function as a Taylor series truncated in the second term.

$$S(\theta) = S(\hat{\theta}) + (\theta - \hat{\theta})^T \nabla S_{\theta} + \frac{1}{2} (\theta - \hat{\theta})^T H_{\theta} (\theta - \hat{\theta}) \quad (5)$$

As the approximation is performed at the minimum point, the gradient of the objective function is null and the Hessian matrix is related to the covariance matrix of the parameters, and Eq. (6) is obtained. (Kang and Bates, 1990).

$$S(\theta) - S(\hat{\theta}) = (\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) \quad (6)$$

Assuming that the deviations between the values obtained experimentally and those predicted by the model are normally distributed, and as these normal deviations are summed and squared in the calculation of the objective function, it is correct to state that the right-hand side of the Eq. (6) is described by the chi square distribution,  $\chi^2$ , with the number of parameters,  $NP$ , as degrees of freedom and confidence level equal to  $1 - \alpha$ , Eq. (7). (Cui et al., 2015).

$$(\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) \leq \chi^2(NP, \alpha) \quad (7)$$

The equation considers that the covariance matrix of the parameters is known with precision, which, almost always is not true, since a large number of observations of the experiments would be necessary. Thus, dividing it by the objective function and weighting the parts on the left-hand side of the equation by their respective degrees of freedom, Eq.(8) the chi-square distribution can be written as an F-Fisher distribution with confidence level equal to  $1 - \alpha$ . (Yoo et al., 2017).

$$\frac{(\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) / NP}{S(\hat{\theta}) / NE - NP} \leq F(NP, NE - NP, \alpha) \quad (8)$$

Considering the algebraic treatment demonstrated by (Donaldson and Schnabel, 1987), the confidence region can be defined like Eq. (9), as a function of the minimum point parameters, the covariance matrix of the parameters, the number of parameters and the F-Fisher distribution with  $NP$  and  $NE-NP$  degrees of freedom and confidence level equal to  $1 - \alpha$ .

$$(\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) \leq NP.F(NP, NE - NP, \alpha) \quad (9)$$

It can be pointed (Donaldson and Schnabel, 1987), that due to the assumed approximations and considerations, Eq. (9) is only valid for linear models, since the deviation between the model variables and the measures experimentally follows the normal distribution and the Hessian matrix and the Gaussian approximation is valid. For non-linear models, the quality of the approximation of the confidence region is not exact and, therefore, will depend on the linear approximation of the model. In this way, the results can lead to wrong predictions about the parameters and the region of confidence. (Schwaab et al., 2008).

Similarly, we can perform the same procedures for the left-hand side of Eq. (6), obtaining the equation for the Likelihood Ratio region, Eq. (10), in which  $s$  represents an approximation of the experimental error. (Donaldson and Schnabel, 1987).

$$S(\theta) - S(\hat{\theta}) \leq s^2.NP.F(NP, NE - NP, \alpha) \quad (10)$$

The Particle Swarm algorithm can also be used to define the region of confidence of the parameters. Taking advantage of the particles generated during the function minimization stage and choosing only those that have the objective function value less than the region's limit, it is possible to use them to delimit the internal space of the confidence region. (Koduru et al., 2007; Schwaab et al., 2008).

The efficiency of the use of the Particle Swarm to define the regions of confidence of the parameters will depend on the choice of values of weight of inertia, which is responsible for conditioning the convergence speed of the algorithm. (Shi and Eberhart, 1998). Thus, the best alternative for the region of confidence to be well defined is to ensure that the value of the weight of inertia can be changed during the iterations of the particles, ensuring that the region is fully explored and converges without difficulties. (Kessentini and Barchiesi, 2015).

### **3. METHODOLOGY**

#### **3.1. Data Set**

The data used in this work were obtained from the Health Ministry, or the local equivalent to that, of each country researched. The main idea was collecting enough data to describe a peak situation. The data that continue after the peak descendence was discarded, because in many situations, there were a second pandemic contagious which could be not be described with these models without some adaptations.

The countries chosen were Italy and Austria. Italy data were observed from February 24<sup>th</sup> to June 26<sup>th</sup>, (Salute, 2020), and Austria data were observed from February 25<sup>th</sup> to June 24<sup>th</sup>. (Konsumentenschutz, 2020).

#### **3.2. Estimation Procedures**

The two models presented in this work, Gaussian and SIR, differs on then mathematical characteristics. The first one is a well-known problem with one equation and has it analytical proprieties developed. On the other hand, the second one is a system of three differential equations with a complex analytical solution, that involves a numerical integral. Although, in the parameter estimation procedure, it is feasible to point some common settings. Both equations had its least squares objective function minimized by a hybrid Particle Swarm and Gauss-Newton optimization algorithm. The estimations were performed for different data sets with 10%, 30%, 50%, 70% and 100% of the available information and, after that, the value of time to achieve 95% of

the curve of infected cases and the confidence regions using the points generated by the Particle Swarm were obtained.

For the Gaussian model, obtain  $t_{95\%}$  value was simpler, due the due the fact that there is a mathematical expression to calculate it. After the estimation of the three optimum parameters values of Eq. (2), we used Eq. (3) to obtain  $t_{95\%}$  value for each data set. To determinate the uncertainty of  $t_{95\%}$  it was used this same equation with the parameter's values generate by the swarm particles. This allowed to build a confidence interval for  $t_{95\%}$ . To demonstrate the statistical uncertainty of this value, its variance was calculated according Eq. (11):

$$Var(t_{95\%}) = u_{95\%}^2 \cdot Var(\sigma) + Var(\mu) + 2u_{95\%} \cdot Cov(\mu, \sigma) \quad (11)$$

where  $Var$  represents the variables' variance and  $Cov$  represents the covariance matrix.

The SIR model estimation was not so simple as its analytical solution is not so simple. To avoid using an auxiliary variable, the solution of three differential equations of this model was calculated numerical with a Runge-Kutta method, RK4. To estimate the value of  $t_{95\%}$ , it was implemented a numerical integral procedure using Simpson's rule to determinate in which point the area of the curve was bigger than 95% of the total area. After that, an interpolation algorithm was used to calculate the value of  $t_{95\%}$ . Results uncertain was calculated with a similar expression to the one used by (Bard, 1974) to define the covariance matrix, Eq. (12).

$$Var(t_{95\%}) = \begin{bmatrix} \frac{\partial t_{95\%}}{\partial \beta} \\ \frac{\partial t_{95\%}}{\partial \gamma} \end{bmatrix}^T V_{\theta} \begin{bmatrix} \frac{\partial t_{95\%}}{\partial \beta} \\ \frac{\partial t_{95\%}}{\partial \gamma} \end{bmatrix} \quad (12)$$

To compare with the Gaussian results,  $t_{95\%}$ . value was obtained from the curve of infected cases.

All simulations performed in this paper were written and implemented in a Fortran's 90 compiler and the figures were generated with R Project's ggplot2 package.

## 4. RESULTS AND DISCUSSION

The values for  $t_{95\%}$  of the original data were calculated using the same approach used to SIR model. The values expected were equal to 108 days for Italy and 86 days for Austria.

### 4.1. Gaussian Model

The results for parameter estimation to different data set of Italy and Austria infected cases are presented at Table 1. The curves presented in Figure 1, Figure 2, Figure 3 and Figure 4 show the comparison of the original samples (blue bars) and the model prediction for its values (gray curve) split in the different data sets. The dates in the x-axis respectively correspond to the first day and the ones representing 10%, 30%, 50%, 70% of the data set and the last information it is the point where 95% of the cases were achieved. The point where there is a change of color on the gray curves represents the value of  $t_{95\%}$  estimated with its results.

First for Italy's cases, it is possible to compare the scale of the objective function, which grows as the more data was used on the estimation. The large number of sample available and the fact that it is difficult to a model fit all data perfectly contributes to higher the objective function's value. From Figure 1, it is possible to evaluate how the different data set estimation response fits the original data. The first idea that it is possible to argue from the graphics is that, for small amounts of the original data, the model fits perfectly them, which is associated to the low values of the objective function obtained for 10% and 30% sets. Although, the remaining values are not correctly represented, due the fact that this sets do not significantly represent the entire information. With 50% of the information there is an improvement in the quality of the estimation, and for 70% and 100% it is possible to obtain a model that fits well the data's scale and tries to represent in the best way all the information. It is also remarkable a problem of the Gaussian models for its situation; generally, the infectious curves are not symmetrical, so this models would not be capable of fit all data if its end it is not similar to it beginning, as we can see in the 70% and 100% curves.

The evaluation of the values of  $t_{95\%}$  for each estimation is presented in Table 1 and can be observed in Figure 1. From the figure, it can be seen that no one of the

estimation match the value obtained with the original data, which is equal to 108 days. The reliability of this value may be observed from its variance, which is expected a low value, allowing to claim that this value has small an error. This affirmation is false only for the case with 10% of the data set, for the other ones,  $t_{95\%}$  results can be trusted. According to the values, the biggest set of data lead to a best estimation of  $t_{95\%}$  value, due the fact that there is a best comprehension of the total curve's extension. Using the parameter's values inside the confidence region and Eq. (3) it was possible to obtain an interval of values that statistically represents the  $t_{95\%}$ . The best situation is an interval that comprehends  $t_{95\%}$  value and, no one of them do that, except for the case with 100% of the data, that miss for one day.

The Austria's cases estimation results are also showed at Table 1 From it is possible to observe a similar situation to the Italy case, in which the value of the objective function grows with the number of samples analyzed, due the fact that there are more data to fit, so the model tries to adequate to all of them in the best way.

The estimated curves for Austria can be observed at Figure 2, in which the data set with initial information, 10%, lead to results that do not represent the information. As the samples number are increased, the predicted results adjust better to the original information and, as it can be seen, one more time the Gaussian model cannot fit all data that due their asymmetrical distribution. The values of  $t_{95\%}$  represented at this figure and in Table 1 indicates that the prediction of the pandemic end for all data sets do not correspond to the observed one, that is equal to 86 days. From its variance results, one more time the value from 10% of the samples cannot be trusted, due their high scale; for the others one, the deviance associated is acceptable. The interval of  $t_{95\%}$  values obtained with the confidence regions points, Table 1, allows to conclude that no one of the estimated ends is correct, due the fact that its values do not comprehend  $t_{95\%}$  of the original data. There is a similar case, for 10 % of the data, that misses for one day, however it is not possible to trust in it, due its high variance, so this result is discarded.

## 4.2. SIR Model

The results for the SIR model were obtained considering the three differential equations of Eq. (4), but to resume the information, it will be only showed the curve of infected cases.

In Table 2 are presented the results obtained from the estimation of the SIR model. Firstly, considering the Italy's results, the evaluation of the objective function's values demonstrates that it is more difficult to fit largest data sets, like it was seen for the Gaussian models, because the increase of points makes the estimation harder. The value of the objective function for 10% differs significantly from the one for 30%, due to the fact that the estimation of this model also considers the number of susceptible and recovered cases, which raises the evaluation of the objective function three times.

From Figure 3 it is possible to compare the SIR model predictions for all Italy's data sets. The curves with 10% and 30% of the initial samples fit well the curve's beginning, but they are not capable of suit the remaining data, so their estimation is not reliable. With 50% of the data it is possible to get a better fit of all set and it gets better with more data, as it is possible to be seen in the curves with 70% and 100%, which suit well the points after the peak. Other interesting aspect is that all SIR curves have their peak scale near to the one from the real data, even for the initial cases, which is a behavior that was not observed in the Gaussian model, whose scale differs drastically from that one.

The evaluation of  $t_{95\%}$  values from Figure 3 and from Table 2 allows to determine that neither of the data sets were able to predict the pandemic's end. The values obtained with 10% and 30% of the data set have a high variance value, so they cannot be trusted. The others result differ from the reference value, that is equal to 108 days.

Austria's estimation results are described at Table 2 and, as expected, the value of the objective function increases with the number of samples on the data set. The scale of the function's value also highly increases from the first to the second data set, due to the number of dependable variables of the problem.

The predicted curves for the infected cases using different data sets for Austria's cases are exhibited at Figure 4. For 10% of the data, the curve aspect tends to run away from the original data standard behavior, due to the fact that it has a low rate of recovering, which means that the infected case will take more time to change their status. The data set with 30% of the data gets a better fit than the previous one, but it also

does not describe the correctly the descendent part of the curve. For the cases with 50%, 70% and 100% of the data, the fit is similar, only getting an improvement when the number of samples is increased.

The estimation of  $t_{95\%}$  value remarkable wrong for the case with 10% of the data, due it high value and its variance. The second curve, 30%, also does not predict this value correctly, with is equal to 86 days. On the other side, the estimation with 50%, 70% and 100% were able to predict the exact number of days to end the pandemic.

As a general comparison of the results obtained, independently of the model, for a, at least, medium quality estimation, it is necessary 50% or more information of the data. With this amount, the predicted values were capable of get distributed in the problem scale. A factor that may help the fit consists in get information with regions near the peak, especially after the inflection point, so, the models will tend to descend the curve.

Other important fact observed was that the curves of SIR models got better distributed at the problem scales, even for low data sets. This can be explained by its capacity of describe an asymmetrical curve, so its parts before and after the peak can shape these regions more realistic, unlike the Gaussian model, that needs to clone the second half due its symmetrical behavior.

## **5. CONCLUSION**

In this work, the prediction of the end time of the COVID-19 pandemic was studied using different portions of the same data set with the numbers of people infected in two countries. For this, the parameters of the Gaussian and the SIR models were estimated and, using the predicted values, an estimate of  $t_{95\%}$  was calculated, which represents the number of days for 95% of infected cases to have occurred; this value was compared with that obtained by the original data. In general, the estimates need at least 50% of the data to be reliable, since the estimates with 10% tend to be very inaccurate and those with 30% do not comprise the complete set. From 50% onwards, the curves have already reached the summit and those are close to it, so the forecasts become more accurate. Estimations with 70% or 100% of the data are

similar, so at least 70% of the data is recommended so that you can have a good representation of the infection curve.

For the prediction of the end of the pandemic the results indicate that the more data, the better the prediction of the end, although they differ in which is the best model. For the data from Italy, the Gaussian model with 100% of the data was the one that best suited the data and approached the forecast of  $t_{95\%}$  with the help of the statistical interval obtained with the confidence regions. As for the data from Austria, the SIR model was the one that best predicted the end, being able to determine it with 50% of the data. The choice of which model is better is a little more suggestive. Although the SIR model should be preferred due to its ability to better represent the asymmetries existing in pandemic curves, the use of Gaussian models cannot be completely ruled out, since it has adapted better to the data set in Italy. Problems of this type end up depending on other factors that are beyond our control, such as social isolation, climate, political situation, so, there is no single model that best suits all regions or countries; the best option is to always compare the results to determine which one best represents the problem.

## REFERENCES

- Almeida, A.M., Castel-Branco, M.M., Falcão, A.C., 2002. Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 774, 215–222. [https://doi.org/10.1016/S1570-0232\(02\)00244-1](https://doi.org/10.1016/S1570-0232(02)00244-1)
- Baeza-Baeza, J.J., Torres-Lapasió, J.R., García-Álvarez-Coque, M.C., 2011. Approaches to estimate the time and height at the peak maximum in liquid chromatography based on a modified Gaussian model. *J. Chromatogr. A* 1218, 1385–1392. <https://doi.org/10.1016/j.chroma.2011.01.029>
- Bard, Y., 1974. *Nonlinear Parameter Estimation*. Academic Press, New York.
- Barmparis, G.D., Tsironis, G.P., 2020. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos, Solitons and Fractals* 135. <https://doi.org/10.1016/j.chaos.2020.109842>
- Benyahia, B., Latifi, M.A., Fonteix, C., Pla, F., 2013. Emulsion copolymerization of styrene and butyl acrylate in the presence of a chain transfer agent. Part 2: Parameters estimability and confidence regions. *Chem. Eng. Sci.* 90, 110–118. <https://doi.org/10.1016/j.ces.2012.12.013>
- Cui, H., Yan, G., Song, H., 2015. A novel curvelet thresholding denoising method based on chi-squared distribution. *Signal, Image Video Process.* 9, 491–498. <https://doi.org/10.1007/s11760-014-0685-9>
- Donaldson, J.R., Schnabel, R.B., 1987. Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics* 29, 67–82. <https://doi.org/10.1080/00401706.1987.10488184>
- Fernández-Villaverde, J., Jones, C., 2020. Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities. *Natl. Bur. Econ. Res.* <https://doi.org/10.3386/w27128>
- Harko, T., Lobo, F.S.N., Mak, M.K., 2014. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* 236, 184–194. <https://doi.org/10.1016/j.amc.2014.03.030>
- Kang, G., Bates, D.M., 1990. Approximate Inferences in Multiresponse Regression Analysis. *Biometrika* 77, 321–331.

- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks. IEEE, pp. 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character 115, 700–721. <https://doi.org/10.1098/rspa.1927.0118>
- Kessentini, S., Barchiesi, D., 2015. Particle Swarm Optimization with Adaptive Inertia Weight. Int. J. Mach. Learn. Comput. 5, 368–373. <https://doi.org/10.7763/IJMLC.2015.V5.535>
- Koduru, P., Welch, S.M., Das, S., 2007. A particle swarm optimization approach for estimating parameter confidence regions. Proc. GECCO 2007 Genet. Evol. Comput. Conf. 70–77. <https://doi.org/10.1145/1276958.1276969>
- Konsumentenschutz, B.S.G.P. und, 2020. Official COVID19 Dashboard [WWW Document].
- Luo, J., 2020. Predictive Monitoring of COVID-19. Singapore Univ. Technol. Des. 1–12.
- Postnikov, E.B., 2020. Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions? Chaos, Solitons and Fractals 135, 109841. <https://doi.org/10.1016/j.chaos.2020.109841>
- Ruder, S., 2016. An overview of gradient descent optimization algorithms.
- Salute, M. della, 2020. Covid-19 -Situazione in Italia [WWW Document].
- Schüttler, J., Schlickeiser, R., Schlickeiser, F., Kröger, M., 2020. Covid-19 Predictions Using a Gauss Model, Based on Data from April 2. Physics (College. Park. Md). 2, 197–212. <https://doi.org/10.3390/physics2020013>
- Schwaab, M., Alberton, A.L., Pinto, J.C., 2010. ESTIMA&PLANEJA: Pacote computacional para estimação de parâmetros e de planejamento de experimentos. Rio de Janeiro.
- Schwaab, M., Biscaia E.C., J., Monteiro, J.L., Pinto, J.C., Biscaia Júnior, E.C., Monteiro, J.L., Pinto, J.C., Biscaia, E.C., Monteiro, J.L., Pinto, J.C., 2008. Nonlinear parameter estimation through particle swarm optimization. Chem. Eng. Sci. 63, 1542–1552. <https://doi.org/10.1016/j.ces.2007.11.024>
- Schwaab, M., Pinto, J.C., 2007. Análise de Dados Experimentais I: Fundamentos de

Estatística e Estimação de Parâmetros. e-Papers, Rio de Janeiro.

- Shi, Y., Eberhart, R.C., 1998. A modified particle swarm optimizer, in: Proceedings of the 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence. IEEE, pp. 69–73. <https://doi.org/10.1109/ICEC.1998.699146>
- Woody, S., Garcia Tec, M., Dahan, M., Gaither, K., Fox, S., Meyers, L.A., Scott, J., 2020. Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones 1–13. <https://doi.org/10.1101/2020.04.16.20068163>
- Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., He, J., 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* 12, 165–174. <https://doi.org/10.21037/jtd.2020.02.64>
- Yoo, S.K., Cotton, S.L., Sofotasios, P.C., Matthaiou, M., Valkama, M., Karagiannidis, G.K., 2017. The Fisher-Snedecor F Distribution: A Simple and Accurate Composite Fading Model. *IEEE Commun. Lett.* 21, 1661–1664. <https://doi.org/10.1109/LCOMM.2017.2687438>
- You, C., Deng, Y., Hu, W., Sun, J., Lin, Q., Zhou, F., Pang, C.H., Zhang, Y., Chen, Z., Zhou, X.H., 2020. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int. J. Hyg. Environ. Health* 228, 113555. <https://doi.org/10.1016/j.ijheh.2020.113555>

## TABLES AND FIGURES

Table 1 - Parameter estimation results for the Gaussian model using different percentages of the same data set to COVID-19's infected cases for Italy and Austria.

Country	Data Set	S( $\theta$ )	$t_{95\%}$	Confidence interval limits ( $t_{95\%}$ )		Variance $t_{95\%}$	Parameters	Estimated value
Italy	10%	1.19E+05	46	29	94	115.92	N	22025.8
							$\mu$	29.7
							$\sigma$	9.65
	30%	7.41E+06	60	59	61	0.14	N	82319.8
							$\mu$	40.7
							$\sigma$	11.8
	50%	9.68E+08	84	81	87	1.12	N	109778.0
							$\mu$	53.5
							$\sigma$	18.5
	70%	2.84E+09	97	95	99	0.57	N	112634.0
							$\mu$	59.4
							$\sigma$	22.9
	100%	7.73E+09	104	102	107	0.61	N	108513.5
							$\mu$	61.8
							$\sigma$	25.9
Austria	10%	28.49	42	15	85	60.79	N	857.9
							$\mu$	29.2
							$\sigma$	8.04
	30%	1.42E+05	51	51	52	0.09	N	9039.51
							$\mu$	37.7
							$\sigma$	8.32
	50%	1.29E+07	57	56	59	0.14	N	8864.0
							$\mu$	40.1
							$\sigma$	10.6
	70%	4.40e+07	59	57	61	0.34	N	8613.7
							$\mu$	40.5
							$\sigma$	11.3
	100%	5.11E+07	59	57	61	0.28	N	8612.9
							$\mu$	40.5
							$\sigma$	11.3

Table 2 - Parameter estimation results for the SIR model using different percentages of the same data set to COVID-19's infected cases for Italy and Austria.

Country	Data Set	S( $\theta$ )	$t_{95\%}$	Confidence interval limits ( $t_{95\%}$ )	Variance $t_{95\%}$	Parameters	Estimated value
Italy	10%	7.12E+05	100		13.28	$\beta$	0.2874
						$\gamma$	0.0424
	30%	5.79E+09	100		10.20	$\beta$	0.2392
						$\gamma$	0.0505
	50%	4.99E+10	117		2.72	$\beta$	0.2031
						$\gamma$	0.0381
	70%	7.00E+10	124		0.62	$\beta$	0.1936
						$\gamma$	0.0319
	100%	7.15E+10	124		0.27	$\beta$	0.1934
						$\gamma$	0.0320
Austria	10%	218.37	269		205.9	$\beta$	0.3250
						$\gamma$	0.0102
	30%	3.35E+07	113		18.11	$\beta$	0.3087
						$\gamma$	0.0337
	50%	2.50E+08	86		1.88	$\beta$	0.3246
						$\gamma$	0.0614
	70%	3.28E+08	86		1.18	$\beta$	0.3242
						$\gamma$	0.0623
	100%	3.49E+08	86		0.84	$\beta$	0.3236
						$\gamma$	0.0620

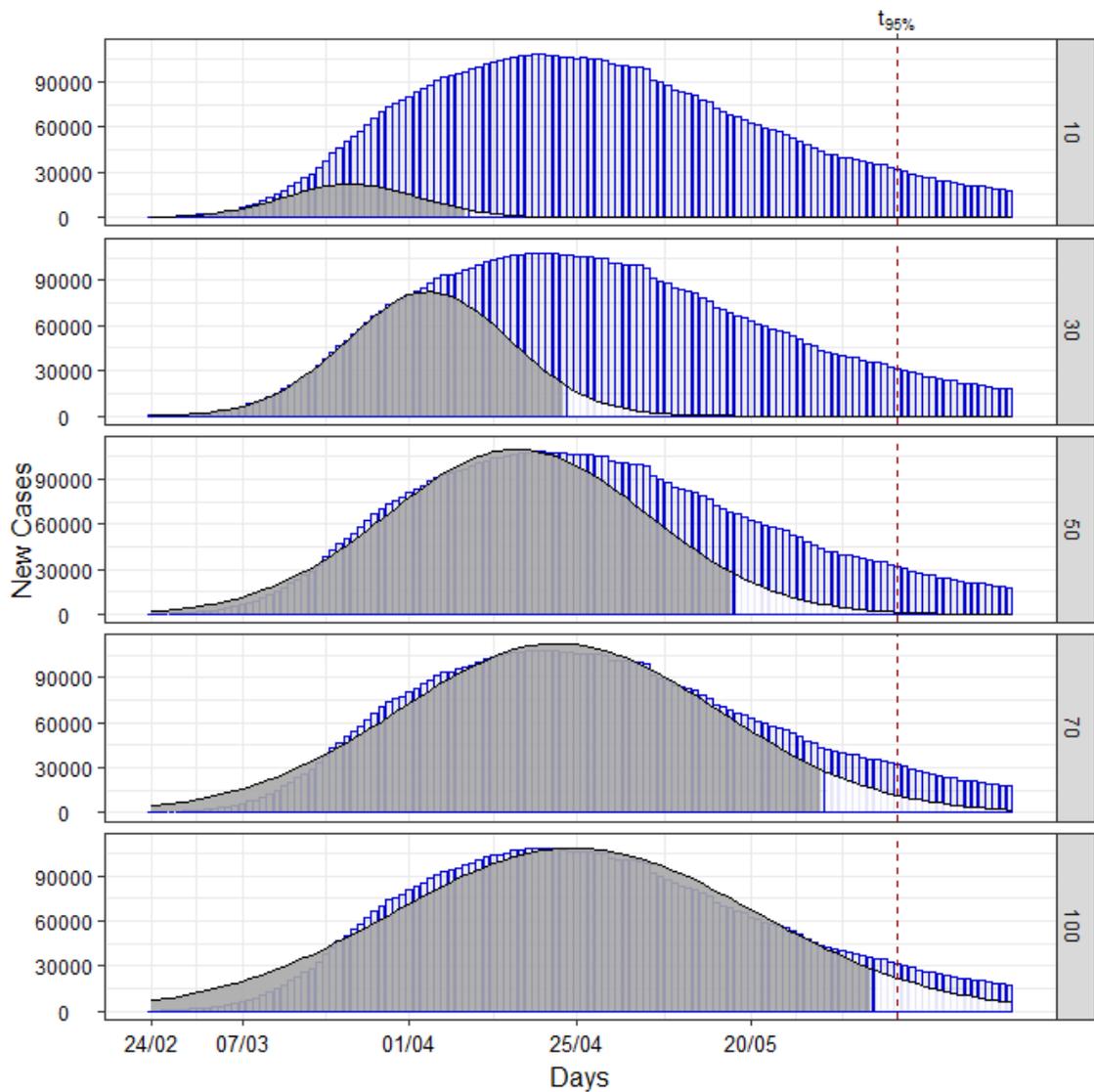


Figure 1 - Comparison between infected cases observed of COVID - 19 in Italy and the estimated Gaussian model curves for different percentages of this data set.

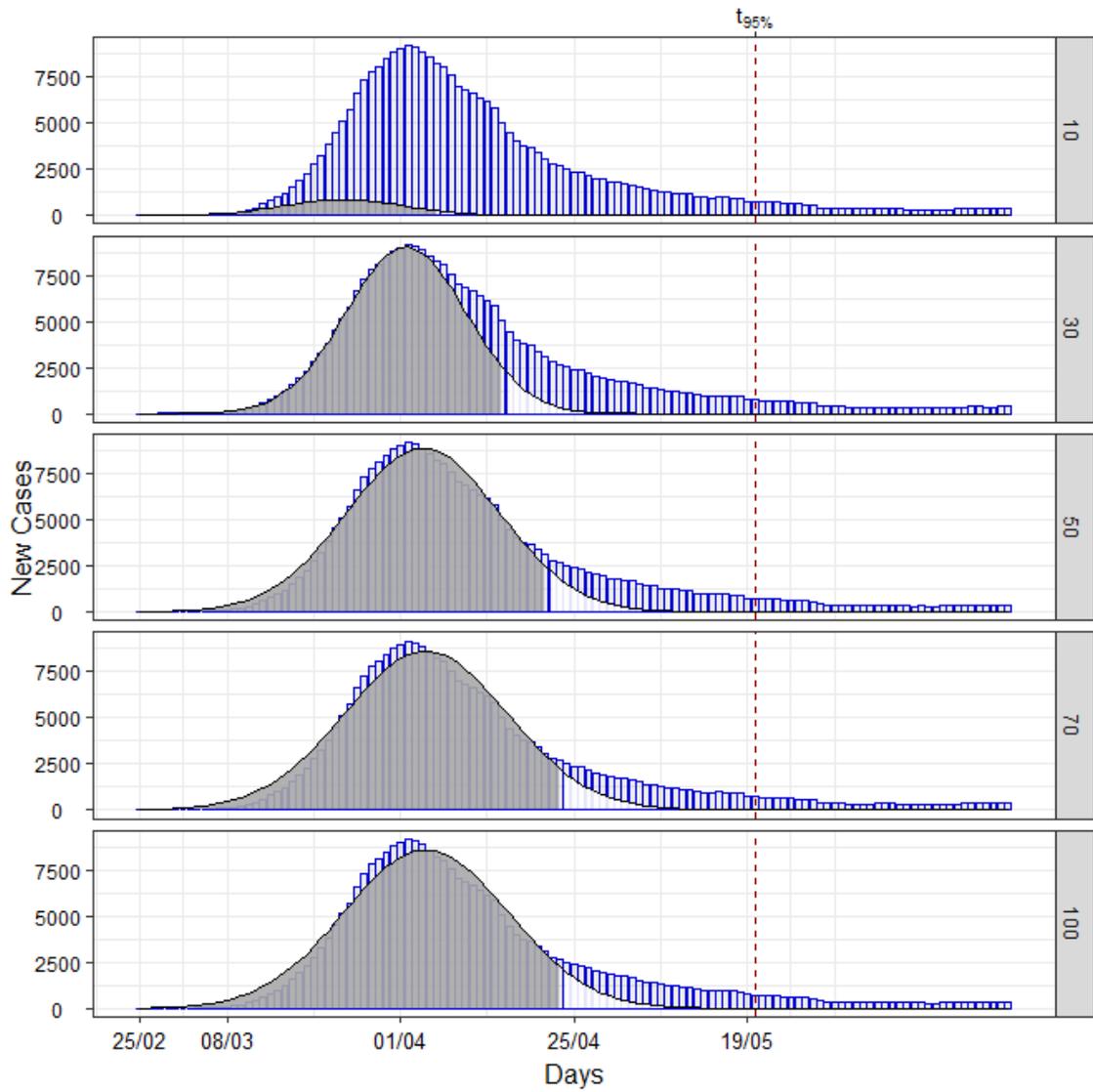


Figure 2 - Comparison between infected cases observed of COVID - 19 in Austria and the estimated Gaussian model curves for different percentages of this data set.

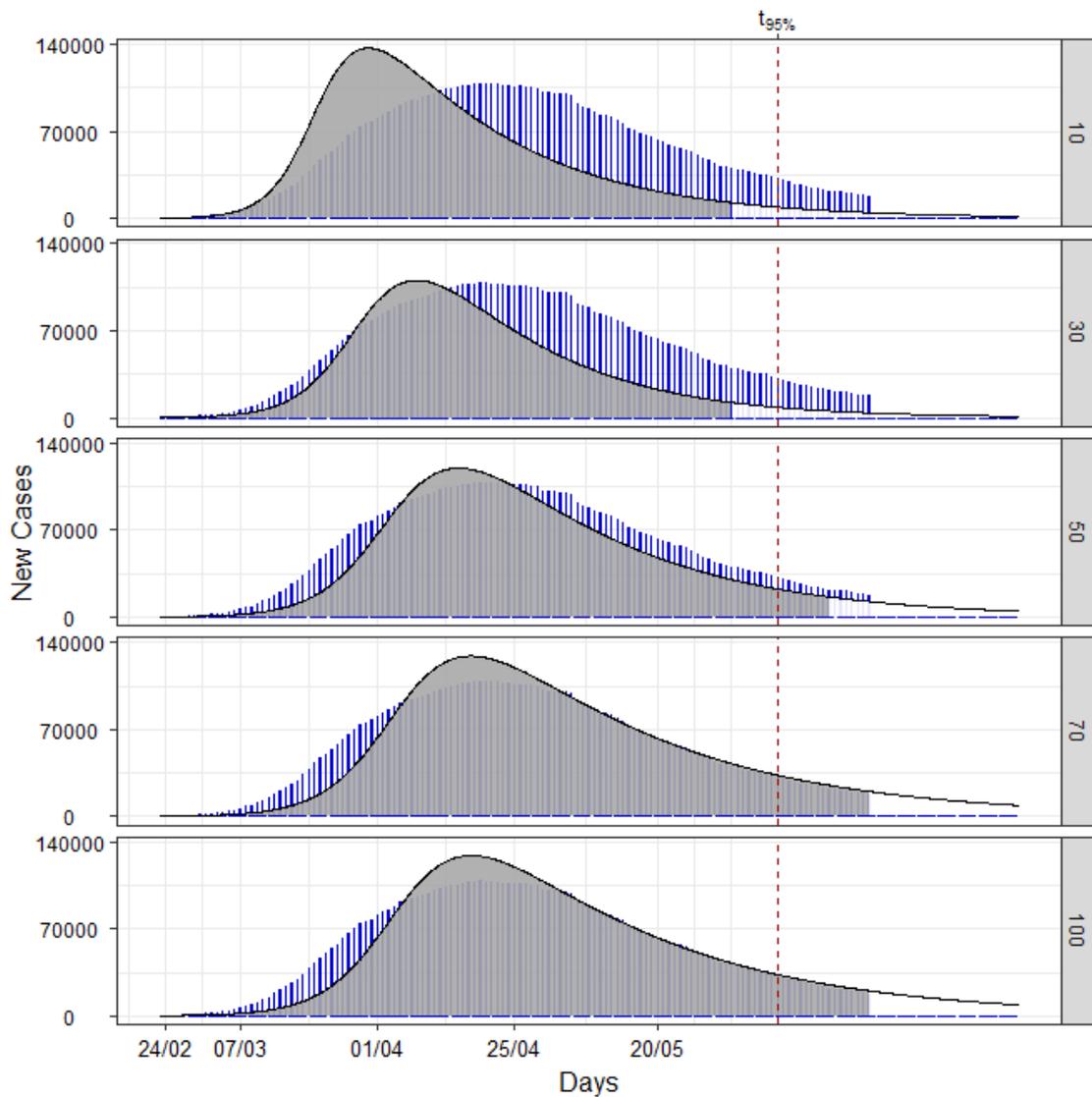


Figure 3 - Comparison between infected cases observed of COVID - 19 in Italy and the estimated SIR model curves for different percentages of this data set.

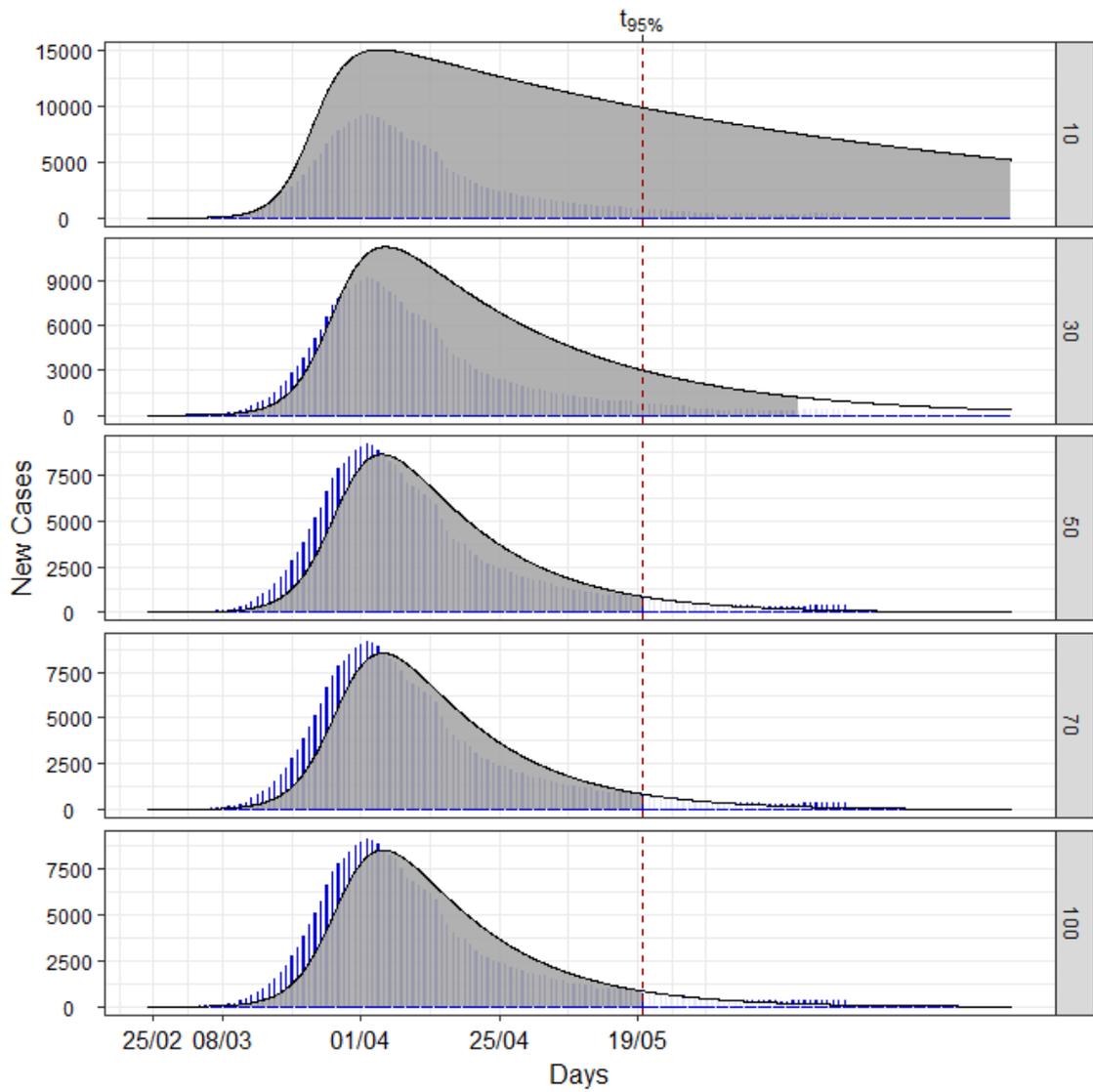


Figure 4 - Comparison between infected cases observed of COVID - 19 in Austria and the estimated SIR model curves for different percentages of this data set.