



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-5548-TDI/534

**ALGUMAS FERRAMENTAS ESTATÍSTICAS
NA SÍNTESE, PROCESSAMENTO E ANÁLISE DE
IMAGENS DE RADAR DE ABERTURA SINTÉTICA**

Alejandro César Frery Orgambide

Tese de Doutorado em Computação Aplicada, orientada pelos Professores Nelson
D. A. Mascarenhas e O. H. Bustos, aprovada em 10 de dezembro de 1993.

INPE
São José dos Campos
10 de dezembro de 1993

Publicado por:

Coordenação de Ensino, Documentação e
Programas Especiais - CEP

Instituto Nacional de Pesquisas Espaciais - INPE

Caixa Postal 515

12201-970 - São José dos Campos - SP - Brasil

Fone: (012) 345.6911

Fax: (012) 345.6919

E-Mail: marciana@sid.inpe.br

- Solicita-se intercâmbio
- Exchange welcome
- Si sollecita intercambio
- Echange souhaité
- Mann bittet un Austausch
- Solicitamos intercambio
- Просим обмена
- 歡迎進行交換
- 出校外交渉のお願い

Publicação Externa - É permitida sua reprodução para interessados.

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-5548-TDI/534

**ALGUMAS FERRAMENTAS ESTATÍSTICAS
NA SÍNTESE, PROCESSAMENTO E ANÁLISE DE
IMAGENS DE RADAR DE ABERTURA SINTÉTICA**

Alejandro César Frery Orgambide

Tese de Doutorado em Computação Aplicada, orientada pelos Professores Nelson
D. A. Mascarenhas e O. H. Bustos, aprovada em 10 de dezembro de 1993.

INPE
São José dos Campos
10 de dezembro de 1993

**ALGUMAS FERRAMENTAS ESTATÍSTICAS
NA SÍNTESE, PROCESSAMENTO E ANÁLISE DE
IMAGENS DE RADAR DE ABERTURA SINTÉTICA**

Alejandro César Frery Orgambide

INPE
São José dos Campos
10 de dezembro de 1993

621.376.5

ORGAMBIDE, A.C.F.

Algumas ferramentas estatísticas na síntese, processamento e análise de imagens de radar de abertura sintética. / Alejandro César Frery Orgambide. -- São José dos Campos: INPE, 1993. 244p. -- (INPE-5548-TDI/534)

1. imagens de radar. 2. modelagem estatística. 3. processamento de imagens. I.Título.

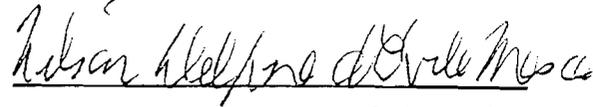
Aprovada pela Banca Examinadora
em cumprimento a requisito exigido
para a obtenção do Título de Doutor
em Computação Aplicada

Dr. Gerald Jean Francis Banon



Presidente

Dr. Nelson Delfino d'Ávila Mascarenhas



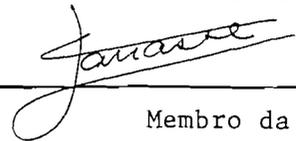
Orientador

Dr. Luciano Vieira Dutra



Membro da Banca

Dra. Corina da Costa Freitas Yanasse



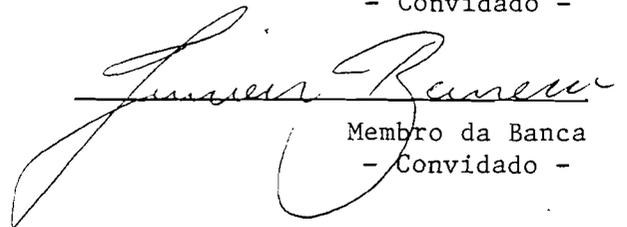
Membro da Banca

Dr. Fernando Toshinori Sakane



Membro da Banca
- Convidado -

Dr. Junior Barrera



Membro da Banca
- Convidado -

Candidato: Alejandro Cesar Frery Orgambide

São José dos Campos, 10 de dezembro de 1993

“Vistas por los ángeles, tal vez las cimas de los árboles
sean raíces que beben de los cielos;
y, en el suelo, las profundas raíces de un haya
les parezcan pináculos silenciosos

¿Acaso no es para ellos la tierra transparente
frente a un cielo macizo como un cuerpo?
Es tierra ardiente donde se lamenta
junto a las fuentes el olvido de los muertos.”

RAINER MARIA RILKE

“En [las] remotas páginas [que el Doctor Franz Kunz atribuye a cierta enciclopedia china que se titula *Emporio Celestial de conocimientos benévolos*] está escrito que los animales se dividen en (a) pertenecientes al Emperador, (b) embalsamados, (c) amaestrados, (d) lechones, (e) sirenas, (f) fabulosos, (g) perros sueltos, (h) incluídos en esta clasificación, (i) que se agitan como locos, (j) innumerables, (k) dibujados con un pincel finísimo de pelo de camello, (l) etcétera, (m) que acaban de romper el jarrón, (n) que de lejos parecen moscas. El Instituto Bibliográfico de Bruselas también ejerce el caos: ha parcelado el universo en 1000 subdivisiones, de las cuales la 262 corresponde al Papa; la 282, a la Iglesia Católica Romana; la 263 al Día del Señor; la 268, a las escuelas dominicales; la 298, al mormonismo, y la 294 al brahmanismo, budismo, shitoísmo y taoísmo. No rehusa las subdivisiones heterogéneas, verbigracia, la 179: “Crueldad con los animales. Protección a los animales. El duelo y el suicidio desde el punto de vista de la moral. Vicios y defectos varios. Virtudes y cualidades varias”.

...notoriamente no hay clasificación del universo que no sea arbitraria y conjetural. La razón es muy simple: no sabemos qué cosa es el universo.”

JORGE LUIS BORGES: El idioma analítico de John Wilkins; en *Otras Inquisiciones*. 1960. Emecé Editores. Argentina.

Lhe devo ao meu irmão Mario a felicidade quase premonitória de conhecer essas estrofes de Rilke, dentre muitas outras coisas. A minha irmã Mirta, a capacidade de apreciá-las. A minha mãe e a minha avó, tudo. Aos meus amigos, as referências afetivas e o apoio permanente e desinteressado. Ao INPE, ao Brasil e ao seu povo, a possibilidade de estar escrevendo esta dedicatória.

E, é claro, também dedico esta tese a μ .

AGRADECIMENTOS

Fazer uma enumeração das pessoas que contribuíram para o desenvolvimento desta tese e às quais eu gostaria de agradecer a sua participação é, necessariamente, uma tarefa ingrata: a minha memória não é suficientemente boa nem tenho o espaço necessário para fazê-lo. Mesmo assim, e correndo o risco de ser injusto, gostaria de expressar a minha profunda gratidão aos meus orientadores, Nelson D. A. Mascarenhas e Oscar Bustos, que ficaram de plantão durante estes anos guiando e ajudando. Vários amigos tiveram um papel central neste trabalho, ora para discutir assuntos técnicos, ora para melhorar a minha qualidade de vida; dentre eles, é impossível não mencionar Lúbia Vinhas, Tania Sausen, Goya Arenas, Ruben Frery, Jorge Abraham, Sidnei Sant'Anna, Maurício Marengoni, Danton Nunes, Alberto Ferrari e Emílio Muno. Os professores Gerald Banon e Maria Eulália Vares fizeram aportes muito importantes.

O ambiente de camaradagem da DPI foi o lugar perfeito para desenvolver este trabalho. A toda essa família, o meu sincero agradecimento.

As palavras de destaque são para a professora e amiga Corina Yanasse, ajuda, estímulo e companhia permanente durante boa parte deste trabalho.

Este trabalho foi parcialmente desenvolvido com recursos do Projeto Temático de Equipe No. 91/3532-2 da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

RESUMO

Apresenta-se o desenvolvimento de ferramentas computacionais, derivadas da modelagem estatística, para a síntese, a análise e o processamento de imagens de radar de abertura sintética. Diversos problemas de imagens são enfocados de forma unificada, dentro do contexto Bayesiano. O modelo de Potts-Strauss é empregado como distribuição *a priori* para as classes, e são estudadas: a técnica de estimação por máxima pseudoverossimilhança, as propriedades de convergência de algoritmos de aglomerados e a invariância da classe de distribuições Markovianas para a distribuição *a posteriori* de uma imagem, dada a sua versão borrada e ruidosa. Especificamente, são desenvolvidos, de forma computacionalmente tratável, os estimadores de pseudoverossimilhança para o modelo de Potts-Strauss e um número arbitrário de classes. As propriedades de convergência são estudadas com uma experiência Monte Carlo, pelo uso dos estimadores de pseudoverossimilhança como critério de parada. É apresentada uma extensão do Teorema de Geman e Geman (1984), sobre a invariância dos campos Markovianos que permite modelar imagens com ruído multiplicativo, entre outras. Utilizando-se os resultados sobre o modelo de Potts-Strauss, é desenvolvida uma ferramenta para a simulação de imagens ópticas e de radar. Estudam-se várias distribuições e estimadores de parâmetros adequados para o ruído em imagens obtidas por iluminação coerente. Propõe-se e desenvolve-se uma modificação de um algoritmo Markoviano (ICM) para segmentação de imagens SAR e ópticas, cujos parâmetros são estimados iterativamente, pelos estimadores de pseudoverossimilhança, para automatizar o seu uso. Propõe-se o uso de filtros robustos, baseados nas propriedades da distribuição Rayleigh e uma analogia entre filtragem e estimação, para redução de ruído *speckle* em imagens de uma visada e detecção linear. Desenvolve-se uma ferramenta computacional para avaliar a aderência de amostras de imagens às distribuições propostas como modelos de ruído, e essa informação é empregada na discriminação de tipos de ocupação do solo. Avalia-se, pelo uso de técnicas Monte Carlo, a influência que o formato dos dados tem no desempenho dos testes de aderência.

SOME STATISTICAL TOOLS FOR SAR IMAGE SYNTHESIS, ANALYSIS AND PROCESSING

ABSTRACT

Computational tools, derived from statistical modelling, for the synthesis, analysis and processing of synthetic aperture radar (SAR) images are presented. Several problems arising in image processing are posed in a unified manner, within the Bayesian framework. The Potts-Strauss model is used as the *a priori* distribution for the classes, and the following topics are studied: the maximum pseudolikelihood estimation, the convergence properties of cluster flip algorithms, and the invariance Markovian property of the *a posteriori* distribution of a random field, given its blurred and noisy version. Specifically, the pseudolikelihood estimators for the Potts-Strauss model and an arbitrary number of classes are calculated in a tractable computational form. The convergence properties are studied using a Monte Carlo experiment, using the pseudolikelihood estimators as stopping rules. An extension of a theorem by Geman and Geman is presented; this extension allows the modelling of images with multiplicative noise, among others. Using these results about the Potts-Strauss model, a computational tool for the simulation of optical and SAR images is presented. Several distributions for the noise present in coherent illuminated images are studied, along with estimation techniques for their parameters. A modification of a Markovian algorithm (ICM) for the segmentation of SAR and optical images is presented, where all the required parameters are iteratively estimated from the data through the use of pseudolikelihood estimators allowing, thus, an automated operation. The speckle noise reduction in one look amplitude images is posed as an estimation problem, and robust filters, based on the Rayleigh distribution, are proposed for this task. A computational tool for two goodness of fit tests for images is presented; that information is used as a discriminating factor for land use. The influence of byte data format on these tests is evaluated through the use of Monte Carlo techniques.

SUMÁRIO

	Pág.
LISTA DE FIGURAS	xix
LISTA DE TABELAS	xxiii
CAPÍTULO 1 INTRODUÇÃO	1
CAPÍTULO 2 ANALOGIAS ENTRE TÉCNICAS DE INFERÊNCIA ESTATÍSTICA E PROBLEMAS NA ANÁLISE E NO PROCESSAMENTO DE IMAGENS	7
2.1 Introdução	7
2.2 Notação	8
2.3 Inferência Clássica	9
2.3.1 O Método dos Momentos	10
2.3.2 O Método da Máxima Verossimilhança	11
2.3.3 Inferência Bayesiana	12
2.4 Algumas Distribuições Importantes e seus Estimadores	15
2.5 O Paradigma de Bayes e Problemas na Análise e no Processamento de Imagens	26
2.6 Problemas de Processamento e Análise de Imagens	28
2.6.1 Segmentação em Modelos Básicos (o Método de Máxima Verossimilhança)	30
2.6.2 Modelos com Dependência Espacial	36
2.6.3 Estimadores MAP, MPM e ICM	39
2.6.4 Outros Estimadores	41

CAPÍTULO 3	O MODELO DE POTTS-STRAUSS, A ESTIMAÇÃO POR PSEUDOVEROSSIMILHANÇA E A SIMULAÇÃO DE IMAGENS SAR	43
3.1	Introdução	43
3.2	Formulação Teórica	45
3.3	Estimação pela Equação de Máxima Pseudoverossimilhança	52
3.4	Simulação do modelo de Potts-Strauss	59
3.4.1	Dinâmicas de <i>spin-flip</i>	60
3.4.1.1	Algoritmo de Metropolis	60
3.4.1.2	Algoritmo de Geman e Geman	61
3.4.2	Dinâmicas de Aglomerados	62
3.4.2.1	Algoritmo de Swendsen-Wang	64
3.4.2.2	Algoritmo de Wolff	65
3.4.2.3	Obtenção das Componentes Conexas de um Grafo	66
3.4.3	Tempos de Convergência	68
3.5	Simulação de Imagens SAR	76
3.5.1	Observações Espacialmente Independentes	78
3.5.2	Observações Espacialmente Dependentes	82
3.6	Simulação de Imagens Ópticas	86
3.7	Uma Extensão do Teorema de Geman e Geman	87
3.7.1	Nota sobre a Extensão	94
3.7.2	Exemplo de Aplicação a Imagens SAR	94
CAPÍTULO 4	IMPLEMENTAÇÕES DO ALGORITMO ICM	99
4.1	Técnicas Markovianas de Segmentação	99

4.2	O algoritmo ICM	102
4.2.1	Implementação do ICM para Segmentação de Imagens SAR de uma Visada e Detecção Linear	103
4.2.2	Implementações do ICM para Segmentação de Imagens Ópticas Multiespectrais no Ambiente Spring	111
 CAPÍTULO 5 A DISTRIBUIÇÃO RAYLEIGH E FILTROS ROBUSTOS PARA DETECÇÃO LINEAR E UMA VISADA		119
5.1	Introdução	119
5.2	Propriedades da Distribuição Rayleigh	123
5.3	Estimação	126
5.4	Estimadores Robustos	129
5.5	Avaliação de Estimadores	130
5.6	Estimador de Máxima Verossimilhança Aproximado para Amostras Aparadas	138
5.7	Exemplo	141
5.8	Filtros Robustos para Detecção Linear e Uma Visada	142
5.8.1	Extensões	148
 CAPÍTULO 6 APLICAÇÃO DE ANÁLISE DE ESTATÍSTICA DE DADOS A IMAGENS DA MISSÃO SAREX SOBRE A RESERVA FLORESTAL DE TAPAJÓS ..		159
6.1	Introdução	159
6.2	Análise Preliminar	161
6.3	Testes de Qualidade de Ajuste	162
6.3.1	Teste χ^2	168
6.3.2	Teste de Kolmogorov-Smirnov	169

6.4	Discriminação da Ocupação do Solo	169
6.5	Influência do Formato dos Dados nos Testes de Ajuste	177
CAPÍTULO 7 CONCLUSÕES		185
REFERÊNCIAS BIBLIOGRÁFICAS		189
APÊNDICE A INFORMAÇÃO COMPUTACIONAL		205
A.1	Geradores de Números Pseudoaleatórios	205
A.2	Aplicativos	206
APÊNDICE B LISTAGENS DE ALGUNS DOS PROGRAMAS IM- PLEMENTADOS		207
B.1	Programa para Simulação do Modelo de Potts-Strauss e Contaminação Rayleigh com Número Arbitrário de Visadas e Detecção Linear	207
B.2	Programa para Segmentação em um Número Arbitrário de Classes, pelo Algoritmo ICM, de Imagens com Contaminação Rayleigh, uma Visada e Detecção Linear	211
B.3	Programa que implementa os Filtros Robustos	223

LISTA DE FIGURAS

		Pág.
2.1	Densidade Beta usada como distribuição <i>a priori</i>	14
2.2	Função a ser minimizada para a estimação de parâmetros da distribuição \mathcal{KAn}	23
2.3	Função a ser minimizada para a estimação de parâmetros da distribuição \mathcal{KAn}	23
2.4	O suporte de uma imagem particionado para o modelo de mosaico	37
3.1	Número de iterações até a convergência do algoritmo de Swendsen-Wang para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura infinita.	71
3.2	Número de iterações até a convergência do algoritmo de Wolff para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura infinita	72
3.3	Número de iterações até a convergência do algoritmo de Swendsen-Wang para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura zero.	74
3.4	Número de iterações até a convergência do algoritmo de Wolff para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura zero.	75
3.5	Tamanho médio dos aglomerados, medidos com simulações obtidas pela dinâmica de Swendsen-Wang.	76
3.6	Valores estimados de média e desvio padrão do estimador de pseudoverossimilhança	77
3.7	Valores teórico e estimado da magnetização média absoluta do modelo de Ising.	78
3.8	Valores da correlação a distância "1".	79
3.9	Valores da correlação a distância "10".	80
3.10	Ocorrência do modelo de Potts-Strauss para $\beta = 0.99$ e três classes.	81

3.11	Simulação de uma imagem observada de uma visada.	82
3.12	Simulação de uma imagem observada de três visadas.	83
3.13	Simulação de uma imagem observada de seis visadas.	84
3.14	Simulação de um campo de <i>speckle</i> correlacionado.	86
3.15	Simulação de uma imagem com duas classes, três bandas e distribuição \mathcal{N}^3	88
3.16	Ocorrência da distribuição <i>a priori</i> borrada em vizinhanças de segunda ordem.	96
3.17	Ocorrência da distribuição <i>a priori</i> borrada em vizinhanças de segunda ordem e elevada ao quadrado.	97
3.18	Ocorrência da distribuição para a imagem observada.	98
3.19	Ilustração das vizinhanças envolvidas no exemplo do Teorema de Extensão.	98
4.1	Seqüência de visitas do algoritmo ICM a um suporte de tamanho 36	106
4.2	Imagem original: banda L, 512 <i>pixels</i> de lado.	107
4.3	Segmentação da imagem original em duas classes pelo algoritmo de máxima verossimilhança.	108
4.4	Segmentação da imagem original em duas classes pelo algoritmo ICM.	108
4.5	Segmentação da imagem original em três classes pelo algoritmo de máxima verossimilhança.	109
4.6	Segmentação da imagem original em três classes pelo algoritmo ICM.	110
4.7	Transformação logarítmica da distribuição Rayleigh: (a) três densidades Rayleigh; (b) as densidades dos respectivos logaritmos; (c) densidade do logaritmo e densidade da normal com mesmas média e variância superpostas	115
4.8	Imagem SAR-580, logaritmo da banda L	116
4.9	Imagem SAR-580, logaritmo da banda X	116
4.10	Segmentação por máxima verossimilhança dos logaritmos das bandas L e X	117

4.11	Segmentação pelo algoritmo ICM dos logaritmos das bandas L e X	117
5.1	Modelo para o SAR	120
5.2	(a) Função de densidade de probabilidade Rayleigh para vários valores do parâmetro. (b) Função de densidade de probabilidade Exponencial para vários valores do parâmetro.	150
5.3	Ajuste das distribuições Rayleigh, Weibull e $\mathcal{KA6}$ a uma amostra florestal.	151
5.4	Ajuste das distribuições Normal e Raiz de Gamma a uma amostra de pastagem.	151
5.5	Histogramas e densidades Rayleigh estimadas para duas áreas homogêneas	152
5.6	Imagem SAR-580, banda L.	152
5.7	Densidades do estimador MAD para amostras de tamanho um, cinco e nove.	153
5.8	Imagem filtrada pelo estimador MV.	153
5.9	Imagem filtrada pelo estimador MO.	154
5.10	Imagem filtrada pelo estimador TMV.	154
5.11	Imagem filtrada pelo estimador TMO.	155
5.12	Imagem filtrada pelo estimador MAD.	155
5.13	Imagem filtrada pelo estimador IQR.	156
5.14	Imagem filtrada pelo estimador Med.	156
5.15	Histogramas das imagens original e filtradas pelos algoritmos MV, MO, TMV, TMO, MAD, IQR e Med.	157
6.1	Média e desvio padrão do número de amostras Rayleigh rejeitadas aos 90%.	182
6.2	Média e desvio padrão do número de amostras Rayleigh rejeitadas aos 95%.	183
6.3	Média e desvio padrão do número de amostras Rayleigh rejeitadas aos 99%.	184

LISTA DE TABELAS

		Pág.
3.1	Médias e desvios padrão dos números de iterações e tempos de CPU até a convergência dos algoritmos de Swendsen-Wang e Wolff para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura infinita	70
3.2	Médias e desvios padrão dos números de iterações e tempos de CPU até a convergência dos algoritmos de Swendsen-Wang e Wolff para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura nula	73
4.1	Principais estimadores Bayesianos e as suas características	101
4.2	Valores (aproximados) da esperança, variância, assimetria e kurtose das distribuições Rayleigh padrão e o seu logaritmo.	113
4.3	Valores do estimador do parâmetro de atratividade e do número de <i>pixels</i> trocados por iteração no algoritmo ICM	114
5.1	Valores amostrais de média, desvio padrão médio e desvio padrão dos estimadores.	134
5.2	Valores amostrais de mínimo, mediana e máximo dos estimadores.	135
5.3	Quartis inferior e superior dos estimadores.	136
5.4	O estimador atinge o valor verdadeiro?	137
5.5	Coefficientes da regressão sobre os estimadores.	138
5.6	Coefficientes do estimador BLUE para pequenas amostras.	140
5.7	Quantidades estimadas.	146
5.8	p -valores percentuais dos Testes Kolmogorov-Smirnov e χ^2	147
5.9	Valores estimados sobre a região escura.	147
5.10	Valores estimados sobre a região clara.	148
6.1	Quantidades estimadas nas amostras de floresta.	163

6.1	Continuação.	164
6.1	Conclusão.	165
6.2	Quantidades estimadas nas amostras de não floresta.	166
6.2	Conclusão.	167
6.3	Distribuições escolhidas por quatro critérios para amostras de floresta	171
6.3	Continuação	172
6.3	Continuação	173
6.3	Conclusão	174
6.4	Distribuições escolhidas para amostras de não floresta por quatro critérios.	175
6.4	Conclusão.	176
6.5	Distribuições <i>preferidas</i> pelas amostras de floresta.	177
6.6	Distribuições <i>preferidas</i> pelas amostras de não floresta.	177
6.7	Média e desvio padrão do número de amostras Rayleigh de tamanho 500 rejeitadas aos 90% pelos testes Kolmogorov-Smirnov e χ^2	180
6.8	Média e desvio padrão do número de amostras Rayleigh de tamanho 500 rejeitadas aos 95% pelos testes Kolmogorov-Smirnov e χ^2	181
6.9	Média e desvio padrão do número de amostras Rayleigh de tamanho 500 rejeitadas ao 99% pelos testes Kolmogorov-Smirnov e χ^2	183

CAPÍTULO 1

INTRODUÇÃO

Apesar do radar ter sido intensamente utilizado desde a Segunda Guerra Mundial, o seu uso pela comunidade civil foi, até recentemente, limitado a poucas áreas, tais como controle de tráfego aéreo e meteorologia. Nos últimos anos tem-se notado um crescente interesse pelo estudo da potencialidade do radar na discriminação de culturas, florestas, desflorestamento e detecção de mudanças. Como exemplos pode-se citar as campanhas AGRISAR 1986 e SAREX 1992, e os satélites ERS-1, JERS-1, SIR-C, X-SAR, Radarsat, etc. (43).

O radar possui uma clara vantagem sobre sensores ópticos, devido ao fato de sua capacidade ser praticamente inafetada pela escuridão, pelas nuvens, pela neblina e/ou pela fumaça. O retroespalhamento do radar é muito sensível a vários parâmetros do alvo (constante dielétrica, rugosidade da superfície, etc.), e do sistema (comprimento de onda, polarização, resolução, ângulo de incidência, etc.). Tal característica fornece, portanto, informações multidimensionais que podem ser usadas em áreas importantes, como monitoramento agrícola, especialmente se um radar de alta resolução (como é o caso do radar de abertura sintética —SAR—), é utilizado.

O sistema de imageamento por radar é ativo e, portanto, não precisa de fontes externas de iluminação como é o caso dos sensores ópticos que só podem operar durante o dia. A frequência de operação dos sistemas de imageamento por microondas é diferente da empregada pelos sistemas convencionais, e condições atmosféricas adversas não afetam (ou afetam pouco) o desempenho do radar.

O potencial do uso do SAR em monitoramento agrícola é de grande interesse para a comunidade civil. Entretanto, muitos estudos são ainda necessários para se descobrir como a informação de campo deve ser extraída dos dados de radar. Uma das informações importantes, por exemplo, é saber quais as

propriedades estatísticas de dados do SAR. O conhecimento da função densidade de probabilidade é altamente relevante para várias técnicas de processamento digital de imagens, tais como classificação e construção de filtros redutores de *speckle*.

Um ponto central é modelar probabilisticamente os dados do SAR, isto é, supor, a menos de parâmetros, qual é a distribuição de onde provêm os dados observados, tanto na distribuição marginal quanto na conjunta. Aceito o modelo proposto, faz-se necessário inferir a respeito desses parâmetros. Para tanto, lança-se mão de estatísticas de diversas ordens. Por exemplo, a de primeira ordem diz respeito à tonalidade média dos *pixels* na região de interesse, enquanto a de segunda ordem fornece informações sobre as relações entre cada *pixel* e os outros. A função de autocovariância, que é uma estatística de segunda ordem que leva em conta relações de vizinhança, é essencial nas estimativas da variância das médias amostrais e variância amostral dos *pixels*. Estas estimativas também contém informação sobre a textura, que é muito útil na discriminação dos diversos tipos de uso do solo. O conhecimento da função de autocorrelação é também requerido em muitos algoritmos de redução de *speckle*. Entretanto, devido à sensibilidade do radar aos diversos parâmetros do alvo e do sistema, e à falta de dados de alta qualidade para avaliar o comportamento de diferentes alvos aos sinais de radar, ainda não foi possível obter-se um conhecimento mais completo das propriedades estatísticas dos dados de radar.

Em (126) conta-se uma estória que ilustra bem as potencialidades do imageamento usando radar de abertura sintética. O veículo orbital Pioneer Venus revelou, pela primeira vez, a superfície do planeta Venus em escala global. Para isto, foi equipado de um sistema de radar para atravessar a camada de nuvens densas que recobrem, em forma permanente, a superfície do planeta.

Um equipamento semelhante ao que seria destinado à exploração de Venus foi testado, previamente, sobre as florestas de Guatemala e Belize. O imageamento dessas florestas, com o uso de radar, revelou uma intrincada rede de linhas retas e curvas, previamente desconhecidas. Pela pesquisa de campo nessas áreas, descobriu-se que as linhas correspondiam a um sistema de antigos canais

mayas (250 antes de Cristo a 900 depois de Cristo). Dessa maneira, foi explicado o mistério de como os mayas mantinham uma sofisticada civilização de vários milhões de pessoas.

O objetivo central deste trabalho é propor e desenvolver ferramentas de síntese, processamento e análise de imagens, com ênfase no uso de metodologia e idéias estatísticas, para o melhor aproveitamento da informação provinda de imagens de radar de abertura sintética. Contudo, é importante frisar que as ferramentas desenvolvidas podem ser aplicadas, também, a sinais de radar (em um sentido um pouco mais amplo que inclui as imagens de radar), como é o caso de radares de tráfego aéreo.

A seguir descreve-se o plano da tese.

Capítulo 2: contém definições e exemplos de duas importantes técnicas de estimação clássicas: os métodos dos momentos e o método de máxima verossimilhança. O paradigma Bayesiano é definido e ilustrado com exemplos; alguns destes exemplos são adaptações do material apresentado em (11). Tanto a teoria como os exemplos são especializados ao problema de segmentação de imagens. Parte deste Capítulo foi desenvolvida em conjunto com Corina C. F. Yanasse.

Capítulo 3: apresenta a teoria e resultados a respeito da distribuição *a priori* que será empregada em boa parte desta tese: o modelo de Potts-Strauss. Neste Capítulo são definidos os campos Markovianos e são apresentados alguns resultados a respeito dos tempos de convergência dos algoritmos para a simulação de ocorrências do modelo de Potts-Strauss. Esses tempos de convergência são empregados para estimar quantidades no modelo de Ising, e para desenvolver técnicas de simulação de imagens SAR de uma ou mais visadas e detecção linear, bem como de imagens ópticas. Apresenta-se, ainda, um resultado importante para o uso de modelos Markovianos em processamento de imagens: uma extensão do teorema de Geman e Geman sobre a distribuição *a posteriori* de uma imagem. No mesmo Capítulo é apresentada uma técnica importante para

inferência nos campos Markovianos, especializada ao modelo de Potts-Strauss: a estimação por máxima pseudoverossimilhança.

Capítulo 4: mostra o uso da modelagem Markoviana no desenvolvimento de um algoritmo de segmentação contextual: o algoritmo ICM. Comentam-se as diversas etapas seguidas na implementação e avaliação desta técnica, até a sua integração a um Sistema de Processamento de Informações Georeferenciadas: o ambiente SPRING. Parte deste Capítulo foi desenvolvida em conjunto com Guaraci J. Erthal.

Capítulo 5: são enunciadas algumas propriedades da distribuição Rayleigh, visando a sua aplicação na estimação do parâmetro de escala. As diversas técnicas de estimação são aferidas com o uso de experiências Monte Carlo, e aplicadas a imagens reais. O problema de filtrar uma imagem SAR de uma visada e detecção linear é formulado em termos de inferência do parâmetro de escala e, assim, são desenvolvidos alguns filtros robustos para a redução de ruído *speckle* neste caso. Apresenta-se, a seguir, o uso destes filtros robustos em imagens reais de radar de abertura sintética, e avalia-se o desempenho dos filtros propostos. Parte deste Capítulo foi desenvolvida em conjunto com Sidnei J. S. Sant'Anna.

Capítulo 6: definem-se os testes Kolmogorov-Smirnov e χ^2 na forma em que serão empregados para discriminar diferentes ocupações do solo numa imagem da missão SAREX sobre a região de Tapajós. Fornecem-se as distribuições ajustadas e os seus parâmetros, por quatro critérios diferentes, a áreas visualmente classificadas como regiões de floresta e de não floresta. Estuda-se, com uma experiência Monte Carlo, a influência que o formato dos dados tem sobre a qualidade dos ajustes obtidos. Parte deste Capítulo foi desenvolvida em conjunto com Corina C. F. Yanasse e Sidnei J. S. Sant'Anna.

Apêndice A: apresenta-se a informação computacional relevante a respeito dos algoritmos de geração de números pseudoaleatórios empregados, e sobre os aplicativos e pacotes que possibilitaram o desenvolvimento e redação deste trabalho.

Apêndice B: encontram-se as listagens de alguns dos programas que implementam os algoritmos apresentados nesta tese.

CAPÍTULO 2

ANALOGIAS ENTRE TÉCNICAS DE INFERÊNCIA ESTATÍSTICA E PROBLEMAS NA ANÁLISE E NO PROCESSAMENTO DE IMAGENS

2.1 Introdução

As técnicas de processamento de imagens têm sofrido um crescimento explosivo nos últimos anos. Isto deve-se, entre outros motivos, ao fato de se dispor de um volume cada vez maior de imagens, provenientes de várias fontes. Sabe-se, por exemplo, que, até hoje, uma percentagem superior a 50% das imagens enviadas por satélites ambientais nunca foi vista por olhos humanos... e a grande maioria delas carrega informação preciosa.

Essa informação, quando recuperada dentro de prazos aceitáveis, está presente no cotidiano de todos nós: desde dados relevantes para a previsão do tempo, até estatísticas que influem em decisões políticas de concessão de créditos internacionais para as safras de uma nação.

Portanto, cada vez é mais importante poder processar imagens em forma rápida, e de tal forma que a informação procurada seja recuperada de forma precisa. É dentro deste contexto que as ferramentas estatísticas estão tendo um papel central. Outras ferramentas, como por exemplo as de Inteligência Artificial, de Morfologia Matemática, etc., também estão contribuindo ao desenvolvimento de técnicas para o processamento e a interpretação automática de imagens.

O imageamento em sensoriamento remoto por radar de abertura sintética (SAR) está tendo uma importância cada vez maior, devido aos vários satélites já lançados ou a serem lançados nesta década, além de missões aerotransportadas. Dar-se-á especial ênfase a esta classe de imageadores.

Uma tendência bastante atual é a de generalizar o processamento estatístico, e fazê-lo mais rico, lançando mão de, novamente, modelos e ferramentas que provêm da teoria de processos estocásticos. Mais adiante esta diferença deverá ficar mais clara; basta por enquanto citar, como exemplo, o problema de segmentar uma imagem: se tal segmentação for feita usando o método do paralelepípedo, tem-se um exemplo de técnica *clássica*, uma operação *estatística* típica seria obtê-la por máxima verossimilhança e, se a técnica empregada for o algoritmo ICM, tem-se um processamento *estocástico*.

Muitos dos assuntos que serão abordados, ou ainda formulam problemas em aberto (tanto do ponto de vista teórico como das aplicações), ou a sua aplicação está restrita a âmbitos experimentais.

2.2 Notação

Seja X uma variável aleatória real definida sobre o espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, onde Ω é o conjunto dos eventos, \mathcal{A} é uma σ -álgebra e \mathbb{P} é uma probabilidade; ver, por exemplo (32,76). Será denotada genericamente como uma função da forma $X: \Omega \rightarrow R$, com $R \subset \mathbb{R}$. Se X é discreta, e for conveniente, se escreverá $f_X(x) = \mathbb{P}(X = x)$. Se X admite densidade (variável de tipo contínuo) f_X denotará essa densidade. Denotar-se-á $F_X(x)$ a função de distribuição acumulada da variável aleatória X , isto é $F_X(x) = \int_{(-\infty, x]} f_X(y) dy$.

Procurar-se-á manter um único estilo de notação. Para tanto, toda vez que apareçam variáveis aleatórias, usar-se-ão maiúsculas; as suas ocorrências serão denotadas ou com minúsculas ou indicando que é um evento do espaço amostral. Isto é:

X denota uma variável aleatória, e
 x ou $X(\omega)$ denotam uma ocorrência de X .

Os vetores e as matrizes, sejam determinísticos ou aleatórios, serão escritos em negrito; para o caso anterior, e a sua correspondente ocorrência, teria-se escrito:

X denota um vetor de variáveis aleatórias, e

\mathbf{x} ou $\mathbf{X}(\omega)$ denotam uma ocorrência de \mathbf{X} .

No decorrer deste trabalho aparecerá com frequência $\mathbb{I}_A(x)$, a função indicadora do conjunto A ; isto é

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{caso contrário.} \end{cases}$$

A diferença entre os conjuntos A e B será denotada $A \setminus B$, isto é, $A \setminus B = B^c \cap A$.

O conjunto dos números reais positivos será denotado \mathbb{R}_+ , isto é, $\mathbb{R}_+ = (0, +\infty)$, \mathbb{N} denotará o conjunto dos números inteiros positivos, isto é, $\mathbb{N} = \{1, 2, \dots\}$ e, finalmente, \mathbf{Z} denotará o conjunto dos números inteiros, isto é, $\mathbf{Z} = \{\dots, -1, 0, 1, \dots\}$. A cardinalidade do conjunto A será denotado $\#A$, isto é, $\#A$ é o número de elementos pertencentes a A .

A função Gamma de Euler é definida por:

$$\Gamma(\nu + 1) = \int_{\mathbb{R}_+} t^\nu \exp(-t) dt, \text{ para todo } \nu > -1,$$

e vale que $\Gamma(k) = (k - 1)!$ se $k \in \mathbb{N}$. Outras propriedades úteis desta função podem ser vistas em (1,65). A função di-Gamma é dada, para todo $\nu > -1$, por:

$$\Psi(\nu) = \frac{d}{d\nu} \ln(\Gamma(\nu)) = \frac{\frac{d}{d\nu} \Gamma(\nu)}{\Gamma(\nu)}.$$

2.3 Inferência Clássica

Suponha-se, apenas por simplicidade notacional, que o parâmetro desconhecido é de dimensão um, isto é, $\boldsymbol{\theta} = \theta$. Considere-se o problema da estimação de $q(\theta)$, uma função do parâmetro desconhecido θ , baseado no vetor de observações. Suponha-se determinada a classe \mathcal{P} , utilizando hipóteses relativas ao problema em particular, e que θ é fixo. Esta última hipótese equivale a dizer que o único conhecimento que possui-se *a priori* sobre o parâmetro, é que ele pertence ao espaço paramétrico.

Dado que se dispõe da observação $\mathbf{Y}(\omega) = \mathbf{y}$, formar-se-á uma estimativa $\delta(\mathbf{y})$ de $q(\theta)$ segundo algum critério apropriado para cada caso. Em geral não é fácil determinar qual é o tal critério que melhor reflete as necessidades particulares: deseja-se que seja não viciado, isto é $\mathbb{E}(\delta(\mathbf{y})) = q(\theta)$, seja qual for a sua variância?, admitir-se-á algum vício na estimação desde que a variabilidade em torno do valor correto seja pequena? ou procurar-se-á alguma outra medida de adequação?

Serão citados apenas dois métodos de estimação que supõem fixo o valor desconhecido a ser estimado: o método dos momentos (ou método de substituição) e o método de máxima verossimilhança.

2.3.1 O Método dos Momentos

Seja o vetor de observações $\mathbf{y} = (y_1, \dots, y_n)$, cujas n componentes são ocorrências de variáveis aleatórias independentes, e identicamente distribuídas segundo uma distribuição dada por $F(y_i, \theta)$ para todo $1 \leq i \leq n$. Suponha-se também que essa distribuição possui momentos finitos de até, pelo menos, ordem $r \geq 1$. Denotar-se-á $m_j(\theta) = \int y^j dF(y, \theta)$ o j -ésimo momento de Y_i , com $j \leq r$. Para todo $1 \leq j$ o estimador para $m_j(\theta)$ será o j -ésimo momento amostral, ou seja $\widehat{m}_j(\theta) = n^{-1}(y_1^j + \dots + y_n^j)$. É possível ver (32) que, pela Lei dos Grandes Números $\widehat{m}_j(\theta) \rightarrow m_j(\theta)$ quase certamente quando $n \rightarrow \infty$.

Se a função que se quer estimar, $q(\theta)$, depende dos momentos de até ordem r , isto é $q(\theta) = h^{(d)}(m_1(\theta), \dots, m_r(\theta))$, $h^{(d)}$ contínua, então um estimador de $q(\theta)$ pelo método dos momentos é $\widehat{q}(\theta) = h^{(d)}(\widehat{m}_1(\theta), \dots, \widehat{m}_r(\theta))$ e, pelo mesmo argumento que anteriormente (11), vale que $\widehat{q}(\theta) \rightarrow q(\theta)$ quase certamente quando $n \rightarrow \infty$. O super-índice “ (d) ” indica a possibilidade de formar mais de um estimador, para a mesma função q , com os mesmos momentos de até ordem r . Para todo d , a função $h^{(d)}$ é chamada *estimador para $q(\theta)$ obtido pelo método dos momentos* ou, em forma abreviada, *EMO*.

Exemplo 2.1 Sejam as observações provindas de variáveis aleatórias independentes e identicamente distribuídas com distribuição Poisson de pa-

râmetro λ , $\lambda \in \Theta = \mathbb{R}_+$, isto é $\text{IPr}(Y_i = k) = \exp(-\lambda) \frac{\lambda^k}{k!}$; assim sendo vale que $m_1 = \lambda$ e $m_2 = \lambda + \lambda^2$ e, portanto, podem ser formados os seguintes estimadores para λ :

$$\begin{aligned} m_1 = \lambda &\Rightarrow \widehat{\lambda^{(1)}} = \bar{y} = \frac{1}{n} \sum_{1 \leq i \leq n} y_i \\ m_2 - m_1^2 = \lambda &\Rightarrow \widehat{\lambda^{(2)}} = \frac{1}{n} \sum_{1 \leq i \leq n} (y_i - \bar{y})^2 \\ \sqrt{m_2 - m_1} = \lambda &\Rightarrow \widehat{\lambda^{(3)}} = \sqrt{\frac{y_1^2 + \dots + y_n^2}{n} - \bar{y}}. \end{aligned}$$

2.3.2 O Método da Máxima Verossimilhança

Este método define, como melhor estimador, aquele que maximiza a probabilidade de obter o vetor de observações \mathbf{y} . O estimador de máxima verossimilhança (EMV) de $q(\theta)$, que será denotado $\widehat{q(\theta)}$, é definido como:

$$\widehat{q(\theta)} = q\left(\sup_{\theta \in \Theta}^{-1} \{\text{IPr}_{\theta}(\mathbf{y})\}\right).$$

Estes estimadores, ou pequenas variações deles, são muito usados devido à (relativa) facilidade com que podem ser obtidos para uma classe muito larga de problemas estatísticos. Além disso, eles possuem boas propriedades assintóticas (eficiência, consistência e normalidade), o que assegura o sucesso desta técnica de estimação quando o tamanho da amostra cresce indefinidamente.

Partilham com os estimadores obtidos pelo método dos momentos a desvantagem de não ser simples a adição de informação prévia ao experimento estatístico. A função $\widehat{q(\theta)}$ é chamada *estimador para $q(\theta)$ obtido pelo método de máxima verossimilhança* ou, em forma abreviada, *EMV*.

Exemplo 2.2 Depois de observar durante anos que, quase sempre que chegamos ao elevador do nosso andar, ele está em um outro piso, decidimos estimar a probabilidade de encontrá-lo ao sair de casa. Para isso coletamos as n observações Bernoulli y_1, \dots, y_n , escrevendo “1” toda vez que o elevador está

em um andar diferente do nosso e “0” caso contrário. Supondo que os dados são uma amostra de tamanho n de variáveis aleatórias independentes identicamente distribuídas com distribuição $\mathcal{B}(\theta)$, $\theta \in \Theta = [0, 1]$, (equação (2.1)) o EMV e o EMO para θ são iguais a: $\hat{\theta} = \bar{y}$ mas, evidentemente, eles não incorporam a nossa experiência prévia. Tal experiência prévia poderia ser incorporada redefinindo $\Theta = [1/2, 1]$, obtendo neste caso um novo EMV para θ dado por $\hat{\theta} = \text{máx}\{1/2, \bar{y}\}$, o que não deixa de ser muito arbitrário pela escolha do valor “1/2”.

2.3.3 Inferência Bayesiana

Os métodos Bayesianos são uma forma de introduzir, em um problema estatístico, um grau de conhecimento *a priori*; desta forma os conceitos clássicos (ou relacionados aos EMVs) já apresentados são generalizados (isto é, a estimação por máxima verossimilhança é um caso particular da estimação Bayesiana: aquele onde a informação *a priori* incorporada é *não informativa*). Também permitem, em princípio, a obtenção de estimadores ótimos no sentido de *custo* mínimo. Para detalhes sobre estes assuntos ver (10). Os problemas de segmentação e de estimação serão apresentados no mesmo formalismo.

Estes métodos consideram que os parâmetros a serem estimados são ocorrências de variáveis aleatórias, que obedecem uma distribuição conhecida *a priori*. Dada(s) a(s) ocorrência(s) desta(s) variável(eis) aleatória(s), a distribuição das observações, sobre as quais será baseada a inferência, fica determinada: tem-se a distribuição *a posteriori*, modificando desta forma a opinião sobre o verdadeiro valor dos parâmetros.

Os problemas de estimação, neste contexto, são formulados dizendo que dado que a variável aleatória Y toma o valor (observado) y , quer-se estimar o valor θ (não observado) adotado pela variável aleatória Θ , sendo que ambas estão relacionadas por uma distribuição conjunta $f_{\Theta, Y}(\theta, y)$. Conhecendo as densidades $f_{Y|\Theta=\theta}(y | \Theta = \theta)$ e $f_{\Theta}(\theta)$, e usando a fórmula de Bayes $f_{\Theta|Y=y}(\theta | Y = y) \propto f_{Y|\Theta=\theta}(y | \Theta = \theta) \cdot f_{\Theta}(\theta)$, pode-se usar $f_{\Theta|Y=y}(\theta | Y = y)$ para obter a regra de estimação $x = \delta(y)$ que minimize um determinado critério de penalidade.

O custo antes indicado é uma função $C(\theta, x)$, por definição não negativa. Um critério de penalidade muito empregado é o custo médio —ou *risco*— dado por $R(x | Y = y) = \mathbb{E}(C(\theta, \delta(Y)) | Y = y)$. Será o adotado neste trabalho.

Generalizando, o problema agora consiste em: dados Ξ , o conjunto das estimações (ou classes) possíveis; $f_{\Theta}(\theta)$, a distribuição *a priori*; $f_{\mathbf{Y}|\Theta=\theta}(\mathbf{Y} | \Theta = \theta)$, a distribuição das observações dado que $\Theta(\omega) = \theta$; e a função de custo $C(\theta, x)$, achar a regra de estimação (ou de segmentação) δ^* que minimize o risco *a posteriori*. Isto é, achar

$$\delta^* = \inf_{\delta}^{-1} \mathbb{E}(C(\Theta, \delta(\mathbf{Y})) | \mathbf{Y} = \mathbf{y}).$$

Se a regra δ^* existir ela é chamada *regra de Bayes* para o problema sob consideração, pois o seu risco total $\mathbb{E}(\mathbb{E}(C(\Theta, \delta(\mathbf{Y})) | \mathbf{Y}))$ é inferior ao obtido com qualquer outra regra δ .

Exemplo 2.3 Voltando ao elevador do Exemplo 2.2, o conhecimento *a priori* será modelado segundo uma distribuição Beta. Desta forma pode-se dar mais verossimilhança ao fato do parâmetro ter valores perto do valor “1”. Estimar a probabilidade de sucesso baseado em n ensaios binomiais; seja $\mathbf{Y} = (Y_1, \dots, Y_n)$ uma seqüência de variáveis aleatórias independentes identicamente distribuídas com distribuição $\mathcal{B}(\theta)$, Θ com distribuição $\beta(r, s)$ (equação (2.3)), o custo dado por $C(\theta, x) = \frac{(\theta-x)^2}{\theta(1-\theta)}$ e a estatística suficiente $S = \sum_{i=1}^n Y_i$. O risco *a posteriori* é, para $x \in \Xi = (0, 1)$,

$$\begin{aligned} R(x | S = k) &= \mathbb{E}\left(\frac{(\Theta - x)^2}{\Theta(1 - \Theta)} \mid S = k\right) = \\ &= \mathbb{E}\left(\frac{\Theta}{1 - \Theta} \mid S = k\right) - 2x\mathbb{E}\left(\frac{1}{1 - \Theta} \mid S = k\right) + x^2\mathbb{E}\left(\frac{1}{\Theta(1 - \Theta)} \mid S = k\right). \end{aligned}$$

Minimizando em relação a x obtém-se a regra de Bayes dada por:

$$\delta^*(k) = \frac{\mathbb{E}((1 - \Theta)^{-1} | S = k)}{\mathbb{E}([\Theta(1 - \Theta)]^{-1} | S = k)}.$$

Por hipótese, supõe-se que $f_{\Theta}(t) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} t^{r-1} (1-t)^{s-1} \mathbb{1}_{(0,1)}(t)$. Pode-se ver que a distribuição *a posteriori* tem a forma seguinte:

$$f_{\Theta|S=k}(t | S = k) \propto t^{k+r-1} (1-t)^{n-k+s-1},$$

logo $\Theta \mid S = k$ tem uma distribuição $\beta(k + r, n - k + s)$. Finalmente tem-se que $\delta^*(k) = \frac{k+r-1}{n+s+r-2}$. Note-se que para $r = s = 1$ tem-se que Θ obedece a uma distribuição $\mathcal{U}[0, 1]$ e que $\delta^*(k) = k/n = \bar{y}$: para uma distribuição *a priori* não informativa o estimador Bayesiano reduz-se ao EMV. Resta salientar que a escolha particular de $C(\theta, x)$ teve por principal motivação a simplificação das contas. A Figura 2.1 é um exemplo do que poderia ser a distribuição *a priori* para este problema.

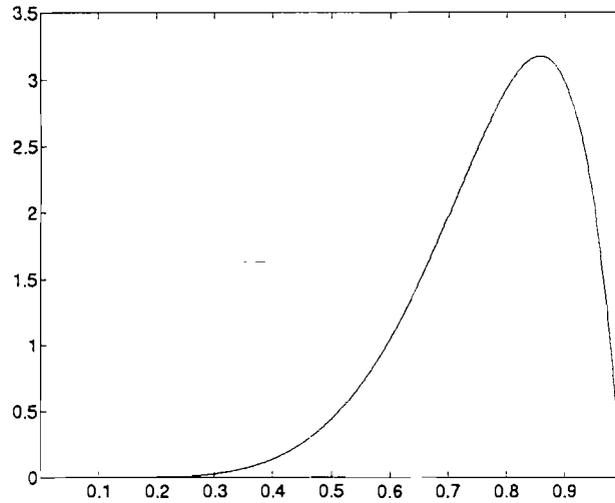


Fig. 2.1 - Densidade $\beta(7, 2)$.

Em muitos problemas pode-se supor que os conjuntos Ξ e Θ são de cardinalidade finita. Assim sendo, sejam dados os conjuntos $\Theta = \{\theta_1, \dots, \theta_K\}$ e $\Xi = \{x_1, \dots, x_q\}$, os reais não negativos $C(\theta_i, x_j) = \lambda_{ij}$, a distribuição $f_{\Theta}(\theta) = \mathbb{P}\text{r}(\Theta = \theta_i) = \pi_i$ valendo $\pi_i \geq 0$ para todo i e $\sum_{i=1}^K \pi_i = 1$, e a densidade $f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y} \mid \Theta = \theta)$. Então pode-se calcular a distribuição *a posteriori*:

$$\mathbb{P}\text{r}(\Theta = \theta_i \mid \mathbf{Y} = \mathbf{y}) = \frac{\pi_i f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y} \mid \Theta = \theta_i)}{\sum_{1 \leq j \leq K} \pi_j f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y} \mid \Theta = \theta_j)},$$

e o risco *a posteriori* será dado por:

$$R(x_j \mid \mathbf{Y} = \mathbf{y}) = \mathbb{E}(C(\Theta, x_j) \mid \mathbf{Y} = \mathbf{y}) = \frac{\sum_{1 \leq i \leq K} \lambda_{ij} \pi_i f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y} \mid \Theta = \theta_i)}{\sum_{1 \leq i \leq K} \pi_i f_{\mathbf{Y}|\Theta=\theta}(\mathbf{y} \mid \Theta = \theta_i)}.$$

A regra ótima $\delta^*(\mathbf{Y})$ deve satisfazer $R(\delta^*(\mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}) = \min_{1 \leq j \leq q} R(x_j \mid \mathbf{Y} = \mathbf{y})$.

Exemplo 2.4 O problema de classificação do padrão \mathbf{y} em uma entre K classes pode colocar-se dentro do contexto do caso anterior. Considere-se o caso $K = q$, $\theta_i = x_i$ para todo $1 \leq i \leq K$ e seja $\lambda_{ij} = \mathbb{I}_{\{j \neq i\}}(i)$; assim sendo, minimizar $R(\theta_i \mid \mathbf{y}) = \mathbb{P}\text{r}(\Theta \neq \theta_i \mid \mathbf{Y} = \mathbf{y})$ é equivalente a maximizar a probabilidade *a posteriori*:

$$\mathbb{P}\text{r}(\Theta = \theta_i \mid \mathbf{Y} = \mathbf{y}) = \frac{\pi_i f_{\mathbf{Y} \mid \Theta = \theta_i}(\mathbf{y} \mid \Theta = \theta_i)}{\sum_{1 \leq j \leq K} \pi_j f_{\mathbf{Y} \mid \Theta = \theta_j}(\mathbf{y} \mid \Theta = \theta_j)}.$$

2.4 Algumas Distribuições Importantes e seus Estimadores

Nesta Seção são lembradas as principais distribuições que aparecerão no decorrer deste trabalho. As mesmas serão especificadas pelas respectivas densidades e funções de distribuição acumuladas, sempre que possível. Para escrever os estimadores obtidos pelos métodos dos momentos e de máxima verossimilhança, em todos os casos, foi suposto que se dispõe de um vetor de observações independentes e identicamente distribuídas de tamanho n . Sempre que possível será omitida a distinção entre estimadores obtidos pelo método de máxima verossimilhança ($\hat{\theta}$) e pelo método dos momentos ($\widehat{\theta}^{(j)}$).

Distribuição 2.1 Dir-se-á que a variável aleatória discreta X tem distribuição Binomial com parâmetros $p \in [0, 1]$ e $q \in \mathbb{N} \setminus \{0\}$ (em símbolos $X \sim \mathcal{B}i(p, q)$) se, para todo $x \in \mathbb{R}$ vale que:

$$\mathbb{P}\text{r}(X = x) = f_X(x) = \binom{q}{x} p^x (1-p)^{q-x} \mathbb{I}_{\{0, 1, \dots, q\}}(x). \quad (2.1)$$

Vale que $\mathbb{E}(X) = pq$ e $\text{Var}(X) = pq(1-p)$. Escrever-se-á $\mathcal{B}(p, 1)$ como $\mathcal{B}(p)$. Se $X \sim \mathcal{B}(p)$, então os estimadores de p obtidos pelo método de máxima verossimilhança e pelo primeiro momento são iguais à média amostral, isto é, $\hat{p} = \widehat{p}^{(1)} = \widehat{m}_1$.

Distribuição 2.2 A variável aleatória U é dita ter distribuição uniforme no conjunto $A \subset \mathbb{R}^d$ (em símbolos $U \sim \mathcal{U}(A)$) se tem densidade dada por:

$$f_U(\mathbf{u}) = \frac{1}{\int_A d\mathbf{u}} \mathbb{I}_A(\mathbf{u}) \text{ para todo } \mathbf{u} \in \mathbb{R}^d.$$

Um caso que aparecerá com muita frequência neste trabalho é aquele em que o conjunto A é um intervalo (aberto, fechado ou semiaberto) da reta. Sejam então os pontos da reta $-\infty < a < b < +\infty$; U é dita ter distribuição uniforme no intervalo (a, b) (ou, equivalentemente, nos intervalos $(a, b]$, $[a, b)$ ou $[a, b]$) se:

$$f_U(u) = \frac{1}{b-a} \mathbb{I}_{(a,b)}(u), \quad \forall x \in \mathbb{R}.$$

Neste caso $\mathbb{E}(U) = (b-a)/2$, $\text{Var}(U) = (b-a)^2/12$ e

$$F_U(u) = \begin{cases} 0 & \text{se } u < a \\ \frac{u-a}{b-a} & \text{se } a \leq u \leq b \\ 1 & \text{se } u > b. \end{cases}$$

Se $U \sim \mathcal{U}(0, \theta)$ tem-se que $\hat{\theta} = \max_{1 \leq i \leq n} \{u_i\}$ e que $\widehat{\theta^{(1)}} = 2\widehat{m}_1$.

Distribuição 2.3 A variável aleatória X é dita ter distribuição Normal com média $\mu \in \mathbb{R}$ e variância $\sigma^2 > 0$ (em símbolos $X \sim \mathcal{N}(\mu, \sigma^2)$) se tem densidade dada por:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \text{ para todo } x \in \mathbb{R}.$$

Se $X \sim \mathcal{N}(0, 1)$ então se dirá que X tem distribuição normal padrão. A função de distribuição acumulada de $X \sim \mathcal{N}(\mu, \sigma^2)$ é dada, para todo $x \in \mathbb{R}$, por

$$F_X(x) = \Phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(\lambda-\mu)^2\right) d\lambda.$$

Distribuição 2.4 Dir-se-á que um vetor aleatório tem distribuição Normal multivariada se possui a mesma distribuição de uma transformação afim de normais padrão independentes. Isto é, se X_1, \dots, X_k são independentes com distribuição comum $\mathcal{N}(0, 1)$, então o vetor aleatório $\mathbf{Y} = (Y_1, \dots, Y_k)$ com

$$Y_j = a_{1j}X_1 + \dots + a_{kj}X_k + \mu_j \quad \forall a_{ij}, \mu_j \in \mathbb{R}, \forall j = 1, \dots, k$$

possui distribuição normal k -variada. Outra forma de escrever a equação anterior, mais compacta e que será muito útil no decorrer deste trabalho, é

$$\mathbf{Y} = \mathbf{X}A + \boldsymbol{\mu}, \text{ com } A = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix} \text{ e } \boldsymbol{\mu} = (\mu_1, \dots, \mu_k). \quad (2.2)$$

Denotar-se-á a distribuição acima $\mathbf{Y} \sim \mathcal{N}k(\boldsymbol{\mu}, \Sigma)$, sendo que Σ é chamada matriz de covariância de \mathbf{Y} , e é dada por:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1k} \\ \vdots & & \vdots \\ \sigma_{k1} & \dots & \sigma_{kk} \end{pmatrix}, \text{ valendo } \sigma_{ij} = \text{Cov}(Y_i, Y_j) = (A^T A)_{ij}.$$

Em (76) pode ver-se que uma matriz $k \times k$ é matriz de covariâncias de algum vetor normal k -variado se, e somente se, é definida não negativa. Isto é, $\mathbf{Y} \sim \mathcal{N}k(\boldsymbol{\mu}, \Sigma)$ se e somente se existe $A_{k,k}$ tal que existe $A_{k,k}^{-1}$ e tal que $(\mathbf{Y} - \boldsymbol{\mu})A_{k,k}^{-1}$ é um vetor de variáveis aleatórias independentes com distribuição $\mathcal{N}(0, 1)$. A densidade de $\mathbf{Y} \sim \mathcal{N}k(\boldsymbol{\mu}, \Sigma)$ é dada, para todo $\mathbf{y} \in \mathbb{R}^k$, por:

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-k/2} (|\Sigma|)^{-1/2} \exp((\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})).$$

Os estimadores de máxima verossimilhança são:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{n} \sum_{1 \leq i \leq n} \mathbf{y}_i, \text{ e} \\ \hat{\Sigma} &= \frac{1}{n} \sum_{1 \leq i \leq n} (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})^T. \end{aligned}$$

Distribuição 2.5 Seja $Y \sim \mathcal{N}(\mu, \sigma^2)$; a variável aleatória $X = \exp(Y)$ é dita ter distribuição Log-Normal com parâmetros μ e σ^2 , e denota-se $X \sim \mathcal{LN}(\mu, \sigma^2)$. A sua densidade é dada por:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right) \mathbb{I}_{\mathbb{R}_+}(x)$$

e a sua função de distribuição acumulada é dada, para todo $x \in \mathbb{R}_+$, por $F_X(x) = \Phi_{\mu, \sigma}(\ln(x))$. Os estimadores de μ e σ^2 , pelo método dos momentos são dados por:

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{1 \leq i \leq n} \ln(x_i), \\ \hat{\sigma} &= \frac{1}{n} \sum_{1 \leq i \leq n} (\ln(x_i) - \hat{\mu})^2. \end{aligned}$$

Distribuição 2.6 A variável aleatória X é dita ter distribuição Beta com parâmetros $\alpha > 0$ e $\beta > 0$ se tem densidade dada, para todo $x \in \mathbb{R}$, por:

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{I}_{(0,1)}(x). \quad (2.3)$$

Portanto, a função de distribuição acumulada é dada, para todo $x \in \mathbb{R}$, por:

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt & \text{se } 0 < x < 1 \\ 1 & \text{se } x \geq 1. \end{cases}$$

Neste caso vale que $\mathbb{E}(X) = \alpha/(\alpha + \beta)$ e $\text{Var}(X) = \alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$. Será empregada a notação $X \sim \beta(\alpha, \beta)$. Ver Figura 2.1. Os estimadores para α e β , pelo método dos momentos, são dados por:

$$\begin{aligned} \hat{\alpha} &= \frac{\widehat{m}_1(\widehat{m}_1 - \widehat{m}_2)}{\widehat{m}_2 - \widehat{m}_1^2} \\ \hat{\beta} &= \frac{(\widehat{m}_1 - 1)(\widehat{m}_2 - \widehat{m}_1)}{\widehat{m}_2 - \widehat{m}_1^2}. \end{aligned}$$

Distribuição 2.7 A variável aleatória X é dita ter distribuição Weibull com parâmetros $\alpha > 0$ e $\beta > 0$ se tem densidade dada, para todo $x \in \mathbb{R}$, por:

$$f_X(x) = \alpha\beta^\alpha x^{\alpha-1} \exp(-(\beta x)^\alpha) \mathbb{I}_{\mathbb{R}_+}(x).$$

Logo, a função de distribuição acumulada é dada, para todo $x \in \mathbb{R}$, por:

$$F_X(x) = [1 - \exp(-(\beta x)^\alpha)] \mathbb{I}_{\mathbb{R}_+}(x).$$

Denota-se, neste caso, $X \sim \mathcal{W}(\alpha, \beta)$, e tem-se que $\mathbb{E}(X^m) = \Gamma(m\alpha^{-1} + 1)/\beta^m$. Os estimadores pelo método dos momentos são obtidos resolvendo-se o sistema de equações:

$$\begin{cases} \Gamma(\hat{\alpha}^{-1} + 1) - \hat{\beta}\widehat{m}_1 = 0 \\ \Gamma(2\hat{\alpha}^{-1} + 1) - \hat{\beta}^2\widehat{m}_2 = 0, \end{cases}$$

ou:

$$\begin{cases} \Gamma(2\hat{\alpha}^{-1} + 1) - \Gamma(\hat{\alpha}^{-1} + 1)^2 \widehat{m}_2 \widehat{m}_1^{-2} = 0 \\ \hat{\beta} - \Gamma(\hat{\alpha}^{-1} + 1) \widehat{m}_1^{-1} = 0. \end{cases}$$

Distribuição 2.8 Seja θ um número real positivo; chama-se de função de distribuição acumulada de uma variável aleatória Rayleigh com parâmetro θ (e se denotará essa variável aleatória como $X \sim \mathcal{R}(\theta)$) à função $F_\theta: \mathbb{R} \rightarrow [0, 1]$ definida por:

$$F_\theta(y) = \left[1 - \exp\left(-\frac{1}{2\theta^2}y^2\right) \right] \mathbb{I}_{\mathbb{R}_+}(y). \quad (2.4)$$

É fácil ver, derivando a equação anterior em relação a y , que F_θ tem densidade dada, para todo $x \in \mathbb{R}$, por:

$$f_\theta(y) = \frac{y}{\theta^2} \exp\left(-\frac{1}{2\theta^2}y^2\right) \mathbb{I}_{\mathbb{R}_+}(y). \quad (2.5)$$

Se $X \sim \mathcal{R}(\theta)$, então tem-se que:

$$\mathbb{E}(X) = \sqrt{\frac{\pi}{2}} \theta \text{ e } \text{Var}(X) = \left(2 - \frac{\pi}{2}\right) \theta^2. \quad (2.6)$$

Daí tem-se que:

$$\widehat{\theta^{(1)}} = \widehat{m}_1 \sqrt{2/\pi}, \quad (2.7)$$

$$\widehat{\theta} = \sqrt{\frac{1}{2n} \sum_{1 \leq i \leq n} y_i^2}. \quad (2.8)$$

Esta distribuição, pelo seu interesse para a modelagem de dados SAR, será estudada com maiores detalhes no Capítulo 5, pois é um caso particular da Distribuição 5.1 (ver página 123). Na literatura sobre SAR, muitas vezes se faz referência ao parâmetro θ como *backscatter*, e denota-se como σ_0 .

Distribuição 2.9 A variável aleatória X é dita ter distribuição Exponencial com parâmetro $\theta > 0$ (em símbolos $X \sim \mathcal{E}(\theta)$) se tem densidade e, portanto, função de distribuição acumulada dadas, para todo $x \in \mathbb{R}$, por:

$$f_X(x) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) \mathbb{I}_{\mathbb{R}_+}(x) \quad (2.9)$$

$$F_X(x) = \left[1 - \exp\left(-\frac{x}{\theta}\right) \right] \mathbb{I}_{\mathbb{R}_+}(x), \quad (2.10)$$

respectivamente. É imediato verificar que $\mathbb{E}(X^m) = \theta^m \Gamma(m + 1)$, e daí $\mathbb{E}(X) = \theta$ e $\text{Var}(X) = \theta^2$. Daí, tem-se que os estimadores obtidos pelos métodos de máxima verossimilhança e pela substituição do primeiro momento coincidem: $\widehat{\theta} = \widehat{\theta^{(1)}} = \widehat{m}_1$.

A seguir serão apresentadas as distribuições K de amplitude e intensidade para múltiplas visadas. O uso destas distribuições para análise de imagens de radar de abertura sintética tem sido bastante restrito na literatura devido, possivelmente, a que as suas funções de distribuição acumuladas, para mais de uma visada, não eram conhecidas em formas computacionalmente tratáveis.

Distribuição 2.10 A variável aleatória X é dita ter distribuição K de Intensidade e parâmetros $\alpha > 0$, $\beta > 0$ e $n > 0$ (em símbolos $X \sim \mathcal{KIn}(\alpha, \beta)$), se a sua densidade é dada, para todo $x \in \mathbb{R}$, por:

$$f_X(x) = \frac{2\alpha n}{\Gamma(\alpha)\beta\Gamma(n)} \left(\frac{\alpha n x}{\beta}\right)^{\frac{\alpha+n}{2}-1} K_{\alpha-n} \left(2\sqrt{\frac{\alpha n x}{\beta}}\right) \mathbb{I}_{\mathbb{R}_+}(x).$$

onde K_ν é a função de Bessel modificada do terceiro tipo. É possível ver que

$$\mathbb{E}(X^m) = \left(\frac{\beta}{\alpha n}\right)^m \frac{\Gamma(m+n)\Gamma(m+\alpha)}{\Gamma(\alpha)\Gamma(n)},$$

e daí é imediato que $\mathbb{E}(X) = \beta$ e que $\text{Var}(X) = \beta^2(1 + \alpha + n)/(\alpha n)$. A função de distribuição acumulada é dada, para todo $x \in \mathbb{R}$, por:

$$F_X(x) = \frac{2^{2-\alpha-n}}{\Gamma(\alpha)\Gamma(n)} \int_0^{2\sqrt{\frac{\alpha n x}{\beta}}} t^{\alpha+n-1} K_{\alpha-n}(t) dt \mathbb{I}_{\mathbb{R}_+}(x).$$

Para se ter uma forma computacionalmente tratável desta função de distribuição acumulada, é necessário impor que $n \in \mathbb{N}$. Nas aplicações de SAR, este parâmetro discreto costuma chamar-se *número de visadas*, e está relacionado ao tipo de processamento da imagem. Assim sendo, e escrevendo:

$$\int_0^{2\sqrt{\frac{\alpha n x}{\beta}}} t^{\alpha+n-1} K_{\alpha-n}(t) dt = f(\nu, k, z),$$

onde $\nu = \alpha - n$, $k = 2n - 1$ e $z = 2\sqrt{\frac{\alpha n x}{\beta}}$, e usando as relações apresentadas em (1), pág. 483, tem-se que:

$$f(\nu, k, z) = \begin{cases} 2^\nu \Gamma(\nu + 1) - z^{\nu+1} K_{\nu+1}(z) & \text{se } k = 1, \\ (k-1)(2\nu + k - 1)g(\nu, k-2, z) + \\ - z^{\nu+k} K_{\nu+1}(z) - (k-1)z^{\nu+k-1} K_\nu(z) & \text{caso contrário.} \end{cases} \quad (2.11)$$

Dado que, na equação (2.11), o argumento da função Γ pode assumir valores no conjunto $\{0, -1, -2, \dots\}$, optou-se pela seguinte forma que evita esse problema (140):

$$F_X(x) = 1 + \frac{2^{2-\alpha-n}}{\Gamma(\alpha)\Gamma(n)}g(\nu, k, z), \quad (2.12)$$

onde:

$$g(\nu, k, z) = \begin{cases} -z^{\nu+1}K_{\nu+1}(z) & \text{se } k = 1, \\ (k-1)(2\nu+k-1)f(\nu, k-2, z) + \\ -z^{\nu+k}K_{\nu+1}(z) - (k-1)z^{\nu+k-1}K_{\nu}(z) & \text{caso contrário.} \end{cases}$$

Na literatura consultada não foi encontrada nenhuma solução ao problema de escrever esta função de distribuição acumulada.

O estimador de β pelo método dos momentos é dado por $\hat{\beta} = \hat{m}_1$. Para se obter os estimadores de α e de n é necessário resolver um sistema de equações, que resulta em $\hat{n} = (-B \pm \sqrt{C})/(2A)$, onde:

$$\begin{aligned} A &= 2\hat{m}_2^2 - \hat{m}_1^2\hat{m}_2 - \hat{m}_1\hat{m}_3, \\ B &= 4\hat{m}_2^2 - 3\hat{m}_1^2\hat{m}_2 - \hat{m}_1\hat{m}_3, \\ C &= \hat{m}_1^4\hat{m}_2^2 - 8\hat{m}_1^2\hat{m}_2^3 - 2\hat{m}_1^3\hat{m}_2\hat{m}_3 + 16\hat{m}_2^4 - 8\hat{m}_1\hat{m}_2^2\hat{m}_3 + \hat{m}_1^2\hat{m}_3^2; \end{aligned}$$

e:

$$\hat{\alpha} = \frac{\hat{m}_1^2(\hat{n} + 1)}{\hat{n}(\hat{m}_2 - \hat{m}_1^2) - \hat{m}_1^2}.$$

Distribuição 2.11 Um caso particular da distribuição $X \sim \mathcal{KIn}(\alpha, \beta)$ é quando $n = 1$. Para esta situação tem-se que:

$$\begin{aligned} f_X(x) &= \frac{2\alpha}{\Gamma(\alpha)\beta} \left(\frac{\alpha x}{\beta}\right)^{\frac{\alpha-1}{2}} K_{\alpha-1}\left(2\sqrt{\frac{\alpha x}{\beta}}\right) \mathbb{I}_{\mathbb{R}_+}(x), \\ F_X(x) &= \left[1 - \frac{2}{\Gamma(\alpha)} \left(\sqrt{\frac{\alpha x}{\beta}}\right)^{\alpha} K_{\alpha}\left(2\sqrt{\frac{\alpha x}{\beta}}\right)\right] \mathbb{I}_{\mathbb{R}_+}(x). \end{aligned}$$

Os estimadores obtidos pelo método de máxima verossimilhança (28) são os valores de $\hat{\alpha}$ e $\hat{\beta}$ que satisfazem o sistema de equações dado por:

$$\left\{ \begin{array}{l} (\hat{\alpha} + 1) \sqrt{\frac{\hat{\alpha}}{\hat{\beta}}} - \frac{1}{n} \sum_{1 \leq i \leq n} \sqrt{x_i} \frac{K_{\hat{\alpha}}\left(2\sqrt{\frac{\hat{\alpha}x_i}{\hat{\beta}}}\right) + K_{\hat{\alpha}-2}\left(2\sqrt{\frac{\hat{\alpha}x_i}{\hat{\beta}}}\right)}{K_{\hat{\alpha}-1}\left(2\sqrt{\frac{\hat{\alpha}x_i}{\hat{\beta}}}\right)} = 0 \\ \Psi(\hat{\alpha}) + \ln\left(\frac{\hat{\alpha}}{\hat{\beta}}\right) - \frac{1}{n} \sum_{1 \leq i \leq n} \frac{\frac{\partial}{\partial \alpha} K_{\hat{\alpha}-1}\left(2\sqrt{\frac{\hat{\alpha}x_i}{\hat{\beta}}}\right)}{K_{\hat{\alpha}-1}\left(2\sqrt{\frac{\hat{\alpha}x_i}{\hat{\beta}}}\right)} = 0, \end{array} \right. \quad (2.13)$$

onde Ψ é a função di-Gamma. Os estimadores obtidos pelo método dos momentos são dados por: $\widehat{\beta}^{(1)} = \widehat{m}_1$ e $\widehat{\alpha}^{(1)} = 2\widehat{m}_1^2/(\widehat{m}_2 - 2\widehat{m}_1^2)$. Resulta interessante comparar estes estimadores com os apresentados no sistema de equações (2.13).

Distribuição 2.12 A variável aleatória X é dita ter distribuição \mathcal{K} de Amplitude com parâmetros $\alpha > 0$, $\beta > 0$ e $n > 0$ (em símbolos $X \sim \mathcal{K}An(\alpha, \beta)$), se a sua densidade é:

$$f_X(x) = \frac{4x\alpha n}{\Gamma(\alpha)\beta\Gamma(n)} \left(\frac{\alpha n x^2}{\beta}\right)^{\frac{\alpha+n-2}{2}} K_{\alpha-n}\left(2x\sqrt{\frac{\alpha n}{\beta}}\right) \mathbb{I}_{\mathbb{R}_+}(x).$$

A função de distribuição acumulada desta variável aleatória é dada por $F_X(x^2)$, com F_X dada na equação (2.12). Os momentos de ordem m são

$$\mathbb{E}(X^m) = \left(\frac{\beta}{\alpha n}\right)^{\frac{m}{2}} \frac{\Gamma(\alpha + \frac{m}{2})\Gamma(n + \frac{m}{2})}{\Gamma(\alpha)\Gamma(n)}.$$

Os estimadores de α , β e n , pelo método dos momentos, são dados por $\beta = \widehat{m}_2$ e pela solução do seguinte sistema de equações:

$$\left\{ \begin{array}{l} f_1(\alpha, n) = \sqrt{\frac{\widehat{m}_2}{\alpha n}} \frac{\Gamma(\alpha + \frac{1}{2})\Gamma(n + \frac{1}{2})}{\Gamma(\alpha)\Gamma(n)} - \widehat{m}_1 = 0 \\ f_2(\alpha, n) = \left(\frac{\widehat{m}_2}{\alpha n}\right)^{\frac{3}{2}} \frac{\Gamma(\alpha + \frac{3}{2})\Gamma(n + \frac{3}{2})}{\Gamma(\alpha)\Gamma(n)} - \widehat{m}_3 = 0. \end{array} \right. \quad (2.14)$$

Os estimadores $(\hat{\alpha}, \hat{n})$ podem ser obtidos, alternativamente, pela minimização da função $f_1^2(\alpha, n) + f_2^2(\alpha, n)$. Foi analisado o comportamento desta nova função, para uma amostra de dados SAR amplitude e várias visadas. As Figuras 2.2 e 2.3 apresentam o gráfico desta função em dois domínios de interesse para

a procura do mínimo global e para a determinação de $(\hat{\alpha}, \hat{n})$. Constata-se, assim, a existência de múltiplos mínimos locais, o que dificulta notavelmente a localização do mínimo global (caso ele exista).

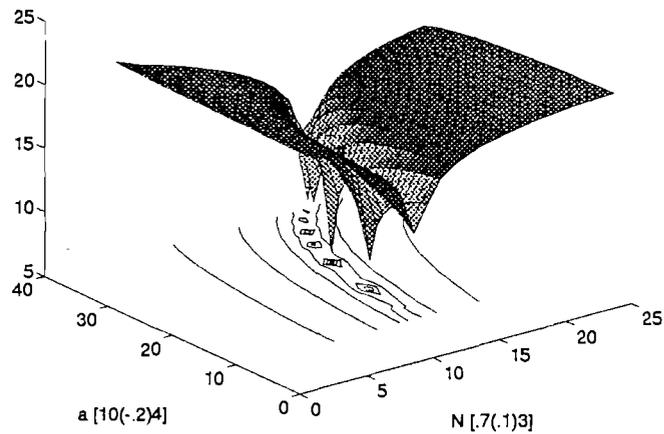


Fig. 2.2 - Função a ser minimizada para a estimação de parâmetros da distribuição $\mathcal{K}An$.

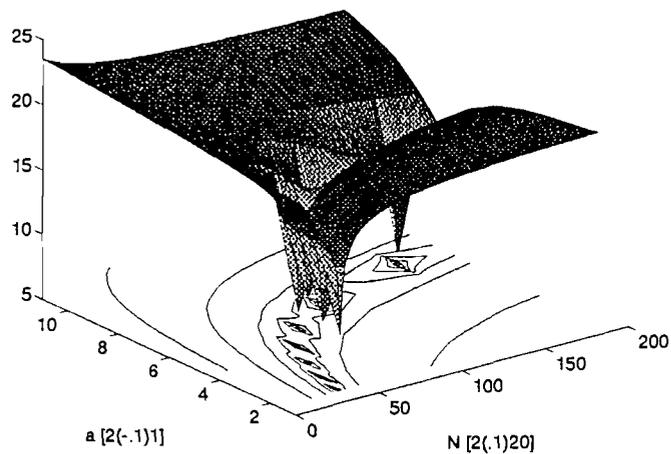


Fig. 2.3 - Função a ser minimizada para a estimação de parâmetros da distribuição $\mathcal{K}An$.

Nas aplicações apresentadas no decorrer deste trabalho não foi feita a estimação simultânea de α e n . Devido à complexidade que esta estimação

simultânea envolve, e dado que se conhecia *a priori* a ordem de n , este parâmetro foi estimado por meio de uma outra técnica, e utilizado diretamente no sistema de equações (2.14).

Distribuição 2.13 Um caso particular da distribuição $X \sim \mathcal{K}An(\alpha, \beta)$ é quando $n = 1$. Para esta situação tem-se que:

$$f_X(x) = \frac{4}{\Gamma(\alpha)} \left(\frac{\alpha}{\beta}\right)^{\frac{\alpha+1}{2}} x^\alpha K_{\alpha-1}\left(2x\sqrt{\frac{\alpha}{\beta}}\right) \mathbb{I}_{\mathbb{R}_+}(x), \quad (2.15)$$

$$F_X(x) = \left[1 - \frac{2x^\alpha}{\Gamma(\alpha)} \left(\frac{\alpha}{\beta}\right)^{\frac{\alpha}{2}} K_\alpha\left(2x\sqrt{\frac{\alpha}{\beta}}\right)\right] \mathbb{I}_{\mathbb{R}_+}(x). \quad (2.16)$$

É possível ver que vale a seguinte relação:

$$\mathbb{E}(X^m) = \left(\frac{\beta}{\alpha}\right)^{\frac{m}{2}} \frac{\Gamma(1+m/2)\Gamma(\alpha+m/2)}{\Gamma(\alpha)},$$

e daí é imediato que:

$$\mathbb{E}(X) = \sqrt{\frac{\beta\pi}{\alpha}} \frac{\Gamma(\alpha + \frac{1}{2})}{2\Gamma(\alpha)}, \text{ e que } \text{Var}(X) = \beta \left[\beta - \frac{\pi}{4\alpha} \left(\frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)} \right)^2 \right].$$

Os estimadores de α e β , pelo método dos momentos são dados por $\hat{\beta} = \widehat{m}_2$ e pela solução de:

$$\sqrt{\frac{\widehat{m}_2\pi}{\alpha}} \frac{\Gamma(\alpha + \frac{1}{2})}{2\Gamma(\alpha)} - \widehat{m}_1 = 0.$$

Distribuição 2.14 A variável aleatória X é dita ter distribuição Gamma com parâmetros $\alpha > 0$ e $\beta > 0$ (em símbolos $X \sim \Gamma(\alpha, \beta)$) se tem densidade dada por:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \mathbb{I}_{\mathbb{R}_+}(x). \quad (2.17)$$

A sua função de distribuição acumulada não possui, em geral, forma fechada pois é dada por:

$$F_X(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} v^{\alpha-1} e^{-v} dv.$$

É possível ver que os momentos de ordem m são dados por:

$$\mathbb{E}(X^m) = \frac{\Gamma(\alpha + m)}{\beta^m \Gamma(\alpha)},$$

e daí é imediato concluir que $\mathbb{E}(X) = \alpha/\beta$ e $\text{Var}(X) = \alpha/\beta^2$. Na terminologia usual dentro da comunidade SAR, se $X \sim \Gamma(\alpha, \beta)$ é usada para modelar dados provindos de regiões homogêneas com detecção quadrática (imagem de intensidade) e N visadas (pois, sendo uma soma de Exponenciais, tem distribuição Gamma), tem-se que $X \sim \Gamma(N, N/\mu_I)$, onde μ_I é a média da intensidade. Isto é: $X = N^{-1} \sum_{i=1}^N X_i$, com $X_i \sim \mathcal{E}(\mu_I)$ para todo $1 \leq i \leq N$.

A seguir é apresentada a distribuição da raiz quadrada de uma variável aleatória com distribuição Γ . O uso desta distribuição para análise de imagens de radar de abertura sintética, justificada mais adiante, não foi encontrado na literatura.

Distribuição 2.15 A variável aleatória X é dita ter distribuição Raiz de Gamma com parâmetros $\alpha > 0$ e $\beta > 0$ (em símbolos $X \sim \sqrt{\Gamma}(\alpha, \beta)$) se tem densidade dada por:

$$f_X(x) = \frac{2\beta^\alpha}{\Gamma(\alpha)} x^{2\alpha-1} \exp(-\beta x^2) \mathbb{1}_{\mathbb{R}_+}(x). \quad (2.18)$$

A sua função de distribuição acumulada não possui, em geral, forma fechada pois é dada por:

$$F_X(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\beta x^2} v^{\alpha-1} e^{-v} dv.$$

É possível ver que os momentos de ordem m são dados por:

$$\mathbb{E}(X^m) = \frac{\Gamma(\alpha + m/2)}{\beta^{m/2} \Gamma(\alpha)},$$

e daí é imediato concluir que:

$$\mathbb{E}(X) = \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)\beta^{1/2}}, \text{ e que } \text{Var}(X) = \frac{1}{\beta} \left[\alpha - \left(\frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)} \right)^2 \right].$$

Na terminologia usual dentro da comunidade SAR, se $X \sim \sqrt{\Gamma}(\alpha, \beta)$ é usada para modelar uma imagem com detecção linear (dados de amplitude) e N visadas, a densidade dada na equação (2.18) costuma escrever-se na forma seguinte:

$$f_A(a) = \frac{2N^N a^{2N-1}}{\Gamma(N)(\mu_I)^N} \exp\left(-\frac{Na^2}{\mu_I}\right) \mathbb{1}_{\mathbb{R}_+}(x), \quad (2.19)$$

onde μ_I é a média da intensidade. Isto é interessante quando pode-se supor que as observações são formadas como $X = \sqrt{\sum_{i=1}^N X_i}$, com $X_i \sim \mathcal{E}(\mu_I)$ para todo $1 \leq i \leq N$. Já, se a formação das observações for dada por $X = N^{-1} \sum_{i=1}^N \sqrt{X_i}$, com $X_i \sim \mathcal{E}(\mu_I)$ para todo $1 \leq i \leq N$ (isto é, uma soma de variáveis aleatórias com distribuição Rayleigh), não é conhecida a densidade (nem a função de distribuição acumulada) de X . Porém, as suas esperança e variância são dadas por

$$\mathbb{E}(X) = \frac{\sqrt{\mu_I \pi}}{2}, \text{ e } \text{Var}(X) = \frac{\mu_I}{N} \left(\frac{4 - \pi}{\pi} \right). \quad (2.20)$$

2.5 O Paradigma de Bayes e Problemas na Análise e no Processamento de Imagens

O problema de segmentação de imagens é um caso muito interessante da teoria geral de classificação. O intuito das técnicas de segmentação é, dada uma imagem observada, atribuir cada um dos seus pixels a uma das classes possíveis pré-determinadas. Exemplos e aplicações disto são inúmeros. Pode-se imaginar o caso de, utilizando uma imagem multiespectral de radar de abertura sintética (SAR) de uma determinada região da Amazônia, ou se ter a necessidade de atribuir cada elemento de imagem a uma das classes “*FLORESTA, SERRADO, CULTURA, QUEIMADA, DESMATADO, ÁGUA, NENHUMA DAS ANTERIORES*”. Isto, que em princípio pode ser feito por um especialista no assunto, ganharia em, pelo menos, tempo se fosse incorporada uma técnica automática e eficiente implementada em computador (que não exclui o especialista).

Neste trabalho será utilizado o termo *segmentação* prioritariamente para fazer referência ao problema de, dada uma imagem, atribuir cada pixel a uma entre K classes possíveis e sem que estas classes façam referência a objetos da natureza. O termo *classificação* será utilizado, toda vez que possível, para designar o problema de, dada uma imagem segmentada, atribuir a cada classe um rótulo correspondente a objetos da natureza. O termo *restauração* será reservado ao problema de, dada uma imagem, processá-la e obter uma nova imagem com menos defeitos de ruído, borramento, etc. Para fixar idéias e introduzir a notação que será

empregada no decorrer deste trabalho, os três problemas anteriores seriam, dada a imagem observada $\mathbf{y} = [y_s]_{s \in S}$, com S um suporte adequado e $y_s \in \{0, \dots, 255\}$ para todo $s \in S$ (uma imagem monoespectral de 256 níveis de cinza), obter:

$$\hat{\mathbf{x}} = [x_s]_{s \in S} \text{ com } x_s \in \{0, \dots, 255\} \text{ seria restauração,} \quad (2.21)$$

$$\hat{\mathbf{x}} = [x_s]_{s \in S} \text{ com } x_s \in \{\xi_1, \dots, \xi_K\} \text{ seria segmentação, e} \quad (2.22)$$

$$\hat{\mathbf{x}} = [x_s]_{s \in S} \text{ com } x_s \in \{\text{trigo, água}\} \text{ seria classificação.}$$

Contudo, é importante frisar que os modelos que serão empregados permitem uma grande flexibilidade notacional, e que os três conjuntos aos quais pode pertencer x_s são matematicamente iguais. Por isto, em particular para o problema descrito pela equação (2.22), aparecerão, entre outras, as seguintes versões do mesmo: $x_s \in \{\xi_1, \dots, \xi_K\}$, $x_s \in \{-1, +1\}$, $x_s \in \{0, \dots, K-1\}$, $x_s \in \{1, \dots, K\}$, etc.

Este trabalho se concentrará nos problemas de segmentação e restauração de imagens de radar de abertura sintética.

Existem dois grandes grupos de técnicas de segmentação: as determinísticas e as estocásticas —incluem-se aqui as técnicas estatísticas clássicas—. As primeiras baseiam-se em noções geométricas (basicamente variações de distância em algum sentido) e aplicam métodos mais ou menos *ad hoc* (grafos e suas componentes conexas, paralelepípedos, o perceptron (40), etc.) para resolver os problemas. Dois exemplos interessantes destas técnicas aplicadas ao processamento de imagens podem ser vistos nos trabalhos de Leguizamón (93,94). As segundas supõem que os dados são amostras de variáveis aleatórias e utilizam o arsenal da inferência.

Como toda vez que é utilizada uma abordagem estocástica, o elemento central vem dado pelo modelo utilizado. Chegar a formular um modelo que, simultaneamente, seja tratável e reflita a realidade é altamente não trivial. Nesta fase devem entrar considerações físicas sobre o problema (em geral, uma modelagem do sistema imageador que, *per se*, envolve outras tantas dificuldades) e delicados aspectos computacionais: um modelo excessivamente detalhado é intratável, e simplificações em demasia levam a modelos computacionalmente pouco eficientes.

Uma exceção a esta regra de modelagem pode ser vista no trabalho (67), onde é proposta uma série de características texturais —com forte sabor de estatísticas de diversas ordens— a serem estimadas, para proceder à segmentação de uma imagem e sem que haja um modelo explícito.

Neste Capítulo serão apresentados alguns modelos conhecidos e consagrados, para introduzir outros, em grau crescente de complexidade.

Será visto como, para os problemas e modelos apresentados, o modelo de Geman e Geman de formação da imagem é sempre adequado, permitindo assim um tratamento unificado —pelo menos do ponto de vista teórico— de uma larga classe de problemas: segmentar sempre será, neste trabalho, um problema de estimação.

Utilizar-se-á o espaço probabilístico clássico mais a família de distribuições possíveis para o vetor de observações, ou seja, a tripla “espaço amostral-classe dos eventos aleatórios-probabilidade definida para os eventos aleatórios”, denotada $(\Omega, \mathcal{A}, \mathbb{P}r)$. A classe de distribuições sob consideração para cada problema será denotada $\mathcal{P} = \{\mathbb{P}r_{\theta} : \theta \in \Theta\}$, onde $\Theta \subset \mathbb{R}^{\ell}$ é o espaço paramétrico.

Considerar-se-ão apenas os modelos paramétricos regulares, onde um número finito de parâmetros determina a distribuição de $\mathbf{Y} = (Y_1, \dots, Y_n)$, o vetor de observações.

2.6 Problemas de Processamento e Análise de Imagens

Será visto como é possível utilizar o paradigma de Bayes para formular problemas típicos do processamento de imagens. Algumas das dificuldades destes problemas são o enorme volume de dados a serem manuseados e, talvez a maior, a complexidade de fornecer distribuições *a priori* que modelem realisticamente o conhecimento prévio do assunto.

A segmentação de uma imagem é um procedimento pelo qual a partir de uma imagem observada y obtém-se uma nova imagem \hat{x} , onde a cada

posição é atribuído um único rótulo, entre K rótulos possíveis. Para fixar idéias, pode-se pensar nesses K rótulos como diferentes cores. O intuito deste procedimento é encontrar uma partição da imagem original \mathbf{y} em K regiões com propriedades estatísticas homogêneas, a partir de hipóteses a respeito da distribuição da imagem verdadeira \mathbf{x} e do processo que a transforma na imagem observada \mathbf{y} . Nesse sentido, a imagem segmentada $\hat{\mathbf{x}}$ é um estimador da imagem verdadeira \mathbf{x} .

Isto é, deseja-se rotular com uma mesma cor as posições que provêm de uma mesma distribuição. É também importante notar que os rótulos já mencionados não fazem referência a objetos da natureza do tipo *água*, *trigo*, etc.; eles apenas separam classes com propriedades diferentes entre si. A respeito do uso de técnicas estatísticas na atribuição de rótulos simbólicos, no trabalho (108) pode ver-se a aplicação de campos Markovianos a tal efeito.

Utilizar-se-á o modelo resultante do teorema de Geman e Geman, e a sua extensão (22,59), para a formação da imagem. Supor-se-á que a imagem observada \mathbf{y} resulta das operações locais \mathbf{H} (os borramentos), pontuais ϕ (as distorções) e das misturas \odot com matrizes de ruído \mathbf{E} , sendo que todas as operações e objetos mencionados estão bem definidos sobre um conjunto suporte S :

$$\tau(\mathbf{x}) = \mathbf{y} = \phi(\mathbf{H}(\mathbf{x})) \odot \mathbf{e}.$$

Os problemas de segmentação e de restauração consistem em inverter a operação τ e recuperar \mathbf{x} baseado na observação de \mathbf{y} . A diferença básica entre as duas classes de problemas citados é que a cardinalidade do conjunto de soluções, isto é $\#\Xi$, nos primeiros é muito menor que nos segundos (comparar as equações (2.21) e (2.22)).

Das definições dos borramentos $[H_s]_{s \in S}$ e das distorções $[\phi_s]_{s \in S}$ para todo $s \in S$, e das especificações de $\mathbb{P}r(\mathbf{X} = \mathbf{x})$ e de $\mathbb{P}r(\mathbf{E} = \mathbf{e})$ dependerá, em geral, a complexidade do problema.

2.6.1 Segmentação em Modelos Básicos (o Método de Máxima Verossimilhança)

O método estatístico mais empregado para fazer esta partição é o método de máxima verossimilhança, que consiste das seguintes etapas:

- 1) Estabelecer o número de populações diferentes que aparecem na imagem y .
- 2) Estabelecer a distribuição de cada uma dessas populações.
- 3) Estimar os parâmetros relevantes, para cada classe.
- 4) Atribuir cada *pixel* à classe cujo valor de verossimilhança é máximo, usando a informação coletada.

Na prática, a etapa 1 é feita pela observação da imagem, ou pelo conhecimento adquirido de outras fontes (fotografias, cartas, estudos de campo, etc., da região a ser segmentada). Na etapa 2 é desejável levar em conta características da parte física do processo de formação da imagem (em geral, imagens no espectro visível com k bandas são modeladas através da distribuição Normal k -variada). A etapa 3 (também chamada *treinamento* ou *aprendizado*) é realizada escolhendo-se pequenas sub-regiões representativas de cada classe, e calculando-se certas estatísticas delas, tais como média amostral, variância amostral, etc. A etapa 4 (segmentação por máxima verossimilhança propriamente dita) consiste em procurar-se, entre as K classes, aquela cuja densidade atinge o valor máximo, em cada valor de *pixel*.

Este esquema tem dado bons resultados. Porém, o uso de métodos mais sofisticados pode ser requerido em algumas situações, tais como:

- A distribuição Normal k -variada, introduzida na etapa 2, não é satisfatória (como no caso de uma classe homogênea apresentar histogramas multimodais ou com forte assimetria) e não há informação adicional sugerindo outras distribuições. Neste caso pode ser necessário lançar-se mão de técnicas não paramétricas para a estimação de densidades.

- A fase de aprendizado não é muito confiável, por exemplo, pela possibilidade de entrarem em uma única amostra observações de diversas populações. Requer-se, neste caso, o uso de, por exemplo, técnicas robustas de estimação de parâmetros.
- Não é possível discernir, *a priori*, o número de classes. Pode-se, neste caso, utilizar técnicas de agregamento de dados, entre outras.

Convém ressaltar, entretanto, que o método pode não funcionar adequadamente em algumas situações, por ser derivado de algumas hipóteses não necessariamente realistas, tais como:

- 1) equiprobabilidade de classes, isto é, dado que há K classes presentes, a probabilidade de que um *pixel* pertença a uma dada classe é igual à probabilidade dele pertencer a qualquer outra, e
- 2) independência total, ou seja, a verdadeira classe de uma posição independe da verdadeira classe das posições vizinhas, e a observação em cada posição independe das observações nas vizinhas, dadas as respectivas verdades terrestres.

Começar-se-á ilustrando algumas aplicações onde a modelagem *a priori* é simples. Por isto entende-se que a distribuição de \mathbf{X} é uma coleção de variáveis aleatórias independentes identicamente distribuídas, isto é $\mathbb{P}\text{r}(\mathbf{X} = \mathbf{x}) = \prod_{s \in S} \mathbb{P}\text{r}(X_s = x_s)$. Considere-se o caso da classificação de padrões k -dimensionais, isto é, a segmentação de uma imagem com k bandas. Este método é conhecido, em geral, como o Método da Máxima Verossimilhança (ou método *max-ver*).

Suponha-se inicialmente que não há nem borramento nem distorção, ou seja $\mathbf{Y} = \mathbf{X} \odot \mathbf{E}$; que o problema é de segmentação, ou seja $\Xi_s \in \{\xi_1, \dots, \xi_K\}$ as $K \geq 2$ classes possíveis valendo $\mathbb{P}\text{r}(X_s = \xi_j) = \pi_j$, com $\pi_j \geq 0$ e $\sum_{j=1}^K \pi_j = 1$; que o ruído é formado por K matrizes (todas indexadas em $s \in S$, e em cada s um vetor k -dimensional) com $\mathbf{E}_s = (\mathbf{E}_s^1, \dots, \mathbf{E}_s^K)$, onde $\mathbf{E}_s^j \sim \mathcal{N}k(\boldsymbol{\mu}^j, \Sigma^j)$, $\boldsymbol{\mu}^j = (\mu_1^j, \dots, \mu_k^j)^T$ ($\mu_i^j \in \mathbb{R}$ para todo $1 \leq j \leq K$ e para todo $1 \leq i \leq k$) e $\Sigma^j = [\sigma_{i\ell}^j]_{1 \leq i, \ell \leq k}$ a matriz de covariância para todo $1 \leq j \leq K$. Suponha-se também

que há independência entre \mathbf{E}_s^j e \mathbf{E}_t^j para todo par de pontos s e t , ambos em S , diferentes e para todo $1 \leq j \leq K$.

A imagem é formada da seguinte maneira: se o pixel X_s adotou o valor ξ_j observa-se a variável aleatória \mathbf{E}_s^j , que tem distribuição $\mathcal{N}_k(\boldsymbol{\mu}^j, \Sigma^j)$, ou seja as operações $\odot = [\odot_s]_{s \in S}$ ficam definidas como $X_s \odot_s \mathbf{E}_s = \mathbf{E}_s^j$ se $X_s(\omega) = \xi_j$, para todo $s \in S$. Colocando o problema em forma construtiva, para cada uma das K classes, o processo de degradação gera (independentemente) uma matriz \mathbf{E}^j ($1 \leq j \leq K$) formada por observações de variáveis aleatórias independentes identicamente distribuídas Normais k -variadas com média $\boldsymbol{\mu}^j$ e matriz de covariância Σ^j . Dada a ocorrência do processo $\mathbf{X}(\omega)$, que provém da distribuição *a priori* produto, decorrente das variáveis aleatórias X_s independentes, antes mencionadas, substitui-se a observação $X_s(\omega) = \xi_j$ pelo vetor $\mathbf{E}_s^j(\omega)$. Note-se que cada matriz $\mathbf{E}^j(\omega)$ corresponde ao que se observaria se a imagem $\mathbf{X}(\omega)$ fosse completamente homogênea e da classe ξ_j .

Uma forma razoável e simples de (tentar) resolver este problema é usar como $\delta(\mathbf{y})$ o estimador MAP, isto é:

$$\delta(\mathbf{y}) = \text{máx}_{\mathbf{x} \in \Xi}^{-1} \{ \text{Pr}(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y}) \};$$

usando a independência entre pixels resulta em uma segmentação ponto a ponto sem supervisão e, para fazê-la, precisa-se conhecer os valores de $\boldsymbol{\mu}^j$ e de Σ^j para todo $1 \leq j \leq K$.

Caso estes parâmetros sejam desconhecidos o problema de segmentação ainda é ponto a ponto, mas necessita-se de uma fase prévia de supervisão ou *aprendizado*. Suponha-se conhecer uma partição de $S = A \uplus B$ onde para todo $s \in B$ conhece-se $X_s(\omega)$ (e tal que existe um número suficiente de pixels correspondentes a cada classe ξ_j para fazer a estimação). Também suponha-se que $\boldsymbol{\mu}$ e Σ são amostras de vetores aleatórios $\mathbf{M}: \Omega \rightarrow \mathbb{R}^k$ e $\mathbf{V}: \Omega \rightarrow \mathcal{M}$ (\mathcal{M} é o conjunto de todas as matrizes semi-definidas positivas de ordem $k \times k$) de distribuições conhecidas $f_{\mathbf{M}}$ e $f_{\mathbf{V}}$, respectivamente.

Usando as amostras $\mathbf{y}_B = (\mathbf{y}_B^1, \dots, \mathbf{y}_B^K)$ a segmentação dos pixels \mathbf{y}_A é feita em duas etapas: no aprendizado usa-se a fórmula de Bayes:

$$\begin{aligned} f_{\mathbf{M}|\mathbf{Y}_B = \mathbf{y}_B^j}(\cdot | \mathbf{Y}_B^j = \mathbf{y}_B^j) &\propto f_{\mathbf{Y}_B^j|\mathbf{M} = \mu^j}(\mathbf{y}_B^j | \mathbf{M} = \mu^j) f_{\mathbf{M}}(\cdot) \\ f_{\mathbf{V}|\mathbf{Y}_B = \mathbf{y}_B^j}(\cdot | \mathbf{Y}_B^j = \mathbf{y}_B^j) &\propto f_{\mathbf{Y}_B^j|\mathbf{V} = \Sigma^j}(\mathbf{y}_B^j | \mathbf{V} = \Sigma^j) f_{\mathbf{V}}(\cdot) \end{aligned}$$

para obter os estimadores de μ^j e de Σ^j para todo $1 \leq j \leq K$. Na segmentação, substitui-se estes valores estimados na equação MAP.

Caso se possa supor apenas que μ^j e Σ^j são (desconhecidos e) fixos para todo $1 \leq j \leq K$, a etapa de aprendizado reduz-se a algum método de estimação (por exemplo, usar os EMV).

Nos exemplos apresentados a seguir, a solução ao problema de atribuição do padrão \mathbf{y}_s a uma entre K classes possíveis será apresentada de uma maneira unificada: serão procuradas K funções $g^\ell: \mathbb{R}^k \rightarrow \mathbb{R}$, $1 \leq \ell \leq K$, tais que se a função $g^j(\mathbf{y}_s) > g^\ell(\mathbf{y}_s)$ para todo $j \neq \ell$, então o padrão \mathbf{y}_s será atribuído à classe j . Ver-se-á como uma larga variedade de problemas de classificação e de reconhecimento de padrões pode ser colocada desta maneira unificada.

Serão vistos três casos supondo conhecidos, ou já estimados, os parâmetros μ^j e Σ^j para todo $1 \leq j \leq K$.

Exemplo 2.5 As observações são variáveis aleatórias independentes identicamente distribuídas $\mathcal{N}(\mu_i^j, \sigma^2)$ para todo $1 \leq j \leq K$ e para todo $1 \leq i \leq k$, ou seja, $\Sigma^j = \sigma^2 \mathbf{I}$ (onde denota-se \mathbf{I} a matriz identidade de ordem $k \times k$). Neste caso a regra de segmentação vem dada por $\delta(\mathbf{y}_s) = \xi_\ell$ onde $g^\ell(\mathbf{y}_s) > g^j(\mathbf{y}_s)$ para todo $\ell \neq j$ e, em particular,

$$g^j(\mathbf{y}_s) = \ln \pi_j - \frac{\|\mathbf{y}_s - \mu^j\|^2}{2\sigma^2}.$$

Exemplo 2.6 As observações são variáveis aleatórias com distribuição Normal k -variada, e todas têm a mesma matriz de covariância; isto é, $\Sigma^j = \Sigma$ para todo $1 \leq j \leq K$. A regra de segmentação é a mesma que para o caso anterior mudando apenas

$$g^j(\mathbf{y}_s) = \ln \pi_j - \frac{1}{2}(\mathbf{y}_s - \mu^j)^T \Sigma^{-1}(\mathbf{y}_s - \mu^j).$$

Exemplo 2.7 As observações são variáveis aleatórias com distribuição Normal k -variada e diferentes vetores de média e matrizes de covariância Σ^j . A regra de segmentação é a dada no Exemplo 2.5 com

$$g^j(\mathbf{y}_s) = \ln \pi_j - \frac{1}{2} \left[(\mathbf{y}_s - \boldsymbol{\mu}^j)^\top (\Sigma^j)^{-1} (\mathbf{y}_s - \boldsymbol{\mu}^j) + \ln |\Sigma^j| \right], \quad (2.23)$$

onde $|\Sigma^j|$ denota o determinante da matriz Σ^j , para todo $1 \leq j \leq K$.

É possível ver (por exemplo em (40)) que os Exemplos 2.5 e 2.6 se reduzem à estimação de funções discriminantes lineares; isto é, podem se achar hiperplanos que particionam \mathbb{R}^k nas K regiões correspondentes a ξ_1, \dots, ξ_K . No Exemplo 2.7 precisa-se de funções discriminantes hiperquádricas para obter tal partição.

Na prática *sempre* se usa a regra dada pela equação (2.23) por um motivo bem simples: os parâmetros são desconhecidos e devem ser estimados. Como as distribuições envolvidas admitem densidade, a probabilidade de se obterem as mesmas observações, mesmo se houver igualdade de parâmetros, é nula.

Exemplo 2.8 Neste Exemplo, apresenta-se a metodologia vista nos Exemplos 2.5, 2.6 e 2.7 para um caso em que a hipótese Gaussiana não é admissível: far-se-á uma modelagem —muito simplificada— de uma imagem de radar de abertura sintética (SAR) como uma coleção de variáveis aleatórias independentes com distribuição Rayleigh, e obter-se-á o segmentador MAP para este caso.

Sejam novamente $\Xi = \{\xi_1, \dots, \xi_K\}$ o conjunto das classes possíveis; $\mathbb{P}\mathbf{r}(\mathbf{X} = \mathbf{x}) = \prod_{s \in S} \mathbb{P}\mathbf{r}(X_s = \xi_j) = \pi_j$ a distribuição *a priori* (de variáveis independentes e identicamente distribuídas), $\mathbf{E}_s = (\mathbf{E}_s^1, \dots, \mathbf{E}_s^K)$ e \odot_s como definidos anteriormente, agora com $\mathbf{E}_s^j \sim (\mathcal{R}(\theta_1^j), \dots, \mathcal{R}(\theta_k^j))(s)$, variáveis aleatórias independentes Rayleigh de parâmetro θ_i^j para todo $s \in S$. Suponham-se conhecidos (ou já estimados como visto nas equações (2.7) e (2.8)) os parâmetros $[\theta_i^j]_{1 \leq j \leq K, 1 \leq i \leq k}$, o segmentador MAP pontual virá dado pela regra apresentada no Exemplo 2.5 com

$$g^j(\mathbf{y}_s) = \ln \pi_j - \left[2 \sum_{1 \leq i \leq k} \ln \theta_i^j + \frac{1}{2} \sum_{1 \leq i \leq k} \frac{y_s^2(i)}{(\theta_i^j)^2} \right],$$

onde se escreve $\mathbf{y}_s = (y_s(1), \dots, y_s(k))^\top$.

Exemplo 2.9 Procurar-se-á modelar (um pouco) mais realisticamente as imagens de radar de abertura sintética com uma visada e detecção linear, embora as funções discriminantes $g^j(\mathbf{y}_s)$ não sejam explicitamente calculadas, devido à sua complexidade; apenas indicar-se-á uma forma de incluir na modelagem básica o fato das observações estarem correlacionadas entre as três bandas consideradas.

Sejam Ξ_s e $\text{Pr}(\mathbf{X} = \mathbf{x})$ como definidas no Exemplo anterior; considere-se a matriz de ruído definida como $\mathbf{E}_s = (\mathbf{E}_s^1, \dots, \mathbf{E}_s^K)$, com

$$\mathbf{E}_s^j = (S_{1\Re}^j, S_{1\Im}^j, S_{2\Re}^j, S_{2\Im}^j, S_{3\Re}^j, S_{3\Im}^j)(s) \sim \mathcal{N}6(\mathbf{0}, \Sigma^j)$$

e $\mathbf{Y}_s = \mathbf{X}_s \odot_s \mathbf{E}_s$ dada por $\mathbf{Y}_s = (Z_1^j, Z_2^j, Z_3^j)(s)$ se $X_s = \xi_j$, dada por

$$Z_i^j = \sqrt{(S_{i\Re}^j)^2 + (S_{i\Im}^j)^2} \quad \text{para todo } 1 \leq j \leq K \text{ e } 1 \leq i \leq 3.$$

Desta forma, e supondo válida a hipótese de que componentes reais e imaginárias são independentes, isto é:

$$(S_{i\Re}^j, S_{i\Im}^j) \sim \mathcal{N}2(\mathbf{0}, (\Sigma'_i)^j) \text{ onde } (\Sigma'_i)^j = \begin{pmatrix} \sigma_{ii}^j & 0 \\ 0 & \sigma_{ii}^j \end{pmatrix},$$

ou seja que é possível decompor a densidade conjunta das seis variáveis aleatórias em um produto de densidades:

$$f_{S_{1\Re}^j, S_{1\Im}^j, S_{2\Re}^j, S_{2\Im}^j, S_{3\Re}^j, S_{3\Im}^j} = f_{S_{1\Re}^j, S_{2\Re}^j, S_{3\Re}^j} \times f_{S_{1\Im}^j, S_{2\Im}^j, S_{3\Im}^j},$$

vale que $Z_i^j \sim \mathcal{R}(\sqrt{2\sigma_i^j})$.

Para calcular-se $g^j(\mathbf{y}_s)$, $\mathbf{y}_s = (y_s(1), y_s(2), y_s(3))$, precisa-se conhecer a densidade condicional $f_{\mathbf{Y}_s | X_s = \xi_j}$. Essa densidade, para Σ^j não diagonais, não é de obtenção imediata (o caso diagonal foi visto no exemplo anterior onde $\theta_i^j = \sqrt{2\sigma_i^j}$). O uso da notação empregada, com componentes reais e imaginárias, ficará plenamente justificado no modelo apresentado na Seção 3.5.2.

Antes de entrar na modelagem espacial do processo de classes, um exemplo muito simples ilustrará como é possível usar o modelo apresentado para formular um problema de reconstrução de imagens.

Exemplo 2.10 Suponha-se agora que $\mathbf{Y} = \mathbf{X} + \mathbf{E}$ com $\Xi = \mathbb{R}$, $[X_s]_{s \in S}$ variáveis aleatórias independentes identicamente distribuídas $\mathcal{N}(\theta, \sigma^2)$; $\mathbf{E} = [E_s]_{s \in S}$ variáveis aleatórias independentes identicamente distribuídas $\mathcal{N}(0, \eta^2)$. Problema: estimar $\mathbf{X}(\omega)$ baseados na observação de $\mathbf{Y}(\omega)$. Solução: usar o fato que (ver o Exemplo 2.4)

$$(X_s \mid Y_s(\omega) = y_s) \sim \mathcal{N}\left(\frac{\sigma^2 y_s + \eta^2 \theta}{\sigma^2 + \eta^2}, \cdot\right).$$

2.6.2 Modelos com Dependência Espacial

Nos Exemplos 2.6, 2.7 e 2.9 foi introduzida na modelagem o fato, intuitivamente razoável, de se ter correlação entre as diferentes bandas do mesmo pixel. Outra hipótese interessante é a da *suavidade espacial*: embora não se tenha evidência de $\pi_j \neq \pi_k$ para algum $j \neq k$, dado que $X_s(\omega) = \xi_j$ espera-se que $\mathbb{P}\text{r}(X_t = \xi_j) > 1/K$ para algum t próximo a s .

Como primeira aproximação ao que se deseja, será visto um modelo estocástico muito simples para modelar suavidade espacial.

Exemplo 2.11 O modelo de mosaico, ou de replicação de *pixels*, para a imagem SAR do Exemplo 2.8. Seja $S = \{0, 1, \dots, 2^n - 1\} \times \{0, 1, \dots, 2^n - 1\}$ e a partição $S = S_e \uplus S_d$, onde \uplus denota a união de conjuntos com intersecção nula, (“e” por estocástico e “d” por determinístico) onde $S_e = \{(s_1, s_2) \in S : s_1 \bmod 2 = s_2 \bmod 2 = 0\}$ e define-se $\Xi_s = \{\xi_1, \dots, \xi_K\}$ para todo $s \in S_e$ e $\Xi_s = \xi_0$ caso contrário, valendo $\mathbb{P}\text{r}(X_s = \xi_j) = \pi_j$ e X_s independente de X_t para todo $s, t \in S_e$ com $s \neq t$, como anteriormente se $s \in S_e$ e $X_s = \xi_0$ se $s \in S_d$. Define-se ainda os borramentos $\mathbf{H} = [H_s]_{s \in S}$ da seguinte forma:

$$H_{s_1, s_2}(X_{s_1, s_2}) = \begin{cases} X_{s_1, s_2} & \text{se } (s_1, s_2) \in S_e \\ X_{s_1-1, s_2} & \text{se } s_1 \bmod 2 = 1 \text{ e } s_2 \bmod 2 = 0 \\ X_{s_1, s_2-1} & \text{se } s_1 \bmod 2 = 0 \text{ e } s_2 \bmod 2 = 1 \\ X_{s_1-1, s_2-1} & \text{se } s_1 \bmod 2 = s_2 \bmod 2 = 1. \end{cases}$$

Tem-se, assim, um modelo que incorpora homogeneidade espacial fixando em, pelo menos, quatro o número de posições próximas que pertencem à mesma classe ξ_j : um

mosaico formado por “azulejos” de cores ξ_j e de tamanho 2×2 pixels. A Figura 2.4 é um desenho de S para $n = 3$ onde com “•” representam-se os elementos de S_e , e com “o” os de S_d :

	0	1	2	3	4	5	6	7
0	•	o	•	o	•	o	•	o
1	o	o	o	o	o	o	o	o
2	•	o	•	o	•	o	•	o
3	o	o	o	o	o	o	o	o
4	•	o	•	o	•	o	•	o
5	o	o	o	o	o	o	o	o
6	•	o	•	o	•	o	•	o
7	o	o	o	o	o	o	o	o

Fig. 2.4 - O suporte S particionado em $S_e \uplus S_d$.

Continuando com o Exemplo 2.9, suponha-se agora que a imagem é um mosaico formado por azulejos de ruído homogêneo: $\mathbf{Y} = \mathbf{H}(\mathbf{X}) \odot \mathbf{E} = \mathbf{X}' \odot \mathbf{E}$ e que se quer estimar o mosaico $\mathbf{X}'(\omega)$ baseados na observação de $\mathbf{Y}(\omega) = \mathbf{y}$. Proceda-se então a estimar $\mathbf{X}(\omega)$ trabalhando com os $2^{2(n-1)}$ grupos de observações $\tilde{\mathbf{y}}_{s_1, s_2}$ para todo $(s_1, s_2) \in S_e$ que, pela modelagem apresentada, são variáveis aleatórias independentes identicamente distribuídas $\mathcal{R}(\theta^j)$, onde

$$\tilde{\mathbf{y}}_{s_1, s_2} = \left(y_{s_1, s_2}, y_{s_1+1, s_2}, y_{s_1, s_2+1}, y_{s_1+1, s_2+1} \right),$$

e a nova função discriminante virá dada, para todo $(s_1, s_2) \in S_e$, por:

$$\tilde{g}^j(\tilde{\mathbf{y}}_{s_1, s_2}) = g^j(y_{s_1, s_2}) + g^j(y_{s_1+1, s_2}) + g^j(y_{s_1, s_2+1}) + g^j(y_{s_1+1, s_2+1}).$$

Obtida assim $\hat{\mathbf{x}}$, a estimativa de $\mathbf{X}(\omega)$, estima-se $\mathbf{X}'(\omega)$ simplesmente com $\hat{\mathbf{x}}' = \mathbf{H}(\hat{\mathbf{x}})$. Note-se que a fase de aprendizado, se necessária, deve também respeitar a modelagem de replicação. Outras definições de S_e , S_d e \mathbf{H} permitiriam trabalhar com diferentes modelos de mosaicos. Ver, por exemplo (60), onde são utilizados mosaicos de 21×21 pixels.

Uma linha de pesquisa muito ativa é a procura de distribuições capazes de fornecer bons modelos *a priori* que incluam a noção de contexto espacial; isto é, *pixels* ou grupos de *pixels* recebem a atribuição de rótulos simbólicos como no exemplo da Amazônia antes citado. Nestes problemas a informação contextual é basicamente local, e os campos Markovianos fornecem um mecanismo flexível para a modelagem de dependência espacial. Estas distribuições também são úteis para a modelagem de atributos não observados, como é o caso das bordas entre regiões homogêneas de tons ou de texturas, por exemplo.

Um modelo estocástico que apresenta propriedades muito interessantes é o modelo de Potts-Strauss. Ele é um caso particular de campo Markoviano. Para uma discussão detalhada sobre os campos Markovianos e as suas aplicações algumas referências completas são (25,29,61,109).

Para o problema de imagens pode-se supor que o suporte é um subconjunto finito $S \subset \mathbf{Z}^2$ e, toda vez que necessário, associa-se a cada $s \in S$ um par de inteiros (s_1, s_2) , sendo que s_1 representa a fila e s_2 a coluna. Nas aplicações, é comum se supor que S é um retângulo de \mathbf{Z}^2 , isto é, $S = S_1 \times S_2$ com $S_i = [a_i, \dots, b_i]$ e $a_i < b_i$ para $i = 1, 2$. O processo de classes ou regiões, com contradomínio no conjunto $\Xi_s \in \{\xi_1, \dots, \xi_K\}$, descreve o agrupamento de pixels em *manchas*, e supõe-se que K , o número total de classes, não é muito grande (2, 3 ou 4, tipicamente).

Assim, em cada *pixel* s a variável aleatória X_s pode assumir um dos K valores $\{\xi_1, \dots, \xi_K\}$ e espera-se algum tipo de continuidade espacial, tal que as chances relativas de um pixel ser da mesma classe que os seus vizinhos é alta. Para modelar isto se dirá que $\mathbf{X} = [X_s]_{s \in S}$ é um campo Markoviano em relação ao sistema de vizinhança $\{\partial_s: s \in S\}$ onde ∂_s , a vizinhança do pixel s , é um subconjunto de S que satisfaz i) $s \notin \partial_s$ e ii) $t \in \partial_s \iff s \in \partial_t$, e por definição, valem as condições de positividade ($\text{IPr}(\mathbf{X} = \mathbf{x}) > 0$ para todo $\mathbf{x} \in \Xi$) e de independência condicional:

$$\text{IPr}(X_s = x_s \mid \mathbf{X}_{S \setminus \{s\}} = \mathbf{x}_{S \setminus \{s\}}) = \text{IPr}(X_s = x_s \mid \mathbf{X}_{\partial_s} = \mathbf{x}_{\partial_s}) \quad \forall s \in S.$$

Equivalentemente, pode-se especificar a distribuição conjunta dizendo que \mathbf{X} é uma medida de Gibbs, ou seja:

$$\mathbb{P}_r(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} V_c(\mathbf{x})\right) \quad \text{para todo } \mathbf{x} \in \Xi,$$

onde $\mathcal{C} = \{A \subset S: a \in A, b \in A \Rightarrow a \in \partial_b\}$ é, por definição, o conjunto dos *cliques* do grafo induzido pelo par $(S, \{\partial_s: s \in S\})$; $\mathbf{y}_c = \mathbf{x}_c \Rightarrow V_c(\mathbf{y}) = V_c(\mathbf{x})$ são os *potenciais*; e $Z = \sum_{\mathbf{x} \in \Xi} \exp(-\sum_{c \in \mathcal{C}} V_c(\mathbf{x}))$ é a função de partição.

O modelo de Potts-Strauss fica definido pelas especificações $\partial_s = \{t \in S: \|s - t\| = 1\}$ para todo $s \in S$, e pelos potenciais:

$$\begin{aligned} V_{\{s\}}(\mathbf{x}) &= \alpha_i \text{ se } x_s = \xi_i \quad \text{para todo } 1 \leq i \leq K \quad \text{e} \\ V_{\{(s,t)\}}(\mathbf{x}) &= \begin{cases} \beta & \text{se } x_s = x_t \\ -\beta & \text{se } x_s \neq x_t, \end{cases} \end{aligned}$$

ou pelas distribuições:

$$\begin{aligned} \mathbb{P}_r(X_s = \xi_i) &\propto \exp(\alpha_i) \\ \mathbb{P}_r(X_s = x_s \mid \mathbf{X}_{\partial_s} = \mathbf{x}_{\partial_s}) &\propto \exp(\beta \cdot \#\{t \in \partial_s: x_s = x_t\}), \end{aligned} \quad (2.24)$$

e é possível ver que é uma reparametrização do *modelo logístico multinível* apresentado em (36,80). Este assunto será abordado com maiores detalhes no Capítulo 3.

2.6.3 Estimadores MAP, MPM e ICM

Sendo a distribuição *a priori* de \mathbf{X} o modelo de Potts-Strauss com parâmetros conhecidos $\theta = (\alpha_1, \dots, \alpha_K, \beta)$ e conhecida τ , a mecânica de formação da imagem, i. e. $\tau(\mathbf{X}) = \mathbf{Y} = \phi(\mathbf{H}(\mathbf{X})) \odot \mathbf{E}$, pode-se, em princípio, calcular a distribuição *a posteriori* $\mathbb{P}_r(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y})$ e proceder a estimar $\mathbf{X}(\omega)$ usando a observação de $\mathbf{Y}(\omega) = \mathbf{y}$.

Os principais estimadores relacionados às técnicas Markovianas são o ICM (*Iterated Conditional Modes*: modas condicionais iterativas), o MAP (*Maximum A Posteriori*: máximo *a posteriori*), o MPM (*Marginal Posterior Mode*:

moda da marginal *a posteriori*) e o ABC (*Adaptive Bayesian Classification: classificação Bayesiana adaptativa*).

Uma vantagem dos estimadores MPM é que eles penalizam $\tilde{\mathbf{x}}$, a estimação de \mathbf{x} , proporcionalmente ao número de pixels erradamente classificados, enquanto os MAP penalizam por igual qualquer segmentação errada (127). Isto é,

$$\begin{aligned}\hat{\mathbf{x}} &= \delta_{\text{MAP}}(\mathbf{y}) = \text{máx}_{\mathbf{x} \in \Xi} \text{IPr}(\mathbf{x} \mid \mathbf{y}) \text{ minimiza } C(\mathbf{x}, \hat{\mathbf{x}}) = \mathbb{I}_{\{\mathbf{x} \neq \hat{\mathbf{x}}\}}, \text{ enquanto} \\ \tilde{\mathbf{x}} &= \delta_{\text{MPM}}(\mathbf{y}) = \left[\tilde{x}_s \in \Xi_s, s \in S: \text{IPr}(\tilde{x}_s \mid \mathbf{y}) \geq \text{IPr}(x_s \mid \mathbf{y}) \forall x_s \in \Xi_s \right] \\ &\text{ minimiza } C(\mathbf{x}, \tilde{\mathbf{x}}) = \#\{s \in S: x_s \neq \tilde{x}_s\}.\end{aligned}$$

Novamente pode-se usar o estimador MAP. Outra opção seria utilizar o estimador da moda da marginal *a posteriori* (MPM) que escolhe o valor modal de X_s para cada $s \in S$. Uma vantagem do estimador MAP é que depende de $\text{IPr}(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y})$, sendo que o estimador MPM requer o conhecimento de $\text{IPr}(X_s = x_s \mid \mathbf{Y} = \mathbf{y})$. Muitas vezes os estimadores MAP e MPM produzem estimações similares para a mesma distribuição *a posteriori*.

Para obter estes estimadores MPM é possível tomar amostras de $\text{IPr}(\mathbf{X} \mid \mathbf{Y} = \mathbf{y})$ e ficar em cada $s \in S$ com o valor mais freqüentemente observado nesse ponto. Esta técnica pode ser computacionalmente aceitável toda vez que Ξ_s , o espaço amostral de cada variável aleatória X_s , seja de baixa cardinalidade, e desde que se disponha de técnicas eficientes para simular as distribuições envolvidas (ver Capítulo 3).

Uma outra forma de obter o estimador da imagem é a proposta de Besag (9): as ICM (*Iterated Conditional Modes*), Modas Condicionais Iterativas. O algoritmo vem definido por: dada $\hat{\mathbf{x}}(k)$ a estimação da imagem \mathbf{x} na iteração k , formar a nova estimação $\hat{\mathbf{x}}(k+1)$ atualizando o valor do pixel s , $\hat{x}_s(k)$, com o valor modal da distribuição condicional dada pela expressão $\text{IPr}(X_s \mid \mathbf{X}_{S \setminus \{s\}} = \mathbf{x}_{S \setminus \{s\}}, \mathbf{y})$; no próximo passo, atualizar o pixel t , e assim por diante; parar quando houver evidência de convergência.

É possível ver que este algoritmo, desde que implementado em forma sequencial, assegura uma seqüência de estimações $(\hat{\mathbf{x}}(k))_{k \geq 0}$ tal que $\text{IPr}(\mathbf{X} = \hat{\mathbf{x}}(k) \mid \mathbf{y})$ é não decrescente em k ; isto é $\text{IPr}(\hat{\mathbf{x}}(0) \mid \mathbf{y}) \leq \text{IPr}(\hat{\mathbf{x}}(1) \mid \mathbf{y}) \leq \text{IPr}(\hat{\mathbf{x}}(2) \mid \mathbf{y}) \leq \dots$. Este assunto será abordado com maiores detalhes no Capítulo 4.

2.6.4 Outros Estimadores

É possível pensar em variações baseadas na técnica de obtenção do estimador MPM (103). Como já foi apresentado, este estimador pode ser aproximado pela seguinte técnica: seja $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ uma amostra de tamanho n de observações independentes e identicamente distribuídas da distribuição *a posteriori* $\mathbf{X} \mid \mathbf{Y} = \mathbf{y}$. Assim, em cada $s \in S$ observam-se n valores $(x_s(1), \dots, x_s(n))$. O estimador MPM aproximado é dado, para cada $s \in S$, por \tilde{x}_s o valor mais freqüentemente observado na amostra $(x_s(1), \dots, x_s(n))$.

Admitindo-se que $\Xi_s = \mathbb{R}$ ou que $\Xi_s = \mathbb{R}_+$, seria possível definir os estimadores $\mathbf{x}^{(1)}$ e $\mathbf{x}^{(2)}$, baseados na amostra $(x_s(1), \dots, x_s(n))$, e dados por $x_s^{(1)} = n^{-1} \sum_{i=1}^n x_s(i)$ e $x_s^{(2)} = Q_2(x_s(1), \dots, x_s(n))$, para cada $s \in S$, respectivamente, onde $Q_2(\mathbf{x})$ é a mediana amostral do vetor \mathbf{x} (ver Capítulo 5).

Esta hipótese pode ser uma boa aproximação da realidade quando a cardinalidade $\#\Xi_s$ é *grande* e quando $\Xi_s \subset \mathbb{R}$ para todo $s \in S$. Estes estimadores, por analogia com o caso unidimensional, talvez minimizem os custos dados por $C(\mathbf{x}, \mathbf{x}^{(1)}) = \sum_{s \in S} |x_s - x_s^{(1)}|^2$ e $C(\mathbf{x}, \mathbf{x}^{(2)}) = \sum_{s \in S} |x_s - x_s^{(2)}|$.

CAPÍTULO 3

O MODELO DE POTTS-STRAUSS, A ESTIMAÇÃO POR PSEUDOVEROSSIMILHANÇA E A SIMULAÇÃO DE IMAGENS SAR

3.1 Introdução

O modelo de Potts-Strauss vem sendo largamente empregado em várias áreas. Um caso particular dele é o famoso modelo de Ising, que tem recebido muita atenção desde a sua proposta nos anos '20 como uma forma para modelar o ferromagnetismo.

Este modelo visa descrever, em forma estatística, dependência espacial entre variáveis aleatórias. Por ser um modelo paramétrico, tal dependência é controlada por um número finito de parâmetros.

Os elementos básicos do modelo de Potts-Strauss, que é uma função definida sobre um domínio com valores em um contradomínio e propriedades probabilísticas, são:

- 1) o seu suporte ou domínio (tamanho e estrutura) sobre o qual é definido (por exemplo, as suas propriedades são muito diferentes para suportes finitos e infinitos; sabe-se muito para o caso quando o suporte é \mathbf{Z}^2 , e muito pouco para outras dimensões);
- 2) o seu contradomínio (no caso do modelo de Ising com suporte S , o contradomínio é o conjunto $\{-1, +1\}^S$);
- 3) os parâmetros que descrevem a dependência.

A generalidade do modelo de Potts-Strauss é enorme, e a escolha de um modelo em particular depende fortemente da aplicação para a qual deseja-se

destiná-lo. O interesse deste trabalho reside na simulação, processamento e análise de imagens; por isto, empregar-se-á exclusivamente o modelo dado pelo suporte finito S : um subconjunto de \mathbf{Z}^2 , e pelo contradomínio definido por $\Xi_s = \{1, \dots, K\}$ para todo $s \in S$. Desta forma pode estabelecer-se uma correspondência entre cada elemento de S e a coordenada de cada *pixel* de uma imagem; da mesma maneira, os elementos de Ξ_s corresponderão às K classes possíveis de cada *pixel* da imagem.

Este modelo é útil para a formulação de técnicas de processamento de imagens pois é evidente que, na maioria delas, o fato de um *pixel* pertencer a uma determinada classe é *afetado* pelo fato dos seus vizinhos mais próximos pertencerem a certas classes. Isto é, ao observar um aglomerado de *pixels* de floresta, na hora de decidir a respeito da classe à qual pertence um *pixel* interior a esse aglomerado, incorpora-se a informação dos seus vizinhos: existirá a *tentação* de atribuí-lo à classe “floresta”.

Pelo mesmo raciocínio, se o objetivo for simular uma imagem, será desejável que *pixels* espacialmente próximos exibam a tendência de pertencer às mesmas classes.

A noção de *dependência espacial* é útil quando incorporada num esquema Bayesiano de processamento de imagens, o que será feito no decorrer deste trabalho; isto é, o modelo de Potts-Strauss será empregado como distribuição *a priori* para o desenvolvimento de algoritmos de segmentação de imagens.

A importância do uso de *boas* distribuições para o tratamento Bayesiano de imagens tem sido frisado por muitos autores. Por exemplo, em (10) pode ver-se uma formulação muito geral dos problemas de restauração e segmentação de imagens dentro do contexto Bayesiano. Na referência (124) o autor apresenta o uso de modelos espaciais para a deconvolução de imagens astronômicas, segmentação de imagens binárias e extração de formas em objetos contidos em imagens digitais. A literatura referente ao uso de modelos espaciais em processamento de imagens é enorme, e cresce dia a dia.

3.2 Formulação Teórica

O modelo de Potts-Strauss é um caso particular de *campo Markoviano*. Os campos Markovianos são modelos probabilísticos que incorporam a noção de dependência espacial entre variáveis aleatórias. Uma das áreas onde estes modelos têm grande sucesso é a Mecânica Estatística.

Seja S um conjunto finito. Para cada $s \in S$ seja $\Xi_s \subset \mathbb{R}$ um conjunto finito. Define-se o contradomínio (ou as configurações possíveis) do campo como:

$$\Xi = \prod_{s \in S} \Xi_s = \left\{ \mathbf{x}: S \longrightarrow \bigcup_{s \in S} \Xi_s: x_s \in \Xi_s, \forall s \in S. \right\}$$

Para cada $A \subset S$ seja $\Xi_A = \prod_{s \in A} \Xi_s$. Identifica-se Ξ_S com Ξ e Ξ_s com $\Xi_{\{s\}}$. Considera-se que Ξ é um espaço mensurável com a σ -álgebra de todos os seus subconjuntos.

Seja π uma probabilidade nunca nula sobre Ξ , isto é, $\pi(\mathbf{x}) > 0$ para todo $\mathbf{x} \in \Xi$. Para cada $s \in S$ seja π_s a função definida sobre Ξ por

$$\pi_s(\mathbf{x}) = \pi(\mathbf{y} \in \Xi: \{y_s = x_s\} \mid \{\mathbf{y}: y_t = x_t, t \neq s\}). \quad (3.1)$$

A família de funções $\{\pi_s: s \in S\}$ será chamada *família de características locais associadas* a π .

O seguinte resultado, cuja validade é garantida apenas para o caso do suporte S finito, assegura que a distribuição conjunta fica unicamente determinada pelas características locais.

Proposição 3.1 Sejam π e μ duas probabilidades sobre Ξ com $\pi(\mathbf{x}) > 0$ e $\mu(\mathbf{x}) > 0$ para todo $\mathbf{x} \in \Xi$. Se π_s e μ_s são iguais para todo $s \in S$, então π coincide com μ .

Para as provas das Proposições apresentadas nesta Seção, e para outras propriedades relevantes de campos Markovianos, uma referência muito completa é (25). Outras referências interessantes são (9,39,52,58,59,61,71,109,118,127).

Definição 3.1 Uma família $\mathcal{V} = \{V_s: s \in S\}$ de subconjuntos de S é um *sistema de vizinhanças* de S se:

- 1) $s \notin V_s$;
- 2) $s \in V_t$ implica que $t \in V_s$.

Neste caso, dir-se-á que $\mathcal{G} = (S, \mathcal{V})$ é um *grafo*, que os pontos de S são os *nodos* do grafo \mathcal{G} , e que (s, t) é um *arco* do grafo \mathcal{G} se $s \in V_t$ ou se $t \in V_s$.

Definição 3.2 Sejam $\mathcal{G} = (S, \mathcal{V})$ um grafo e π uma probabilidade nunca nula sobre Ξ . Dir-se-á que π é \mathcal{G} -Markoviana se

$$\pi_s(\mathbf{x}) = \pi\left(\mathbf{y} \in \Xi: \{y_s = x_s\} \mid \{\mathbf{y}: y_t = x_t, t \in V_s\}\right).$$

Para cada $s \in S$ seja $V_s^0 = S \setminus \{s\}$. Então $\mathcal{V}^0 = \{V_s^0: s \in S\}$ é um sistema de vizinhanças de S , $\mathcal{G} = (S, \mathcal{V}^0)$ é um grafo, e

$$\pi_s(\mathbf{x}) = \pi\left(\mathbf{y} \in \Xi: \{y_s = x_s\} \mid \{\mathbf{y}: y_t = x_t, t \in V_s^0\}\right),$$

qualquer que seja π , probabilidade nunca nula sobre Ξ . Isto é, π é \mathcal{G}^0 -Markoviana.

Sejam $(\Omega, \mathcal{A}, \mathbb{P}_r)$ um espaço de probabilidade e $\mathbf{X}: \Omega \rightarrow \Xi$ uma função mensurável, que será chamada *campo aleatório*. Supor-se-á que $\mathbb{P}_r(\mathbf{X} = \mathbf{x}) = \mathbb{P}_r(\mathbf{X} = \mathbf{x}) > 0$ para todo $\mathbf{x} \in \Xi$.

Definição 3.3 Seja $\mathcal{G} = (S, \mathcal{V})$ um grafo. Dir-se-á que \mathbf{X} é um *campo aleatório \mathcal{G} -Markoviano* (é um \mathcal{G} -CAM) se $\mathbb{P}_r(\mathbf{X})$ é \mathcal{G} -Markoviana.

Definição 3.4 Seja Λ um conjunto finito, considerado como um espaço mensurável com a σ -álgebra de todos os seus subconjuntos. Sejam (X_0, \dots, X_N) funções mensuráveis definidas sobre Ω com valores em Λ tais que $\mathbb{P}_r(X_n = \lambda) > 0$ para todo $0 \leq n \leq N$ e para todo $\lambda \in \Lambda$. Dir-se-á que as funções (X_0, \dots, X_N) formam uma \mathbb{P}_r -cadeia de Markov com espaço de estados Λ se para cada $0 \leq n \leq N - 1$ e cada $(\lambda_0, \dots, \lambda_{n+1})$ em Λ vale que:

$$\mathbb{P}_r(X_{n+1} = \lambda_{n+1} \mid X_n = \lambda_n, \dots, X_0 = \lambda_0) = \mathbb{P}_r(X_{n+1} = \lambda_{n+1} \mid X_n = \lambda_n).$$

Nesse caso, para cada $1 \leq n \leq N$ define-se a função $\mathbb{P}r_n: \Lambda \times \Lambda \longrightarrow [0, 1]$, chamada *probabilidade de transição de X_{n-1} no estado ξ a X_n no estado λ* por:

$$\mathbb{P}r(\xi | \lambda) = \mathbb{P}r(X_n = \lambda | X_{n-1} = \xi),$$

e chamar-se-á *probabilidade inicial da cadeia* à probabilidade $\mathbb{P}r_0 = \mathbb{P}r_{X_0}$.

Definição 3.5 Seja $\mathcal{P}(S)$ a coleção dos subconjuntos do conjunto S (as partes de S); uma função $V: \Xi \times \mathcal{P}(S) \longrightarrow \mathbb{R}$ é um potencial se satisfaz as propriedades dadas por:

- 1) $V(\mathbf{x}, \emptyset) = 0$ para todo $\mathbf{x} \in \Xi$,
- 2) Se $x_s = y_s$ para todo $s \in A$, então $V(\mathbf{x}, A) = V(\mathbf{y}, A)$.

Suponha-se agora que $0 \in \Xi_s$ para todo $s \in S$.

Definição 3.6 Dir-se-á que um potencial V está *normalizado* se vale que para qualquer $A \subset S$, e para qualquer $\mathbf{x} \in \Xi$ tal que existe $t \in A$ para o qual $x_t = 0$, $V(\mathbf{x}, A) = 0$.

Definição 3.7 Seja $\mathcal{G} = (S, \mathcal{V})$ um grafo. Diz-se que um subconjunto A de S é \mathcal{G} -completo se vale uma das seguintes condições:

- A é um conjunto de um elemento só, isto é, $\#A = 1$.
- Se a e b são elementos distintos de A , então $a \in V_b$.

Os conjuntos \mathcal{G} -completos também recebem o nome de *cliques* do grafo \mathcal{G} . Porém, é importante notar que para alguns autores ((77), por exemplo) para o conjunto A , \mathcal{G} -completo, ser um *clique* ele tem que ser também maximal em relação à inclusão.

Seja $\mathcal{C}(\mathcal{G})$ a família de todos os conjuntos \mathcal{G} -completos.

Definição 3.8 Dado um potencial V , a função $H_V: \Xi \longrightarrow \mathbb{R}$ definida por:

$$H_V(\mathbf{x}) = - \sum_{A \in \mathcal{C}S} V(\mathbf{x}, A)$$

é chamada *energia* da configuração \mathbf{x} sob o potencial V .

Definição 3.9 Sejam $\mathcal{G} = (S, \mathcal{V})$ um grafo e V um potencial. Dir-se-á que V é um \mathcal{G} -potencial se $V(\mathbf{x}, A) = 0$ toda vez que A não é \mathcal{G} -completo.

Note-se que se V é um \mathcal{G} -potencial, então tem-se que:

$$H_V(\mathbf{x}) = - \sum_{A \in \mathcal{C}(\mathcal{G})} V(\mathbf{x}, A).$$

Definição 3.10 Sejam $\mathcal{G} = (S, \mathcal{V})$ um grafo e V um potencial. Diz-se que uma probabilidade π sobre Ξ é uma *distribuição de Gibbs em relação a \mathcal{G} e V* se:

$$\pi(\mathbf{x}) = Z^{-1} \exp(-H_V(\mathbf{x})),$$

onde a constante Z é a chamada *função partição* do potencial V , e é dada por:

$$Z = \sum_{\mathbf{y} \in \Xi} \exp(-H_V(\mathbf{y})).$$

Para cada $A \subset S$ e cada $\mathbf{x} \in \Xi$ seja $\mathbf{x}^A \in \Xi$ dado por $\mathbf{x}^A = x_s \mathbb{I}_A(s)$

Proposição 3.2 Seja π uma probabilidade nunca nula sobre Ξ . Define-se, para todo $\mathbf{x} \in \Xi$, a função potencial $VC(\pi)$ por:

$$VC(\pi)(\mathbf{x}, A) = \begin{cases} 0 & \text{se } A = \emptyset \\ \sum_{B \subset A} (-1)^{\#(A \setminus B)} \ln(\pi(\mathbf{x}^B)) & \text{caso contrário.} \end{cases}$$

Sejam \mathcal{V}^0 e \mathcal{G}^0 como previamente definidos. Então:

- 1) $VC(\pi)$ é um \mathcal{G}^0 -potencial normalizado, chamado *potencial canônico* de π .
- 2) π é uma distribuição de Gibbs em relação a $VC(\pi)$ e \mathcal{G}^0 .
- 3) Se U é outro \mathcal{G}^0 -potencial normalizado tal que π é uma distribuição de Gibbs em relação a $VC(\pi)$ e \mathcal{G}^0 , então $VC(\pi)$ e U coincidem.

Com estes elementos definidos, passa-se a enunciar um teorema clássico desta área: o Teorema de Hammersley e Clifford. Nele prova-se a equivalência entre campos Markovianos e distribuições de Gibbs.

Teorema 3.1 Suponha-se que $0 \in \Xi_s$ para todo $s \in S$. Sejam o grafo $\mathcal{G} = (S, \mathcal{V})$ e π uma probabilidade nunca nula sobre Ξ . Então:

- 1) Se existe V , \mathcal{G} -potencial, tal que π é de Gibbs com respeito a \mathcal{G} e V , então π é \mathcal{G} -Markoviana.
- 2) Reciprocamente, se π é \mathcal{G} -Markoviana, então $VC(\pi)$ é um \mathcal{G} -potencial e π é de Gibbs em relação a \mathcal{G} e $VC(\pi)$.

Especializando um pouco a teoria anteriormente apresentada, para formular modelos de interesse para processamento de imagens, chega-se ao modelo de Potts-Strauss. Para tanto, seja $S \subset \mathbf{Z}^2$ o suporte (finito) do processo. Supõe-se que é da forma $S = S_1 \times S_2$, por exemplo, $S = \{1, \dots, m\} \times \{1, \dots, n\}$. Às vezes será conveniente escrever $S = \{s_1, \dots, s_N\}$. Sejam Ξ_s conjuntos com um número finito de elementos, iguais para todo $s \in S$, seja $\Xi = \prod_{s \in S} \Xi_s$, e seja o número real $\beta \in \mathbb{R}$. Seja $\mathbf{X}: \Omega \rightarrow \Xi$ uma variável aleatória com distribuição dada por

$$\Pr_{\mathbf{X}}(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{Z_\beta} \exp \left(\beta \sum_{s, t \in S: \|s-t\|=1} \mathbb{1}_{\{x_t\}}(x_s) \right), \quad (3.2)$$

onde $\|s - t\|$ é a distância euclidiana entre os pontos s e t . Diz-se, então, que \mathbf{X} obedece a lei do processo de Potts-Strauss com função de partição Z_β . Quando $\beta < 0$ tem-se um *modelo repulsivo*; quando $\beta > 0$ tem-se um *modelo atrativo*; finalmente, quando $\beta = 0$ tem-se um modelo de independência.

Este modelo, quando considerado $S = \mathbf{Z}^2$, apresenta um comportamento chamado *transição de fase* em um valor especial do parâmetro chamado *parâmetro crítico*. Esse valor, em função de $K = \#\Xi_s$, é dado pela relação $\beta_c = \sqrt{1 + \ln K}$. A transição de fase está relacionada à não unicidade da distribuição de \mathbf{X} , e se reflete no aparecimento de descontinuidades em algumas funções típicas de \mathbf{X} , tais como a média, a correlação entre variáveis aleatórias sobre diferentes coordenadas, etc.

A transição de fase está relacionada à seguinte situação: já foi visto que, para S finito, a família de características locais $\{\pi_s: s \in S\}$ define unicamente a distribuição conjunta μ de \mathbf{X} . Uma pergunta natural é como está caracterizada uma distribuição de Gibbs, μ sobre \mathbf{Z}^2 ? A resposta surge do teorema de Kolmogorov (32): caracterizando a distribuição sobre os retângulos finitos (cilindros) de Ξ^S e cuidando os aspectos referidos à consistência. Quando é empregada esta caracterização usando a família de características locais, é possível fracassar ao tentar definir a distribuição conjunta através das distribuições dos cilindros, pois podem existir probabilidades diferentes μ e π sobre Ξ^S , ambas satisfazendo a propriedade apresentada na equação (3.1) para todo $s \in S$. Esta não unicidade da distribuição conjunta dada uma família de características locais é conhecida como *transição de fase*.

A seguir será visto este fenômeno, apresentado para o modelo de Ising (20,117). Sejam $S = \mathbf{Z}^2$ e $\Xi = \{-1, +1\}$ para todo $s \in S$. Para cada $\Lambda \subset S$ seja Ξ^Λ o conjunto de todas as funções definidas sobre Λ com valores em Ξ . Seja $\Omega = \Xi^S$. Para cada $s = (s_1, s_2)$ e cada $t = (t_1, t_2)$ em \mathbf{Z}^2 seja $\|s-t\| = \sqrt{(s_1 - t_1)^2 + (s_2 - t_2)^2}$. Para cada $s = (s_1, s_2) \in S$ seja $\partial_s = \{t \in S: \|s-t\| = 1\}$. Sejam ω^+ e ω^- os elementos de Ω definidos por $\omega^+(s) = 1$ e $\omega^-(s) = -1$ para cada $s \in S$, e δ^+ e δ^- as probabilidades sobre Ω concentradas sobre ω^+ e ω^- , respectivamente.

Será definida uma distância entre probabilidades sobre Ω . Para cada $n = 1, 2, \dots$ seja $\Lambda(n) = \{(i, j): -n \leq i, j \leq n\}$, a *caixa* de lado $2n+1$ centrada na origem. Se μ e π são duas probabilidades sobre Ω , define-se como distância entre μ e π ao número dado por:

$$d(\mu, \pi) = \sum_{n=1}^{\infty} \sum_{\xi \in \Xi^{\Lambda(n)}} \left| \mu(\sigma_{\Lambda(n)} = \xi) - \pi(\sigma_{\Lambda(n)} = \xi) \right| / 2^n,$$

onde $\sigma_{\Lambda(n)}$ é a projeção de Ω sobre $\Xi^{\Lambda(n)}$, isto é, para cada $\emptyset \neq A \subset S$, a função $\sigma_A: \Xi \rightarrow \Xi_A$ é dada por $\sigma_A(\omega)(t) = x(t)$ para todo $t \in A$ e para todo $\omega \in \Xi$. Para cada $s \in S$ será escrito σ_s no lugar de $\sigma_{\{s\}}$

A seguir, será definido um potencial em relação ao sistema de vizinhanças $\mathcal{V} = \{\partial_s: s \in S\}$. Para tanto, seja \mathcal{C} a família de subconjuntos de S dada

por $\mathcal{C} = \{\{s, t\}: \|s - t\| = 1\}$. Para cada $\Lambda \subset S$ finito, seja ϕ_Λ a função definida sobre Ω por:

$$\phi_\Lambda(\omega) = \begin{cases} -\omega(s)\omega(t) & \text{se } \Lambda = \{s, t\} \in \mathcal{C}, \\ 0 & \text{caso contrário.} \end{cases}$$

Para cada $\beta > 0$ a família de funções $\beta\phi$ definidas sobre $\Lambda \subset S$ finito será chamada *potencial* em relação a \mathcal{V} parametrizado por β .

Uma probabilidade π sobre Ω é de Gibbs em relação a $\beta\phi$ se satisfaz a forma dada na Definição 3.10. Com $G(\beta\phi)$ denota-se o conjunto de todas as distribuições de Gibbs em relação a $\beta\phi$. Poderia acontecer que $\#G(\beta\phi) = 0$ (o potencial não induz nenhuma probabilidade) ou que $\#G(\beta\phi) = 1$ (o potencial induz uma única probabilidade) ou que $\#G(\beta\phi) \geq 2$ (o potencial induz mais de uma probabilidade). Diz-se que o potencial $\beta\phi$ apresenta o fenômeno de transição de fase se existem duas ou mais probabilidades de Gibbs em relação a $\beta\phi$, em símbolos, se $\#G(\beta\phi) \geq 2$.

As provas dos seguintes Teoremas podem ser vistas em (61).

Teorema 3.2 *Dado $\varepsilon > 0$ existe $\beta(\varepsilon) > 0$ tal que se $\beta > \beta(\varepsilon)$ então existem π e μ de Gibbs em relação a $\beta\phi$ tais que $d(\pi, \delta^+) < \varepsilon$ e $d(\mu, \delta^-) < \varepsilon$.*

Teorema 3.3 *Existe $0 < \beta_c < +\infty$ tal que $\beta\phi$ apresenta o fenômeno de transição de fase se e somente se $\beta > \beta_c$. Em particular, para todo $\beta \geq 0$ existem π e μ de Gibbs em relação a $\beta\phi$ tais que:*

- 1) $\mu(\sigma_{(0,0)} = 1) \geq \mu(\sigma_{(0,0)} = -1)$.
- 2) $\pi(\omega) = \mu(-\omega)$ para todo $\omega \in \Omega$, logo $\pi(\sigma_{(0,0)} = 1) < \pi(\sigma_{(0,0)} = -1)$.
- 3) Se ν está em $G(\beta\phi)$ então existe $0 \leq \lambda \leq 1$ tal que $\nu = \lambda\pi + (1 - \lambda)\mu$.

Note-se que, do Teorema 3.3, tem-se que $\pi = \mu$ se e somente se $\beta \leq \beta_c$. O Teorema 3.2 diz que, para valores de β suficientemente grandes (temperaturas suficientemente baixas, na linguagem da mecânica estatística) quando a caixa $\Lambda(n)$ cresce em direção a \mathbf{Z}^2 , as probabilidades μ e π concentram-se em

ω^+ e em ω^- , respectivamente, muito embora as distribuições π e μ possuam as mesmas características locais. Uma forma de se interpretar este resultado consiste em dizer (39) que as correlações entre *pixels* vizinhos se alastram até *pixels* muito afastados. Na prática, embora nas aplicações somente sejam usados como suportes retângulos finitos de \mathbf{Z}^n e, para $\#S < \infty$ não ocorra transição de fase, poderiam ocorrer as seguintes situações:

- quando se simula uma ocorrência da distribuição *a priori* modelada, por exemplo, pelo modelo de Potts-Strauss para $K \geq 2$ classes, se $\beta > \beta_c$ obter $\max_{1 \leq i \leq K} \#\{s \in S: x_s = i\} \approx \#S$, isto é, ocorrências com forte predomínio de uma única classe;
- quando essa distribuição *a priori* é empregada como o termo de suavidade em um problema Bayesiano, se $\beta > \beta_c$ serem obtidas soluções excessivamente suaves, pois predomina o termo de regularidade sobre o termo dos dados.

Na linguagem da Física, a uma temperatura inferior a um certo valor (que, para o modelo de Ising, é conhecido como *temperatura de Curie*) existe um estado de equilíbrio de magnetização positiva sem que haja campo externo, e onde a magnetização desse modelo é definida por $M_\beta = \pi(\sigma_{(0,0)} = 1) - \pi(\sigma_{(0,0)} = -1)$ (ver a equação (3.10) para uma forma explícita deste valor.).

3.3 Estimação pela Equação de Máxima Pseudoverossimilhança

O problema de estimar β dada uma ocorrência de \mathbf{X} é, em geral, intratável computacionalmente devido a que Z_β depende de β de uma forma muito complexa. Lança-se mão, então, da solução da equação de pseudoverossimilhança. Para detalhes a respeito desta técnica de inferência, pode ver-se o trabalho de Jensen e Møller (77) e as referências aí indicadas, e o Capítulo 4 de (25).

Seja $S \subset \mathbf{Z}^2$ um retângulo finito, isto é, existem N_1 e N_2 inteiros positivos tais que $S = \{(i, j): 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$. Sejam $\Xi_s \subset \mathbb{R}$ finito para

cada $s \in S$ e $\Xi = \prod_{s \in S} \Xi_s$. Seja $\Theta \subset \mathbb{R}^k$ um subconjunto aberto.

Para cada $\theta \in \Theta$ seja $p_\theta: \Xi \rightarrow (0, 1)$ tal que $\sum_{\mathbf{x} \in \Xi} p_\theta = 1$, isto é, $p_\theta > 0$ para todo $\mathbf{x} \in \Xi$ e para todo $\theta \in \Theta$. Seja $(\Omega, \mathcal{F}, \mathbb{P}_r)$ um espaço de probabilidade e a variável aleatória $\mathbf{X}: \Omega \rightarrow \Xi$ uma função mensurável. Suponha-se que existe um único valor $\theta_0 \in \Theta$ tal que $\mathbb{P}_r(\mathbf{X} = \mathbf{x}) = p_{\theta_0}(\mathbf{x})$ para todo $\mathbf{x} \in \Xi$; assim sendo, o parâmetro é identificável.

Para cada $\theta \in \Theta$ e cada $s \in S$ seja a função $p_{\theta,s}: \Xi \rightarrow (0, 1)$ definida por $p_{\theta,s}(\mathbf{x}) = \mathbb{P}_r(\sigma_s = x_s \mid \sigma_t = x_t, t \neq s)$, onde \mathbb{P}_r é a probabilidade sobre $(\Xi, \mathcal{P}(\Xi))$ induzida por p_θ , e $\mathcal{P}(\Xi)$ são as partes do conjunto Ξ .

Definição 3.11 *Seja $T \subset S$ não vazio; chama-se função de pseudoverossimilhança sobre T a $PL_T: \Theta \times \Xi \rightarrow (0, 1)$ dada por:*

$$PL_T(\theta, \mathbf{x}) = \prod_{t \in T} p_{\theta,t}(\mathbf{x}).$$

Seja a função $v: \Xi \rightarrow \mathbb{R}^k$. Suponha-se que, para cada $\theta \in \Theta$ vale que:

$$p_\theta(\mathbf{x}) = \frac{1}{Z(\theta)} \exp(\langle \theta, v(\mathbf{x}) \rangle),$$

onde $\langle \theta, v(\mathbf{x}) \rangle = \sum_{j=1}^k \theta_j v_j(\mathbf{x})$ se $\theta = (\theta_1, \dots, \theta_k)^T \in \Theta$, $v(\mathbf{x}) = (v_1(\mathbf{x}), \dots, v_k(\mathbf{x}))^T$ e $Z(\theta) = \sum_{\mathbf{y} \in \Xi} \exp(\langle \theta, v(\mathbf{y}) \rangle)$. Será empregada a seguinte notação para o logaritmo da pseudoverossimilhança: $lPL_T(\theta, \mathbf{x}) = \ln(PL_T(\theta, \mathbf{x})) = \sum_{t \in T} \ln(p_{\theta,t}(\mathbf{x}))$.

Proposição 3.3 *Seja $T \subset S$ como na Definição 3.11. Então:*

$$lPL_T(\Xi, \mathbf{x}) = \sum_{t \in T} \left[\langle \theta, v(\mathbf{x}) \rangle - \ln \left(\sum_{\xi \in \Xi} \exp(\langle \theta, \xi_{\mathbf{x}_{S \setminus \{t\}}} \rangle) \right) \right],$$

onde a configuração $\xi_{\mathbf{x}_{S \setminus \{t\}}} \in \Xi$ é definida para todo $s \in S$ e para todo $\xi \in \Xi_t$ por:

$$\xi_{\mathbf{x}_{S \setminus \{t\}}}(s) = \begin{cases} \xi & \text{se } s = t, \\ \mathbf{x}(s) & \text{caso contrário.} \end{cases}$$

Será empregada a seguinte notação para cada $T \subset S$, cada $\theta \in \Theta$ e cada $\mathbf{y} \in \Xi$:

$$\begin{aligned} \mathbb{E}_{\theta_0, T}(\theta, \mathbf{y}) &= \mathbb{E}_{\theta_0}(l\text{PL}_T(\theta, \mathbf{y}) \mid \mathbf{x}_{S \setminus T} = \mathbf{y}) = \\ &= \sum_{\mathbf{z} \in \Xi} l\text{PL}_T(\theta, \mathbf{z}\mathbf{y}) \mathbb{P}_{r_{\theta_0}}(\sigma_T = \mathbf{z} \mid \sigma_{S \setminus T} = \mathbf{y}). \end{aligned}$$

Proposição 3.4 *Seja $T \subset S$ como na Definição 3.11.*

- 1) *Para cada $\mathbf{x} \in \Xi$ a função $\theta \mapsto l\text{PL}(\theta, \mathbf{x})$ é côncava.*
- 2) *Para cada $\mathbf{y} \in \Xi_{S \setminus T}$ a função $\theta \mapsto \mathbb{E}_{\theta_0, T}(\theta, \mathbf{y})$ é estritamente côncava e possui um máximo em θ_0 se e somente se para cada $\theta \in \Theta$, $\theta \neq \theta_0$ existem $t \in T$ e $\mathbf{z} \in \Xi_T$ tais que $\exists \xi \in \Xi_t$ satisfazendo $\mathbb{P}_{r_{\theta}}(\sigma_t = \xi \mid \sigma_{T \setminus \{t\}} = \mathbf{z}, \sigma_{S \setminus T} = \mathbf{y}) \neq \mathbb{P}_{r_{\theta_0}}(\sigma_t = \xi \mid \sigma_{T \setminus \{t\}} = \mathbf{z}, \sigma_{S \setminus T} = \mathbf{y})$ para algum $\xi \in \Xi_t$, isto é, se o valor θ_0 a ser estimado é único.*

Definição 3.12 *Seja $T \subset S$ como na Definição 3.11. Diz-se que a pseudoverossimilhança é identificável sobre T se para cada $\theta \in \Theta$, $\theta \neq \theta_0$ existem $t \in T$ e $\mathbf{z} \in \Xi_T$ tais que $\mathbb{P}_{r_{\theta}}(\sigma_t = \xi \mid \sigma_{T \setminus \{t\}} = \mathbf{z}, \sigma_{S \setminus T} = \mathbf{y}) \neq \mathbb{P}_{r_{\theta_0}}(\sigma_t = \xi \mid \sigma_{T \setminus \{t\}} = \mathbf{z}, \sigma_{S \setminus T} = \mathbf{y})$ para algum $\xi \in \Xi_t$.*

Seja $\Lambda \subset \mathbb{R}$ um conjunto finito; identificar-se-ão os conjuntos Ξ_s com Λ para todo $s \in S$, para se provar a consistência do estimador de máxima pseudoverossimilhança. Seja $\partial_{(0,0)}$ a vizinhança do ponto $(0,0)$, finita, não vazia e tal que se $t \in \partial_{(0,0)}$ então $-t \in \partial_{(0,0)}$. Seja $D = \sup\{\|t\| : t \in \partial_{(0,0)}\}$. Para cada $s \in \mathbf{Z}^2$ seja $\partial_s = s + \partial_{(0,0)}$. Seja $\mathcal{V} = \{\partial_s : s \in \mathbf{Z}^2\}$; é imediato verificar que $t \in \partial_s \iff s \in \partial_t$.

Definição 3.13 *O conjunto não vazio $C \subset \mathbf{Z}^2$ é chamado “clique” se $\#C = 1$ ou se $s \in C$, $t \in C$ e $s \neq t$ implicam que $s \in \partial_t$. Equivalentemente, para este tipo de vizinhança, se $s - t \in \partial_{(0,0)}$.*

Sejam os conjuntos $\mathcal{C} = \{C \subset \mathbf{Z}^2 : C \text{ é um clique}\}$ e $\mathcal{C}' = \{C \in \mathcal{C} : (0,0) \in C\}$.

Proposição 3.5

$$1) C \in \mathcal{C} \iff \exists C_0 \in \mathcal{C}', t \in \mathbf{Z}^2: C = t + C_0.$$

$$2) C \in \mathcal{C} \implies \#C < \infty.$$

Para cada $\emptyset \neq A \subset \mathbf{Z}^2$ definem-se os conjuntos:

$$\begin{aligned} \mathcal{V}(A) &= \bigcup_{t \in A} (\partial_t \setminus A), \\ \bar{A} &= A \cup \mathcal{V}(A), \\ \overset{\circ}{A} &= \{s \in A: t \notin \partial_s, \forall t \in \mathbf{Z}^2\}. \end{aligned}$$

Será empregada a notação $\mathcal{V}(t) = \mathcal{V}(\{t\})$ e $\bar{t} = \overline{\{t\}}$ para todo $t \in \mathbf{Z}^2$.

Seja $\Xi = \Lambda^{\mathbf{Z}^2} = \{x: \mathbf{Z}^2 \rightarrow \Lambda\}$. Para cada $s \in S$ seja $\tau^s: \Xi \rightarrow \Xi$ a translação de x por s , dada por $\tau^s(x)(t) = x(s+t)$ para todo $t \in \mathbf{Z}^2$. Seja $\mathcal{S} = \{A \subset \mathbf{Z}^2: 0 < \#A < \infty\}$. Para cada $A \in \mathcal{S}$ seja $v_A \in M(\sigma_A)$ com $M(\sigma_A) = \{\phi \circ \sigma_A \text{ tal que } \phi: \Lambda^A \rightarrow \mathbb{R}^k\}$. Seja $v = (v_A)_{A \in \mathcal{S}}$; diz-se que v é um potencial. Suponha-se que:

- 1) v é invariante por translações, isto é, $v_A \circ \tau^s = v_{A+s}$ para todo $s \in \mathbf{Z}^2$ e todo $A \in \mathcal{S}$;
- 2) v é um \mathcal{V} -potencial, isto é, $v_A = 0$ se $A \notin \mathcal{C}$.

Para cada $S \subset \mathbf{Z}^2$, retângulo finito, seja v^S a única função sobre Λ^S tal que:

$$v^S(e_S) = \sum_{C \in \mathcal{C}: C \subset S} v_C(e) = \sum_{C \in \mathcal{C}: C \in \mathcal{S}} \frac{1}{\#(A(C))} \sum_{(C_0, t) \in A(C)} v_{C_0+t}(e) \forall e \in \Xi,$$

onde $A(C) = \{(C_0, t) \in \mathcal{C}_0 \times \mathbf{Z}^2: C_0 + t = C\}$.

Seja $(S(n))_{n \geq 1}$ uma sucessão de retângulos finitos em \mathbf{Z}^2 tal que:

- $\mathbf{Z}^2 = \bigcup_{n \geq 1} S(n)$;
- Seja $\nu \geq 1$ inteiro e sejam os naturais $m_{n,i} \in \mathbb{N}$ para cada $n = 1, 2, \dots$ e para cada $i = 1, \dots, \nu$ tais que $m_{n,i} \rightarrow +\infty$ quando $n \rightarrow +\infty$ para todo $i = 1, \dots, \nu$;

- para cada $1 \leq i \leq \nu$, cada $n = 1, 2, \dots$ e cada $1 \leq j \leq m_{n,i}$ tem-se um conjunto $A_{n,i,j} \subset \mathbf{Z}^2$ satisfazendo:
 - 1) $A_{n,i,j} \cup A_{n,i',j'} = \emptyset$ se $i \neq i'$ ou se $j \neq j'$;
 - 2) $\delta = \sup\{\#(A_{n,i,j})\} > \infty$;
 - 3) $d(A_{n,i,j}, A_{n,i,k}) > D$ se $j \neq k$;
 - 4) $S(n) = \bigcup_{i=1}^{\nu} \bigcup_{j=1}^{m_{n,i}} A_{n,i,j}$.

É possível ver que $S(n)$ é um retângulo finito para todo n . Para cada $\theta \in \Theta$ e cada $n = 1, 2, \dots$ seja $p_{\theta}^{(n)}: \Lambda^{S(n)} \rightarrow (0, 1)$ definida por:

$$p_{\theta}^{(n)}(\mathbf{x}_{S(n)}) = \frac{\exp\left(\langle \theta, v^{S(n)}(\mathbf{x}_{S(n)}) \rangle\right)}{Z_n(\theta)},$$

onde $Z_n(\theta) = \sum_{\mathbf{y} \in \Lambda} \exp\left(\langle \theta, v^{S(n)}(\mathbf{x}_{S(n)}) \rangle\right)$.

Seja $(\Omega, \mathcal{F}, \mathbb{P}\text{r})$ um espaço de probabilidade. Para cada $n = 1, 2, \dots$ seja $X_n: \Omega \rightarrow \Lambda^{S(n)}$ uma variável aleatória, função mensurável. Suponha-se que existe um único valor $\theta_0 \in \Theta$ tal que $\mathbb{P}\text{r}(\mathbf{X}_n = \mathbf{x}_{S(n)}) = p_{\theta_0}^{(n)}(\mathbf{x}_{S(n)})$, para todo $\mathbf{x}_{S(n)} \in \Lambda_{S(n)}$, e suponha-se que se $\theta \neq \theta_0$ com $\theta \in \Theta$, então existe $\mathbf{x}_{(0,0)} \in \Lambda_{(0,0)}^{\overline{(0,0)}}$ tal que, para todo n vale que:

$$\mathbb{P}\text{r}_{\theta}^{(n)}(\sigma_{(0,0)} = x_{(0,0)} \mid \sigma_{\partial(0,0)} = \mathbf{x}_{\partial(0,0)}) \neq \mathbb{P}\text{r}_{\theta_0}^{(n)}(\sigma_{(0,0)} = x_{(0,0)} \mid \sigma_{\partial(0,0)} = \mathbf{x}_{\partial(0,0)}),$$

isto é, uma condição de identificabilidade.

Teorema 3.4 Para cada $\varepsilon > 0$, se $n \rightarrow \infty$, vale que:

$$\mathbb{P}\text{r}\left(\left\{\omega \in \Omega: \text{ a função } \theta \mapsto \text{lPL}_{S(n)}^{\circ}(\theta, X_n(\omega)) \text{ tem um único ponto de máximo } \widehat{\theta}_n(\omega), \text{ com } \|\widehat{\theta}_n(\omega) - \theta_0\| < \varepsilon \right\}\right) \rightarrow 1.$$

Este Teorema prova a consistência assintótica do estimador de máxima pseudoverossimilhança na sua forma geral. Nas aplicações de interesse para este trabalho, será empregado o estimador do parâmetro de atratividade do modelo de Potts-Straus.

A seguir é desenvolvido o cálculo do estimador de máxima pseudoverossimilhança para o modelo de Potts-Strauss, de forma tal a se ter estimadores computacionalmente tratáveis para todos os casos deste modelo. Desconhecem-se resultados análogos na literatura.

Pode calcular-se $\hat{\beta} = \hat{\beta}(\mathbf{x})$, o estimador de máxima pseudoverossimilhança do modelo de Potts-Strauss (definido na equação (3.2)), como a solução da seguinte equação:

$$\sum_{s \in S} \left[v_s(x_s) - \frac{\sum_{y \in \Xi_s} v_s(y) \exp(\hat{\beta} v_s(y))}{\sum_{y \in \Xi_s} \exp(\hat{\beta} v_s(y))} \right] = 0, \quad (3.3)$$

onde $v_s(t) = \#\{u \in \partial_s: x_u = t\}$, e ∂_s é a vizinhança do ponto $s \in S$.

Este estimador possui boas propriedades de consistência e eficiência assintóticas, porém, nessa forma geral, é computacionalmente muito *caro*.

A primeira simplificação que pode ser introduzida consiste em se supor que os conjuntos Ξ_s são iguais para todo $s \in S$. Assim, tem-se que $\Xi = (\Xi_s)^S$. Ainda mais, por simplicidade notacional e sem perda de generalidade, suponha-se que Ξ_s é um subconjunto de inteiros, por exemplo, $\Xi_s = \{0, \dots, K - 1\}$.

Outra simplificação de ordem computacional consiste em se usar somente aquelas coordenadas de S que tenham o mesmo número de vizinhos. Se $S = \{1, \dots, m\} \times \{1, \dots, n\}$ e $\partial_s = \{t \in S: \|s - t\| = 1\}$, definem-se $E = \{(1, 1), (1, m), (1, n), (m, n)\}$ o conjunto das esquinas, $W = \{s = (i, j) \in S: 1 < i < m, 1 < j < n\}$ o *corpo*, e $B = S \setminus (W \cup E)$, as bordas. Pode redefinir-se, então, $\hat{\beta}(\mathbf{x})$ como a solução da seguinte equação:

$$\sum_{s \in W} \left[v_s(x_s) - \frac{\sum_{y \in \Xi_s} v_s(y) \exp(\hat{\beta} v_s(y))}{\sum_{y \in \Xi_s} \exp(\hat{\beta} v_s(y))} \right] = 0. \quad (3.4)$$

Desta maneira, o problema de se achar $\hat{\beta}$ pode dividir-se em três casos simples, para os quais será empregada a seguinte notação:

$$\Upsilon(t) = \frac{\exp(t)}{1 + \exp(t)}.$$

Pode ver-se facilmente que $\Upsilon(t) = 1 - \Upsilon(-t)$.

Quando $\Xi_s = \{0, 1\}$ para todo $s \in S$, tem-se que $\hat{\beta}$ é dado por:

$$\begin{aligned} & 2(C_4^1 - C_0^1 + C_0^0 + C_4^1 + C_3^1) - C_1^1 + C_1^0 - 2(C_0^0 + 2C_4^1 + C_4^0)\Upsilon(4\hat{\beta}) - \\ & - (C_1^0 + 2C_3^1 + C_3^0)\Upsilon(2\hat{\beta}) = 0, \end{aligned} \quad (3.5)$$

onde $C_k^\ell = \#\{s \in W : x_s = \ell, \sum_{t \in \partial_s} x_t = k\}$.

Seja a vizinhança do ponto s escrita como $\partial_s = \{t, u, v, w\}$.

Serão de interesse os seguintes valores para o caso $\Xi_s = \{0, 1, 2\}$:

$$\begin{aligned} V^4 &= \#\{x_{\partial_s} \subset W : x_t = x_u = x_v = x_w\}, \\ V^3 &= \#\{x_{\partial_s} \subset W : x_t = x_u = x_v \neq x_w\}, \text{ e} \\ V^2 &= \#\{x_{\partial_s} \subset W : x_t = x_u \neq x_v = x_w\}. \end{aligned}$$

Desta forma, podemos escrever a equação (3.4) como:

$$\sum_{s \in W} v_s(x_s) - 4V^4\Upsilon(4\hat{\beta}) - V^3 \frac{3\exp(3\hat{\beta}) + \exp(\hat{\beta})}{\exp(3\hat{\beta}) + \exp(\hat{\beta})} - 2V^2 = 0. \quad (3.6)$$

Para 4 ou mais classes, i.e., para o caso $\Xi_s = \{0, 1, 2, \dots, K-1\}$ e $K \geq 3$ basta calcular os seguintes valores:

$$\begin{aligned} K^4 &= \#\{x_{\partial_s} \subset W : x_t = x_u = x_v = x_w\}, \\ K^3 &= \#\{x_{\partial_s} \subset W : x_t = x_u = x_v \neq x_w\}, \\ K^2 &= \#\{x_{\partial_s} \subset W : x_t = x_u \neq x_v = x_w\}, \\ K^1 &= \#\{x_{\partial_s} \subset W : x_t = x_u \neq x_v \neq x_w \neq x_t\}, \\ K^0 &= \#\{x_{\partial_s} \subset W : x_i \neq x_j \forall i \neq j\}. \end{aligned}$$

Assim, a equação (3.4) reduz-se a:

$$\sum_{s \in W} v_s(x_s) - 4K^4 \frac{\exp(4\hat{\beta})}{\exp(4\hat{\beta}) + K} - K^3 \frac{3\exp(3\hat{\beta}) + \exp(\hat{\beta})}{\exp(3\hat{\beta}) + \exp(\hat{\beta}) + K - 1} -$$

$$\begin{aligned}
& -K^2 \frac{4 \exp(2\hat{\beta})}{2 \exp(2\hat{\beta}) + K - 1} - K^1 \frac{2(\exp(2\hat{\beta}) + \exp(\hat{\beta}))}{\exp(2\hat{\beta}) + 2 \exp(\hat{\beta}) + K - 2} - \\
& -K^0 \frac{4 \exp(\hat{\beta})}{4 \exp(\hat{\beta}) + K - 3} = 0.
\end{aligned} \tag{3.7}$$

É importante se notar que as constantes envolvidas nas equações (3.5), (3.6) e (3.7) podem ser determinadas com uma única leitura da configuração sob estudo. Assim, o custo computacional de se determinar $\hat{\beta}$ é relativamente baixo.

3.4 Simulação do modelo de Potts-Strauss

A simulação de ocorrências do processo de Potts-Strauss tem despertado atenção de pesquisadores de diferentes áreas. Para os de Mecânica Estatística, a simulação é interessante para se estimar certas quantidades características (por exemplo, predomínio de uma classe sobre as outras, e correlações a curta e longa distância, entre outras). Para os de Ciência da Computação, a simulação usando certas técnicas pode fornecer informação preciosa a respeito da qualidade dos algoritmos de geração de números pseudoaleatórios: um assunto muito importante, levantado em (68). Para os de Processamento de Imagens o interesse gira em torno de duas aplicações principais (50,66):

- 1) simular eventos de uma das distribuições que podem ser usadas como modelos *a priori* para inferência Bayesiana;
- 2) empregar os algoritmos de simulação como formas de se obter restaurações e segmentações de imagens (o estimador MAP, implementado com o algoritmo de *simulated annealing*, o estimador MPM, implementado com o algoritmo *Gibbs sampler* (59), por exemplo).

Em (35) é apresentada uma técnica parecida às que serão comentadas neste trabalho para simulação de ocorrências do modelo de Ising e, usando essas ocorrências, são formadas misturas para gerar ocorrências do processo que modela as imagens observadas. O problema nesse trabalho é que o algoritmo empregado converge a ocorrências de uma distribuição diferente da proposta, devido a que o

número de *pixels* por classe é mantido fixo entre iterações sucessivas. Isto é, não é satisfeita a condição de positividade, necessária para o modelo de Potts-Strauss, pois muitas configurações não são permitidas. Em outras palavras, a dinâmica de simulação não é reversível em relação à distribuição alvo, isto é, existem configurações da distribuição alvo inatingíveis pela dinâmica empregada e, portanto, é impossível garantir a convergência.

O estudo e o uso de dinâmicas de aglomerados têm ficado restritos aos aspectos teóricos e a aplicações em Mecânica Estatística. Desconhece-se, até a data de redação deste trabalho, a aplicação destes algoritmos à simulação de imagens.

3.4.1 Dinâmicas de *spin-flip*

As dinâmicas de *spin-flip* são aquelas que permitem a mudança de estado de apenas uma coordenada por passo. Em geral, elas apresentam uma convergência menos rápida do que as dinâmicas de *cluster-flip*, porém, permitem a simulação de distribuições mais gerais do que as que podem ser simuladas com estas últimas. Dispõe-se de duas dinâmicas deste tipo: os algoritmos de Metropolis e o de Geman e Geman.

3.4.1.1 Algoritmo de Metropolis

Este algoritmo, proposto nos anos '50, é o mais geral para a simulação de distribuições com muitas componentes. Consiste de, em cada passo, *tentar* trocar o estado de uma coordenada. Se o novo estado aumenta a probabilidade da configuração total (isto é, diminui a sua energia), aceitá-lo. Se o novo estado diminui a probabilidade da configuração total (isto é, aumenta a sua energia), aceitá-lo com uma probabilidade que depende da diferença de energias.

Descrito em forma algorítmica, passar da configuração $\mathbf{x}(k)$ à configuração $\mathbf{x}(k+1)$ com probabilidade de transição dada por:

$$\mathbb{P}\text{r}(\mathbf{X}(k+1) = \mathbf{x}(k+1) \mid \mathbf{X}(k) = \mathbf{x}(k)) = \begin{cases} 1 & \text{se } \frac{\mathbb{P}\text{r}(\mathbf{X}(k+1)=\mathbf{x}(k+1))}{\mathbb{P}\text{r}(\mathbf{X}(k)=\mathbf{x}(k))} > 1 \\ \frac{\mathbb{P}\text{r}(\mathbf{X}(k+1) = \mathbf{x}(k+1))}{\mathbb{P}\text{r}(\mathbf{X}(k) = \mathbf{x}(k))} & \text{caso contrário.} \end{cases}$$

Para garantir a convergência à distribuição alvo é necessário, a partir de qualquer configuração inicial, atualizar infinitas vezes todas as coordenadas de suporte do processo.

Dado que a mudança de $\mathbf{x}(k)$ para $\mathbf{x}(k+1)$ é arbitrária, este algoritmo não necessariamente é da classe *spin-flip*. Porém, se as configurações $\mathbf{x}(k)$ e $\mathbf{x}(k+1)$ diferirem em mais de uma coordenada, o cômputo do quociente $\mathbb{P}\text{r}(\mathbf{X}(k) = \mathbf{x}(k))/\mathbb{P}\text{r}(\mathbf{X}(k+1) = \mathbf{x}(k+1))$ é computacionalmente inviável. Por isto, costuma-se considerar este algoritmo como pertencente à classe *spin-flip*.

3.4.1.2 Algoritmo de Geman e Geman

Este algoritmo, também conhecido como *Gibbs sampler*, aproveita a idéia do algoritmo de Metropolis e, em vez de deixar livre a maneira de se mudar de $\mathbf{x}(k)$ para $\mathbf{x}(k+1)$, a especifica em função das probabilidades condicionais que caracterizam um campo Markoviano.

Assim sendo, a probabilidade de transição entre as configurações $\mathbf{x}(k)$ e $\mathbf{x}(k+1)$, quando visitada a coordenada s , é dada por:

$$\mathbb{P}\text{r}(\mathbf{X}(k+1) = \mathbf{x}(k+1) \mid \mathbf{X}(k) = \mathbf{x}(k)) = \begin{cases} 0 & \text{se } \mathbf{x}_{S \setminus \{s\}}(k+1) \neq \mathbf{x}_{S \setminus \{s\}}(k) \\ \mathbb{P}\text{r}(X_s(k+1) = x_s(k+1) \mid \mathbf{X}_{\partial_s}(k) = \mathbf{x}_{\partial_s}(k)) & \text{se } \mathbf{x}_{S \setminus \{s\}}(k+1) = \mathbf{x}_{S \setminus \{s\}}(k). \end{cases}$$

Para garantir a convergência à distribuição desejada a partir de uma configuração inicial arbitrária, todas as posições $s \in S$ devem ser visitadas infinitas vezes (ver (25), por exemplo).

Como este algoritmo permite incorporar explicitamente a influência de campo externo nas probabilidades condicionais, é o mais empregado (embutido num esquema de *simulated annealing*) para resolver problemas de processamento e análise de imagens de porte combinatório. O campo externo incorpora a informação dos dados (imagem observada), enquanto a distribuição *a priori* (o modelo de Potts-Strauss, por exemplo) impõe condições de regularidade à solução do problema.

3.4.2 Dinâmicas de Aglomerados

No final dos anos '80 apareceram duas propostas novas para a simulação de ocorrências do modelo de Potts-Strauss: as dinâmicas de Swendsen-Wang e Wolff (38,130,131,137). Até então dispunha-se apenas de algoritmos de *spin-flip*, isto é, algoritmos que passo a passo podiam alterar o valor de apenas uma coordenada do suporte. As novas dinâmicas, chamadas de *cluster-flip*, permitiam a mudança de mais de uma posição simultaneamente. Estes algoritmos de aglomerados vieram permitir convergências mais rápidas que as obtidas com os algoritmos de Metropolis e de Geman e Geman (21,51,53) e, além disso, menor correlação entre ocorrências de iterações sucessivas. Desta maneira, foi possível realizar experiências Monte Carlo para a medição de quantidades sobre o modelo de Potts-Strauss a menor custo computacional que o permitido pelas dinâmicas de *spin-flip*.

Para ver a prova rigorosa da convergência dos algoritmos de aglomerados pode-se ver, entre outras, as referências (101,102).

O esquema que se apresenta a seguir é o dos algoritmos de Swendsen-Wang e de Wolff para a geração de uma amostra de tamanho um de \mathbf{X} , modelo de Potts-Strauss com $\Xi_s = \{1, \dots, K\}$, $K \geq 2$, para todo $s \in S$ com distribuição dada na equação (3.2). Denota-se $p = \text{IPr}_{\mathbf{X}}(\mathbf{x}) = \text{IPr}(\mathbf{X} = \mathbf{x})$, e supõe-se p conhecida.

Em palavras, o algoritmo de Swendsen-Wang consiste de, a partir de uma configuração inicial qualquer, iterar da seguinte maneira enquanto alguma

condição de convergência não for satisfeita:

- 1) Formar um grafo sobre a configuração atual colocando um arco, entre todas as posições vizinhas que exibem o mesmo valor, com probabilidade $1 - \exp(-\beta)$.
- 2) Obter o conjunto de todas as componentes conexas do grafo obtido no passo anterior.
- 3) Se os elementos da componente conexa j , formada no passo anterior, têm o valor $x_j \in \Xi_s$, substituir esses valores pelo novo valor resultante de observar a variável aleatória $X \sim \mathcal{U}(\Xi_s)$. Note-se, no entanto, que os valores da componente conexa podem não ser alterados.

Em palavras, o algoritmo de Wolff (que nada mais é do que uma modificação do algoritmo de Swendsen-Wang) consiste de, a partir de uma configuração inicial qualquer, iterar da seguinte maneira enquanto alguma condição de convergência não for satisfeita:

- 1) Formar um grafo sobre a configuração atual colocando um arco, entre todas as posições vizinhas que exibem o mesmo valor, com probabilidade $1 - \exp(-\beta)$.
- 2) Obter o conjunto de todas as componentes conexas do grafo obtido no passo anterior.
- 3) Escolher uma única componente conexa do grafo obtido no passo anterior, sendo que cada uma das componentes conexas tem uma probabilidade de ser escolhida proporcional ao seu tamanho.
- 4) Se os elementos da componente conexa j , formada no passo anterior, têm o valor $x_j \in \Xi_s$, substituir esses valores pelo novo valor resultante de observar a variável aleatória $X \sim \mathcal{U}(\Xi_s \setminus \{x_j\})$. Note-se que os valores da componente conexa escolhida são sempre trocados.

A configuração inicial $\mathbf{x}(0)$ para ambos algoritmos pode ser obtida, a baixo custo computacional, tomando uma amostra de tamanho um de

uma das seguintes distribuições:

$$p_0 = p_0(\mathbf{x}(0)) = \begin{cases} \frac{1}{\#\Xi_s} & \text{se } x_s(0) = \xi \in \Xi_s \forall s \in S \\ 0 & \text{caso contrário,} \end{cases} \quad (3.8)$$

$$\text{ou } p_\infty = p_\infty(\mathbf{x}(0)) = \frac{1}{\#\Xi} \mathbb{1}_{\Xi}(\mathbf{x}(0)), \quad (3.9)$$

e se dirá que $\mathbf{x}(0)$, provindo da distribuição p_T , com $T = 0$ ou $T = \infty$, foi *obtido à temperatura* T . Para se obter $\mathbf{x}(0)$ a $T = 0$ basta observar uma variável aleatória com distribuição $\mathcal{U}(\Xi_s)$, e atribuir o valor observado a cada coordenada de $s \in S$. Para se obter $\mathbf{x}(0)$ a $T = \infty$, é preciso atribuir em cada coordenada $s \in S$ a ocorrência de uma variável aleatória distribuída como $\mathcal{U}(\Xi_s)$, em forma independente.

Estas duas formas de se obter $\mathbf{x}(0)$ aparecerão na Seção 3.4.3, onde são empregadas para a determinação de k_* e t_* , e de k^* e t^* : o número de iterações e tempos de CPU para a convergência dos algoritmos de *cluster-flip* quando a configuração inicial é obtida das distribuições (3.9) e (3.8), respectivamente.

Para escrever com rigor os algoritmos de aglomerados é necessário definir um novo elemento sobre o grafo (S, \mathcal{V}) : uma *ordenação* de nodos. Assim sendo, seja $\mathcal{L}' \in S \times S$ tal que:

- 1) $(s, t) \in \mathcal{L}'$ implica que $(t, s) \notin \mathcal{L}'$;
- 2) $(s, t) \in \mathcal{L}'$ implica que $t \in \partial_s$.

Seja ainda $\mathcal{L} = \mathcal{L}' \cup \{(o, o) : o \in S\}$.

3.4.2.1 Algoritmo de Swendsen-Wang

Sejam: $\beta > 0$, $\mathbf{x}(0)$ uma amostra de tamanho um de $\mathbf{X}(0)$ com distribuição p_0 ou p_∞ ; $M > 0$ um inteiro fixado pelo critério de parada escolhido (que será o objeto de estudo da Seção 3.4.3); e o número de iterações $k = 0$. Seja ainda $\mathcal{G} = (S, \mathcal{V})$ o grafo tal que \mathbf{X} é um \mathcal{G} -CAM (isto é, $\mathcal{V} = \{\partial_s : s \in S\}$, o conjunto das vizinhanças de S).

- 1) Definir o conjunto $\mathcal{L}_1(k) = \{(o, q) \in \mathcal{L} : x(k)_o = x(k)_q\}$. Se $\mathcal{L}_1(k) = \emptyset$ ir ao passo 4).

- 2) Gerar $\{\varepsilon_{o,q}: (o, q) \in \mathcal{L}_1(k)\}$ amostra de tamanho $\#\mathcal{L}_1(k)$ da distribuição $\mathcal{B}(1 - \exp(-\beta))$.
- 3) Definir o conjunto: $\mathcal{L}_3(k) = \{(o, q) \in \mathcal{L}_1(k): \varepsilon_{o,q} = 1\}$.
- 4) Definir o conjunto $\tilde{\mathcal{L}}(k) = \{(o, o) \in S \times S: (o, q) \notin \mathcal{L}_3(k) \forall q \in \partial_o\} \cup \mathcal{L}_3(k)$.
- 5) Obter a família $\Lambda(k) = \{\lambda_{1,k}, \dots, \lambda_{L_k,k}\}$ de todas as componentes conexas do grafo $(S, \tilde{\mathcal{L}}(k))$. Note-se que:
 - (a) $(S, \tilde{\mathcal{L}}(k))$ pode não ser um subgrafo de S, \mathcal{L} .
 - (b) A família $\Lambda(k)$ é uma partição de S pois $S = \cup_{m=1}^{L_k} \lambda_{m,k}$ e $\lambda_{m,k} \cap \lambda_{m',k} = \emptyset$ se $m \neq m'$.
- 6) Gerar $\{\eta_1, \dots, \eta_{L_k}\}$, amostra de tamanho L_k da $\mathcal{U}(\Xi_s)$.
- 7) Para cada $m = 1, \dots, L_k$ e cada $o \in \lambda_{m,k}$ atribuir $x_o(k+1) = \eta_m$. Note-se que, dado que o novo valor para cada componente conexa resulta de observar uma variável aleatória uniformemente distribuída em Ξ_s , as coordenadas de cada componente conexa podem não trocar de valor entre iterações sucessivas. De fato, a probabilidade disto acontecer é $1 - (\#\Xi_s)^{-1}$.
- 8) Se $k+1 < M$ então fazer $k = k+1$ e voltar ao passo 1. Caso contrário finalizar atribuindo $\mathbf{x} = \mathbf{x}(k)$.

3.4.2.2 Algoritmo de Wolff

Sejam: $\beta > 0$, $\mathbf{x}(0)$ uma amostra de tamanho um de $\mathbf{X}(0)$ com distribuição p_0 ou p_∞ ; $M > 0$ um inteiro fixado pelo critério de parada escolhido e o número de iterações $k = 0$. Seja ainda $\mathcal{G} = (S, \mathcal{V})$ o grafo tal que \mathbf{X} é um \mathcal{G} -CAM (isto é, $\mathcal{V} = \{\partial_s: s \in S\}$, o conjunto das vizinhanças de S).

- 1) Definir o conjunto $\mathcal{L}_1(k) = \{(o, q) \in \mathcal{L}: x(k)_o = x(k)_q\}$. Se $\mathcal{L}_1(k) = \emptyset$ ir ao passo 4).
- 2) Gerar $\{\varepsilon_{o,q}: (o, q) \in \mathcal{L}_1(k)\}$ amostra de tamanho $\#\mathcal{L}_1(k)$ da distribuição $\mathcal{B}(1 - \exp(-\beta))$.
- 3) Definir o conjunto: $\mathcal{L}_3(k) = \{(o, q) \in \mathcal{L}_1(k): \varepsilon_{o,q} = 1\}$.

- 4) Definir o conjunto $\tilde{\mathcal{L}}(k) = \{(o, o) \in S \times S: (o, q) \notin \mathcal{L}_3(k) \forall q \in \partial_o\} \cup \mathcal{L}_3(k)$.
- 5) Obter a família $\Lambda(k) = \{\lambda_{1,k}, \dots, \lambda_{L_k,k}\}$ de todas as componentes conexas do grafo $(S, \tilde{\mathcal{L}}(k))$. Note-se que:
 - (a) $(S, \tilde{\mathcal{L}}(k))$ pode não ser um subgrafo de S, \mathcal{L} .
 - (b) A família $\Lambda(k)$ é uma partição de S pois $S = \cup_{m=1}^{L_k} \lambda_{m,k}$ e $\lambda_{m,k} \cap \lambda_{m',k} = \emptyset$ se $m \neq m'$.
- 6) Gerar $\{u_1, \dots, u_{L_k}\}$ amostra de tamanho L_k da $\mathcal{U}(0, 1)$, e seja $L \in \{1, \dots, L_k\}$ tal que $u_L > u_{L'}$ para todo $L \neq L'$ (a existência de L está garantida pelo fato de provir de uma densidade).
- 7) Gerar y , amostra de tamanho um da distribuição $\mathcal{U}(\Xi_s \setminus \{x_q\})$, com $q \in \lambda_L$.
- 8) Atribuir $x_q(k+1) = y$ para todo $q \in \lambda_L$. Note-se que, dado que o novo valor para cada componente conexa resulta de observar uma variável aleatória uniformemente distribuída em Ξ_s , menos o valor atual das coordenadas dessa componente, os valores das coordenadas de cada componente conexa necessariamente mudam de valor entre iterações sucessivas.
- 9) Se $k+1 < M$ então fazer $k = k+1$ e voltar ao passo 1. Caso contrário finalizar atribuindo $\mathbf{x} = \mathbf{x}(k)$.

É possível implementar versões mais gerais destes algoritmos: para valores negativos de β , para outras definições de vizinhança e para diferentes valores de atratividade dentro da vizinhança. Porém, até a data de redação deste trabalho, não se sabia como introduzir um campo externo não homogêneo. Esta limitação faz com que este algoritmo (e o de Wolff também) não possa ser empregado para implementar a dinâmica de *simulated annealing*.

3.4.2.3 Obtenção das Componentes Conexas de um Grafo

Um possível algoritmo para a obtenção das componentes conexas de um grafo é o seguinte. Seja $T = \{t_1, \dots, t_M\}$, com $M \geq 2$, um conjunto, e $\mathcal{G} = \{T, \mathcal{L}\}$ um grafo definido sobre T . Isto é, $\mathcal{L} \subset \mathcal{P}_2(T)$, onde $\mathcal{P}_2 = \{T_1: T_1 \subset T, \#T_1 \leq 2\}$ e satisfaz que se $t \in T$ então existe $s \in T$ tal que $\{s, t\} \in \mathcal{L}$ (s poderia

ser o mesmo t). Noutras palavras, \mathcal{L} é uma família de subconjuntos de um ou dois elementos de T tais que para $t \in T$, $\{t\} \in \mathcal{L}$ ou $\{t, s\} \in \mathcal{L}$ para algum $s \in T$ com $s \neq t$. Seja V a matriz $M \times M$ definida por

$$V(i, j) = \begin{cases} 0 & \forall i = j, \dots, M, 1 \leq i \leq M \\ \mathbb{I}_{\mathcal{L}}(\{t_i, t_j\}) & \text{caso contrário;} \end{cases}$$

isto é, V é a matriz com um um na posição (i, j) se $i \neq j$ e se t_j é vizinho de t_i : a matriz de conexidade do grafo. Observe-se que V é simétrica com zeros na diagonal.

Sejam s e t em T (s poderia ser igual a t). Diz-se que existe um caminho entre s e t se existem t_0, \dots, t_m , com $m \geq 1$, em T tais que:

$$s = t_0,$$

$$t = t_m \text{ e}$$

$$\{t_0, t_1\}, \{t_1, t_2\}, \dots, \{t_{m-1}, t_m\} \text{ estão em } \mathcal{L}.$$

Denote-se agora " \sqsubseteq " a relação em T definida por $s \sqsubseteq t$ se existe um caminho entre s e t ; então " \sqsubseteq " é uma relação de equivalência. Para cada $t \in T$ seja $\sqsubseteq(t) = \{s \in T : s \sqsubseteq t\}$. Diz-se que $\lambda \subset T$ é uma componente conexa do grafo \mathcal{G} se existe $t \in T$ tal que $\lambda = \sqsubseteq(t)$.

A seguir apresenta-se uma algoritmo para calcular as componentes conexas do grafo \mathcal{G} , que é uma pequena modificação do algoritmo apresentado em (122). Nessa referência, o algoritmo é apresentado em forma recursiva, e foi essa a maneira em que foi implementado (ver Apêndice B). Porém, para facilitar a compreensão do algoritmo, e para possibilitar implementações em linguagens de programação que não admitem estruturas de recursividade (como FORTRAN e BASIC, por exemplo), a seguir apresenta-se uma variante do algoritmo original.

Seja C um vetor M -dimensional, e sejam $C(1) = \dots = C(M) = 0$, $c = 0$ e $n = 1$.

- 1) Se $n > M$ então finalizar. Se não, prosseguir no passo 2.
- 2) Se $C(n) = 0$ prosseguir no passo 3. Se não atribuir $n = n + 1$ e prosseguir no passo 1.

- 3) Atribuir $c = c + 1$. Chamar ao procedimento $C'(n, V(n, \cdot), c)$, onde $V(n, \cdot) = (V(n, 1), \dots, V(n, M))$. Ao retornar do procedimento C' prosseguir no passo 4.
- 4) Atribuir $n = n + 1$.
- 5) Voltar ao passo 1.

Procedimento $C'(x, V, \ell)$: x é um inteiro entre 1 e M ; V é um vetor M -dimensional com $V(i) = 0$ ou 1 para todo $1 \leq i \leq M$; ℓ é um inteiro. Este procedimento tem acesso às variáveis C e V , definidas no procedimento que o chama.

- 1) Atribuir $C(x) = \ell$.
- 2) Atribuir $i = 1$.
- 3) Se $V(i) = 1$ e se $C(i) = 0$ então chamar $C'(i, V(i, \cdot), \ell)$. Se não, prosseguir no passo 4 deste mesmo procedimento.
- 4) Atribuir $i = i + 1$.
- 5) Se $i > M$ então retornar ao procedimento que chamou C' . Se não, retornar ao passo 3 deste procedimento.

3.4.3 Tempos de Convergência

O problema computacional das dinâmicas de aglomerados que mais atenção tem recebido é o que diz respeito à correlação entre quantidades medidas sobre iterações sucessivas (4,138). De fato, esse é um dos problemas centrais quando se deseja estudar processos que exibem transição de fase (12,13,110). Empregando-se essa informação é possível inferir a respeito de propriedades interessantes destes processos (106,107,121)

Um assunto relevante, e não encontrado na literatura, é o que diz respeito ao tempo que diferentes algoritmos iterativos levam para atingir uma configuração *suficientemente* parecida com uma provinda da verdadeira distribuição sob estudo. Nos trabalhos (21,53) verificou-se que a dinâmica de Swendsen-Wang

converge mais rapidamente ao equilíbrio, para o modelo de Ising, do que a do *Gibbs sampler*, quando a configuração inicial é uma ocorrência a temperatura infinita e o parâmetro é supercrítico. Em (51) foi feita uma experiência Monte Carlo para se determinar qual das duas dinâmicas de aglomerado, Swendsen-Wang ou Wolff, é a mais veloz para configurações iniciais a temperatura infinita.

A técnica empregada em (51) foi a seguinte: para a determinação de k^* , o número de iterações necessárias para atingir o equilíbrio a partir de $\mathbf{x}(0)$ simulada a $T = \infty$, definir:

$$k^* = \inf\{k \in \mathbb{N}: \hat{\beta}(\mathbf{x}(k)) \geq \beta\}.$$

Como o valor do parâmetro β ao qual foi obtida a ocorrência $\mathbf{x}(0)$ é zero (pois $T = \beta^{-1} = \infty$) espera-se, pelas boas propriedades do estimador de máxima pseudo-verossimilhança, que $\hat{\beta}(\mathbf{x}(0)) \approx 0$. Em iterações sucessivas, esse valor deve aumentar para atingir o equilíbrio; a experiência, então, visa determinar a primeira iteração, k^* , na qual o valor $\hat{\beta}(\mathbf{x}(k))$ atinge ou ultrapassa para acima o valor correto de β .

Todos os resultados apresentados no decorrer desta subseção foram obtidos sobre um suporte da forma $\{1, \dots, 128\} \times \{1, \dots, 128\}$. Para cada valor de β em $[0.70(0.005)1.20]$ foram simulados cem eventos independentes, e as quantidades k^* registradas. Também foram guardados t^* , os tempos de CPU em segundos, para cada simulação. A média e desvio padrão de k^* e t^* , para cada dinâmica e cada valor de β são mostrados na Tabela 3.1, reduzindo-se a informação aos valores $[0.70(0.05)1.20]$.

Nas Figuras 3.1 e 3.2 mostra-se, em diagrama semilogarítmico, o resultado de todas as medições de k^* , para os algoritmos de Swendsen-Wang e Wolff, respectivamente. Daí conclui-se que a dinâmica de Swendsen-Wang é mais rápida em atingir o equilíbrio sob as condições especificadas.

TABELA 3.1 - MÉDIAS E DESVIOS PADRÃO DOS NÚMEROS DE ITERAÇÕES E TEMPOS DE CPU ATÉ A CONVERGÊNCIA DOS ALGORITMOS DE SWENDSEN-WANG E WOLFF PARA DIFERENTES VALORES DO PARÂMETRO DE ATRATIVIDADE, COM CONFIGURAÇÃO INICIAL A TEMPERATURA INFINITA

β	Swendsen-Wang		Wolff	
	$\bar{k}^*(s_{k^*})$	$\bar{t}^*(s_{t^*})$	$\bar{k}^*(s_{k^*})$	$\bar{t}^*(s_{t^*})$
0.70	6.79(2.49)	0.49(0.21)	11447.21(4874.93)	377.72(299.84)
0.75	7.99(3.36)	0.54(0.20)	7960.50(1933.68)	185.58(46.60)
0.80	9.52(3.58)	0.67(0.25)	4921.27(1061.16)	230.16(49.94)
0.85	11.82(4.86)	0.91(0.45)	3015.48(435.82)	155.78(58.34)
0.90	16.12(6.19)	1.14(0.44)	1163.37(109.90)	65.11(13.82)
0.95	17.76(5.94)	1.26(0.43)	644.27(62.68)	35.17(2.92)
1.00	16.89(5.43)	1.41(0.63)	459.74(58.39)	26.33(2.66)
1.05	17.32(6.79)	1.38(0.53)	365.13(44.88)	34.13(5.91)
1.10	17.44(6.69)	1.54(0.58)	306.78(46.57)	28.08(3.60)
1.15	17.16(5.22)	1.38(0.55)	262.10(40.64)	24.62(2.80)
1.20	17.84(6.99)	1.49(0.61)	228.06(35.85)	26.24(4.00)

FONTE: (51), p. 487.

Já para a determinação de k_* , o número de iterações necessárias para atingir o equilíbrio a partir de $\mathbf{x}(0)$ simulada a $T = 0$, define-se:

$$k_* = \inf\{k \in \mathbb{N} : \hat{\beta}(\mathbf{x}(k)) \leq \beta\}.$$

Como o valor do parâmetro β ao qual foi obtida a ocorrência $\mathbf{x}(0)$ é infinito (pois $T = \beta^{-1} = 0$) espera-se, pelas boas propriedades do estimador de máxima pseudo-verossimilhança, que $\hat{\beta}(\mathbf{x}(0)) \approx \infty$. Em iterações sucessivas, esse valor deve diminuir para atingir o equilíbrio; a experiência, então, visa determinar a primeira iteração, k_* , na qual o valor $\hat{\beta}(\mathbf{x}(k))$ atinge ou ultrapassa para abaixo o valor correto de β .

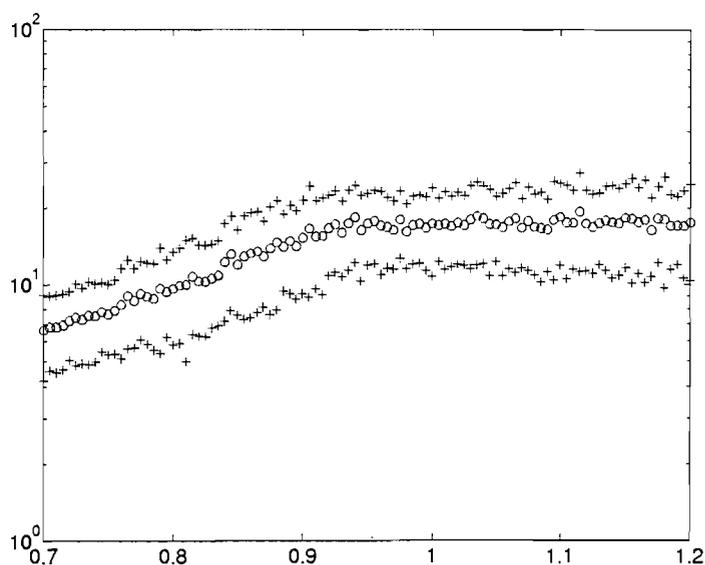


Fig. 3.1 - Número de iterações até a convergência do algoritmo de Swendsen-Wang para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura infinita (média amostral “o” e desvio padrão amostral “+”).

Para cada valor de β em $[0.70(0.005)1.20]$ foram simulados cem eventos independentes, e as quantidades k_* registradas. Também foram guardados t_* , os tempos de CPU em segundos, para cada simulação. A média e desvio padrão de k_* e t_* , para cada dinâmica e cada valor de β são mostrados na Tabela 3.2, reduzido-se a informação aos valores $[0.70(0.05)1.20]$.

Nas Figuras 3.3 e 3.4 mostra-se, em diagrama semilogarítmico, o resultado de todas as medições de k_* , para os algoritmos de Swendsen-Wang e Wolff, respectivamente. Daí conclui-se que, quando a configuração inicial é obtida da distribuição a $T = 0$, a dinâmica de Swendsen-Wang é mais rápida em atingir o equilíbrio para valores $\beta \leq 0.9 \approx \beta_c = \ln(1 + \sqrt{2})$, caso contrário a dinâmica de Wolff é a mais rápida.

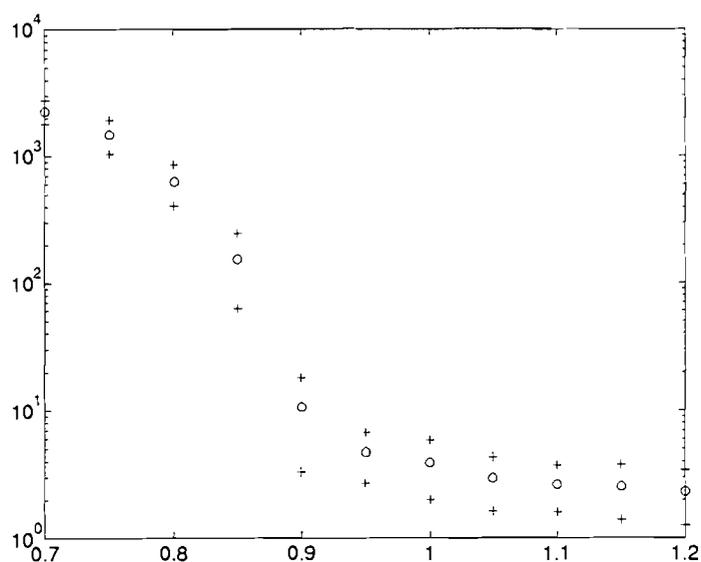


Fig. 3.2 - Número de iterações até a convergência do algoritmo de Wolff para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura infinita (média amostral “o” e desvio padrão amostral “+”).

As Tabelas 3.1 e 3.2 sugerem que a convergência ao equilíbrio é atingida mais rapidamente, por ambos algoritmos, quando a configuração inicial $x(0)$ é obtida a temperatura zero. O resultado é intuitivamente razoável, pois ambas dinâmicas precisam formar aglomerados, de tamanho dependente do parâmetro β (ver, por exemplo, as referências (38,107,121)). Essa dependência é ilustrada na Figura 3.5, obtida de medições feitas em k_* iterações com a dinâmica de Swendsen-Wang.

Se a configuração inicial possuir muitos aglomerados pequenos, qualquer um dos dois algoritmos deverá iterar várias vezes até conseguir aglomerados do tamanho adequado. Por outro lado, se a configuração inicial possuir poucos aglomerados grandes (como é o caso quando a mesma é gerada a $T = 0$), em umas poucas iterações qualquer um dos dois algoritmos construirá os aglomerados do tamanho adequado.

TABELA 3.2 - MÉDIAS E DESVIOS PADRÃO DOS NÚMEROS DE ITERAÇÕES E TEMPOS DE CPU ATÉ A CONVERGÊNCIA DOS ALGORITMOS DE SWENDSEN-WANG E WOLFF PARA DIFERENTES VALORES DO PARÂMETRO DE ATRATIVIDADE, COM CONFIGURAÇÃO INICIAL A TEMPERATURA NULA

—	Swendsen-Wang		Wolff	
	$\bar{k}_*(s_{k_*})$	$\bar{t}_*(s_{t_*})$	$\bar{k}_*(s_{k_*})$	$\bar{t}_*(s_{t_*})$
0.70	6.85(2.23)	0.67(0.47)	2272.63(479.19)	172.07(36.96)
0.75	6.99(2.44)	0.72(0.51)	1480.61(444.39)	108.25(32.28)
0.80	7.09(2.30)	0.68(0.47)	629.87(226.03)	46.55(16.34)
0.85	7.27(2.41)	0.70(0.54)	155.55(92.61)	12.82(7.04)
0.90	6.98(2.48)	0.70(0.52)	10.65(7.36)	2.27(1.09)
0.95	6.19(2.52)	0.66(0.53)	4.73(2.01)	1.57(0.73)
1.00	6.26(2.73)	0.66(0.55)	3.91(1.91)	1.41(0.61)
1.05	5.76(2.24)	0.64(0.50)	2.95(1.33)	1.14(0.53)
1.10	5.34(2.72)	0.57(0.52)	2.66(1.05)	1.11(0.51)
1.15	5.13(2.08)	0.55(0.50)	2.57(1.18)	1.04(0.62)
1.20	5.02(1.91)	0.56(0.54)	2.33(1.07)	0.97(0.59)

O equipamento empregado para estas experiências foi uma estação de trabalho SPARCstation 2 e, portanto, os tempos de CPU nas Tabelas 3.1 e 3.2 referem-se a essa máquina.

Empregando-se a informação apresentada sobre os tempos de convergência, foi feita uma experiência Monte Carlo para a determinação de algumas constantes típicas do modelo de Ising. Tais constantes são conhecidas, em forma exata, para o caso de suporte infinito; para o caso de suportes finitos com mais de 36 posições, é necessário lançar-se mão de ferramentas de simulação.

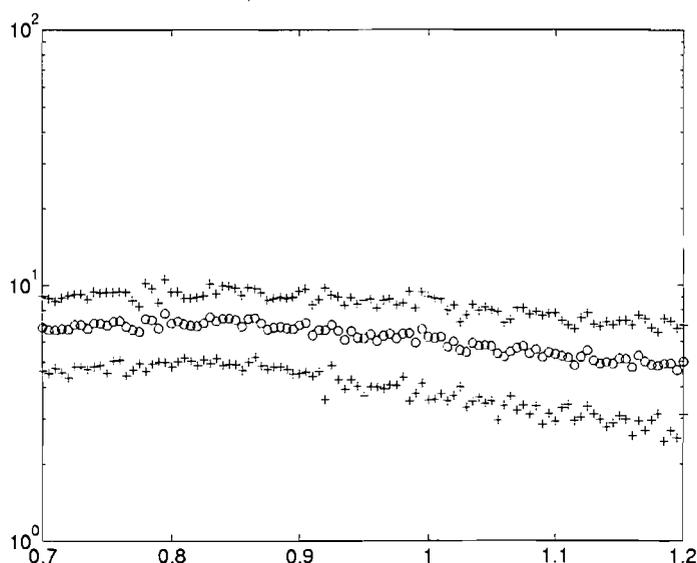


Fig. 3.3 - Número de iterações até a convergência do algoritmo de Swendsen-Wang para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura zero (média amostral “o” e desvio padrão amostral “+”).

Para tanto, e sempre considerando o suporte finito dado por $S = \{1, \dots, 128\} \times \{1, \dots, 128\}$, foram simuladas cem ocorrências independentes para cada valor do parâmetro β em $[0.70(0.005)1.20]$, e medidas as médias e desvios padrões do estimador de pseudoverossimilhança, da magnetização absoluta, da correlação à distância (euclidiana) “1” e da correlação à distância (euclidiana) “10”.

Cada ocorrência, independente das outras, é gerada com configuração inicial a $T = 0$ e cem iterações da dinâmica de Swendsen-Wang. Os resultados são mostrados nas Figuras 3.6, 3.7, 3.8 e 3.9.

Da Figura 3.6 pode concluir-se que a qualidade da estimação pelo método de pseudoverossimilhança é boa, pois não há evidência nem de vício nem de mudanças abruptas da variância do estimador. Também é claro que o intervalo de confiança aproximado, formado em torno de cada valor médio, é pequeno.

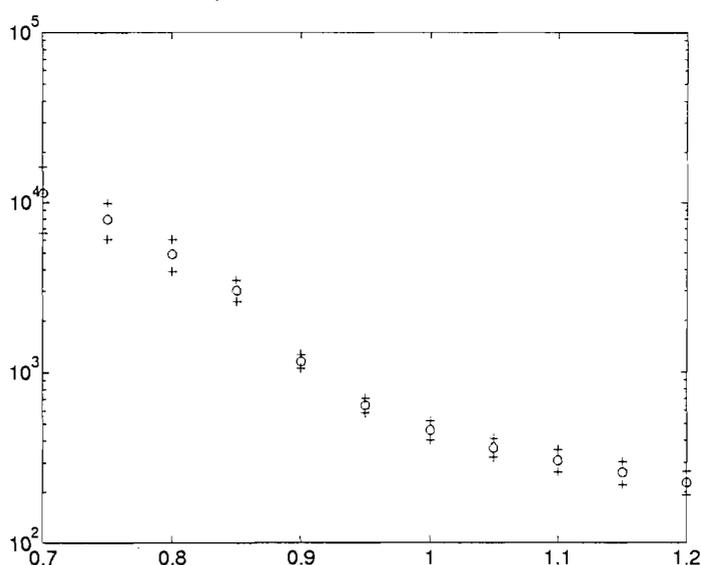


Fig. 3.4 - Número de iterações até a convergência do algoritmo de Wolff para diferentes valores do parâmetro de atratividade, com configuração inicial a temperatura zero (média amostral “o” e desvio padrão amostral “+”).

A Figura 3.7 mostra, superpostos, os resultados teóricos e estimados da magnetização média do modelo de Ising em função do parâmetro β . O valor teórico é dado pela seguinte equação (ver o trabalho (71)):

$$\langle M_\beta \rangle = \begin{cases} 0 & \text{se } \beta < \beta_c \\ \sqrt{\frac{\sqrt{(1 + \exp(-2\beta))\sqrt{1 - 6\exp(-2\beta) + \exp(-4\beta)}}}{1 - \exp(-2\beta)}} & \text{se } \beta \geq \beta_c, \end{cases} \quad (3.10)$$

Das Figuras 3.7, 3.8 e 3.9, comparadas com os resultados teóricos da magnetização média e correlações a curto e longo alcance do modelo de Ising com suporte infinito (71,118), pode concluir-se que as curvas construídas com experiências Monte Carlo sobre suportes finitos se ajustam bem aos resultados teóricos.

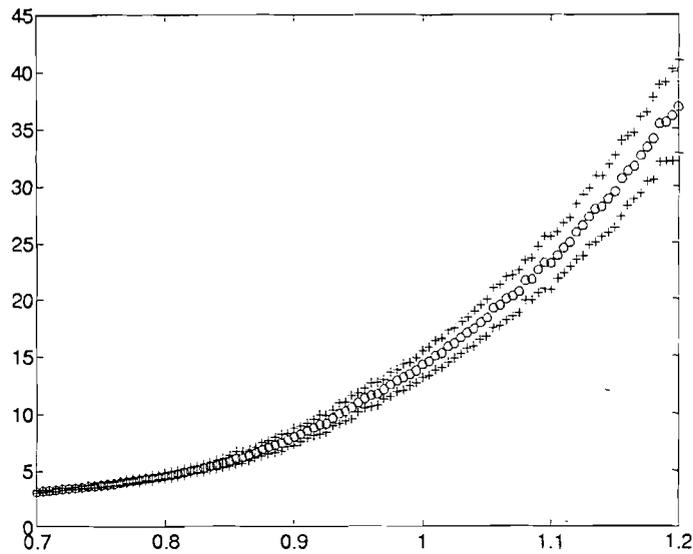


Fig. 3.5 - Tamanho médio dos aglomerados, medidos com simulações obtidas pela dinâmica de Swendsen-Wang (média amostral “o” e desvio padrão amostral “+”).

Pretende-se prosseguir, nesta linha, com outras experiências Monte Carlo para a estimação dos valores de k^* , k_* , t^* e t_* , para modelos de Potts-Strauss com mais de duas classes. Usando esta informação, poder-se-á construir gráficos análogos aos apresentados nas Figuras 3.6, 3.7, 3.8 e 3.9, para vários casos do modelo de Potts-Strauss.

3.5 Simulação de Imagens SAR

Em muitas aplicações é importante se dispor de imagens sintéticas. Um caso típico consiste na avaliação de técnicas de processamento e análise, em geral. Para tanto, se uma nova técnica for testada sobre uma imagem real, não será possível avaliar o seu desempenho comparando o resultado com a *realidade*, pois essa realidade não está disponível. Por outro lado, se fosse possível se dispor da realidade e da imagem observada, seria viável uma avaliação do desempenho dessa nova técnica pela comparação entre a primeira e o resultado obtido de aplicar a técnica sobre a imagem observada.

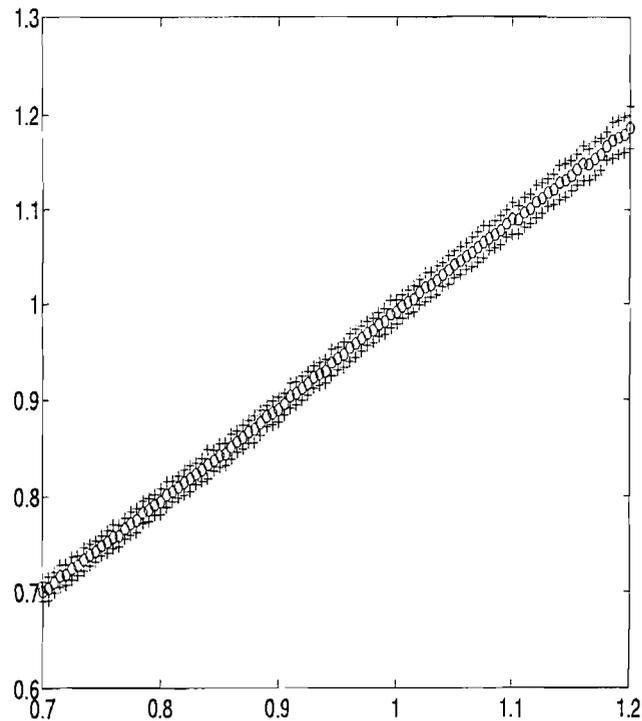


Fig. 3.6 - Valores estimados de média (“o”) e desvio padrão (“+”) do estimador de pseudoverossimilhança.

Para o problema de segmentação, portanto, seria de interesse a capacidade de se simular ocorrências da *verdade terrestre* (classes), a partir delas ocorrências das correspondentes *imagens observadas* e, sobre estas últimas, aplicar os algoritmos de segmentação. O mesmo procedimento poderia aplicar-se para o problema de restauração, desde que fossem simuladas ocorrências adequadas.

Utilizando a informação apresentada sobre os tempos de convergência dos algoritmos de Swendsen-Wang e Wolff, foi implementado um programa que gera ocorrências do modelo de Potts-Strauss, para um número arbitrário de classes e que substitui a ocorrência de cada classe pela ocorrência de uma variável aleatória cuja distribuição é a de uma média (de um ou mais termos) de variáveis aleatórias Rayleigh, com diferentes parâmetros para cada classe considerada.

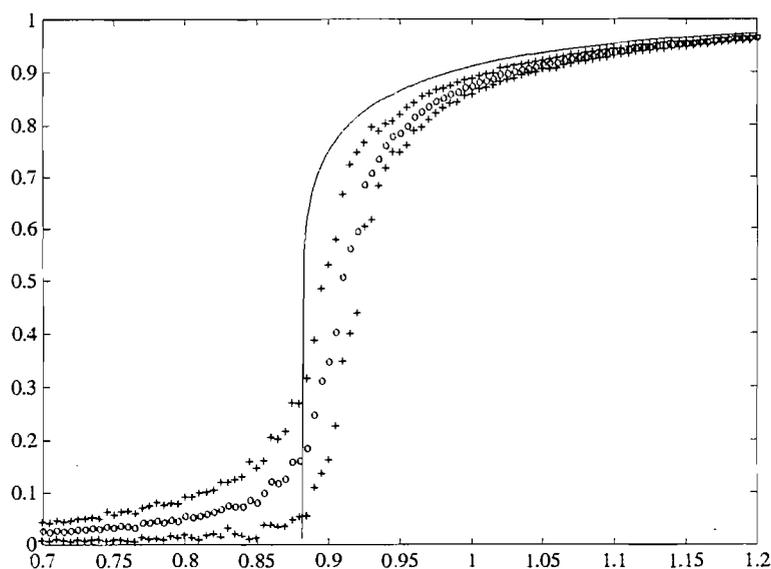


Fig. 3.7 - Valores teórico (linha contínua) e estimado (média: “o” e desvio padrão: “+”) da magnetização média absoluta do modelo de Ising.

A listagem deste programa encontra-se no Apêndice B. Resta, então, simular as ocorrências do modelo de degradação. Isto é, com os algoritmos de Swendsen-Wang e de Wolff pode gerar-se uma (ou mais) ocorrência de \mathbf{X} com distribuição de Potts-Strauss, por exemplo $\mathbf{X}(\omega) = \mathbf{x}$. Para se ter uma “imagem observada” simulada, faz-se necessária a geração de eventos provindos da distribuição Rayleigh, entre outras.

3.5.1 Observações Espacialmente Independentes

Uma maneira direta de se simular ocorrências de variáveis aleatórias com distribuição Rayleigh é o uso da relação apresentada na equação (5.4): se Z_1 e Z_2 são variáveis aleatórias independentes com distribuição $\mathcal{N}(0, 1)$, então $\xi\sqrt{Z_1^2 + Z_2^2} \sim \mathcal{R}(\xi)$. Mas, evidentemente, este método requer a simulação das variáveis Z_1 e Z_2 .

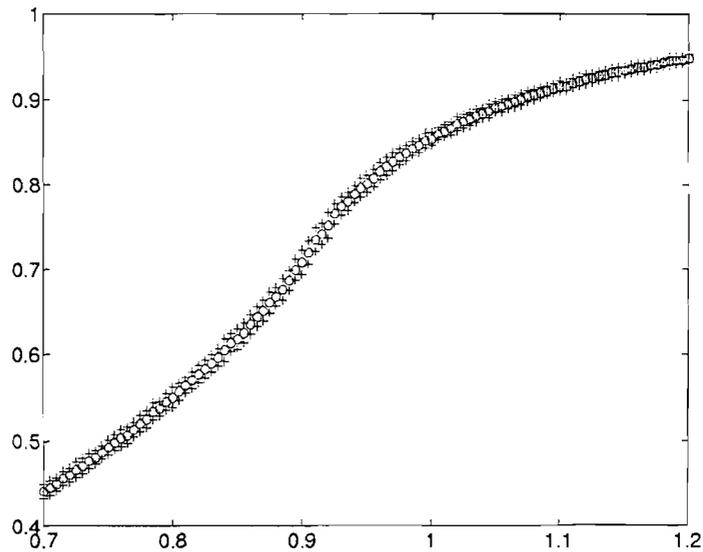


Fig. 3.8 - Valores da correlação a distância "1".

Por outro lado, é imediato que $Z_1^2 + Z_2^2 \sim \mathcal{E}(2)$, e também é claro que se $U \sim \mathcal{U}(0, 1)$, então $-\ln(U) \sim \mathcal{E}(1)$. Usando estes resultados, seria possível obter uma ocorrência de $Y \sim \mathcal{R}(\xi)$ pela relação $Y = \xi\sqrt{-2\ln(U)}$, com $U \sim \mathcal{U}(0, 1)$. Mas este resultado provém, diretamente, da técnica de simulação por inversão, que é comentada a seguir.

A técnica de inversão é de aplicabilidade geral (23,37) e, em muitos casos, é a mais *barata* computacionalmente entre as diversas técnicas de simulação disponíveis.

A técnica de geração por inversão é um método que funciona bem para geração de variáveis de tipo contínuo, embora possa servir, em princípio, para gerar variáveis quaisquer. Baseia-se no Teorema a seguir:

Teorema 3.5 Seja $F: \mathbb{R} \rightarrow [0, 1]$ uma função de distribuição acumulada; $F^{-1}(t): [0, 1] \rightarrow \mathbb{R}$ definida por $F^{-1}(t) = \inf\{x: t \leq F(x)\}$. Se $U \sim \mathcal{U}(0, 1)$ e $Y = F^{-1}(U)$, então F é a função de distribuição da variável aleatória Y .

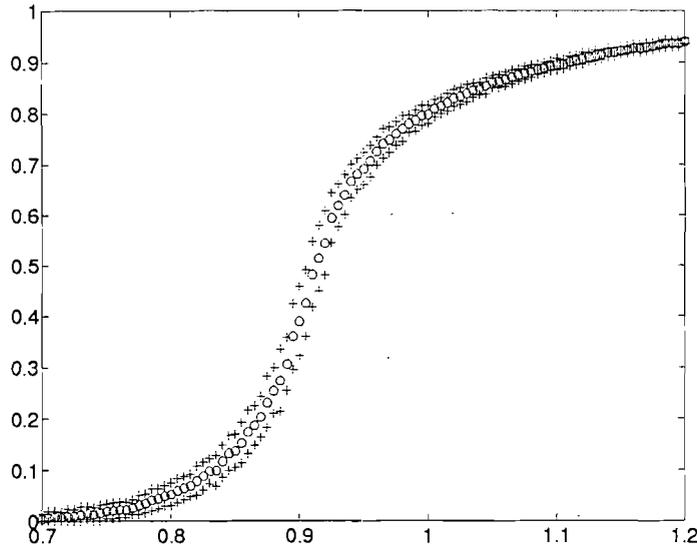


Fig. 3.9 - Valores da correlação a distância "10".

Usando a função de distribuição acumulada dada na equação (2.4), e com $U \sim \mathcal{U}(0,1)$, tem-se que $Y = \sqrt{-2\xi^2 \ln(1-U)} \sim \mathcal{R}(\xi)$. Mas, aproveitando o fato que se $U \sim \mathcal{U}(0,1)$ então $1-U \sim \mathcal{U}(0,1)$ (poupa-se, desta maneira, uma subtração por cada ocorrência desejada), tem-se que:

$$Y = \xi \sqrt{-2 \ln(U)} \sim \mathcal{R}(\xi).$$

Porém, é importante frisar que, tal como formulado acima, este método somente permite a geração de eventos provindos de variáveis aleatórias independentes.

Para se gerar imagens com eventos deste tipo foi implementado um programa, cuja listagem encontra-se no Apêndice B.1. Esse programa gera um evento do modelo de Potts-Strauss, com um número arbitrário de classes, que se constitui na ocorrência da distribuição *a priori*. Na Figura 3.10 tem-se uma ocorrência deste tipo, para três classes e $\beta = 0.99$.

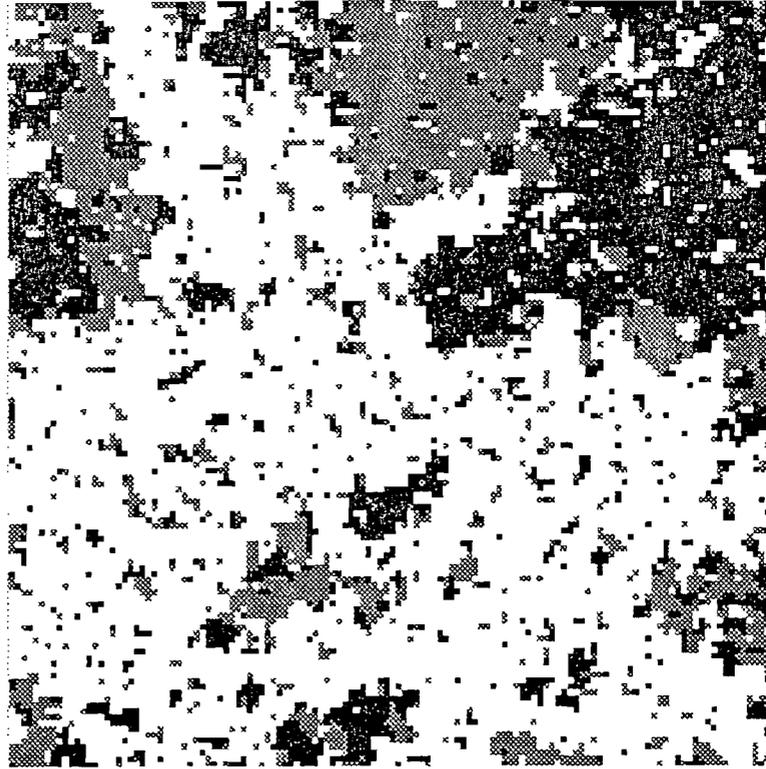


Fig. 3.10 - Ocorrência do modelo de Potts-Strauss para $\beta = 0.99$ e três classes.

Uma vez gerado o evento da distribuição *a priori*, por exemplo $\mathbf{x} = [x_s]_{s \in S}$, com $S = \{1, \dots, 128\} \times \{1, \dots, 128\}$ e $\Xi_s = \{24.68, 49.35, 82.26\}$, para se gerar o evento observado \mathbf{y} substitui-se cada x_s pela ocorrência da variável aleatória $Y = n^{-1} \sum_{i=1}^n Y_i$, com $Y_i \sim \mathcal{R}(\xi_s)$, e $\xi_s \in \Xi_s$. O valor de n pode ser qualquer inteiro. Desta maneira, e para $n = 1, 3, 6$, foram geradas as imagens mostradas nas Figuras 3.11, 3.12 e 3.13, respectivamente.

Esses valores de n foram escolhidos para se ilustrar o efeito do número de visadas na discriminabilidade de diferentes alvos em imagens SAR. Por outro lado, o sensor aerotransportado SAREX 580 fornece imagens de uma visada, as imagens ERS1 podem ser fornecidas com processamento de três visadas, e as imagens SAREX têm, aproximadamente, seis visadas.

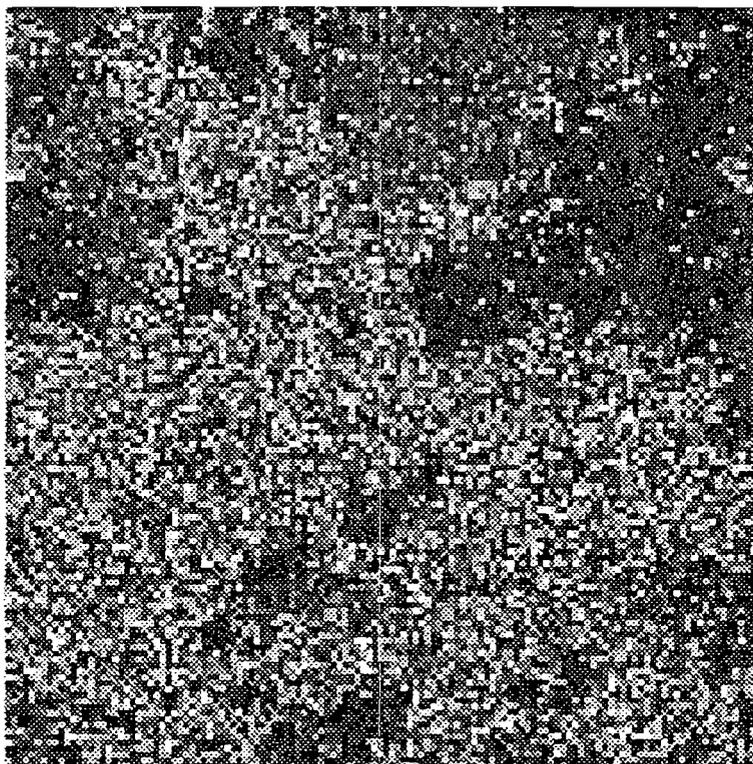


Fig. 3.11 - Simulação de uma imagem observada de uma visada.

3.5.2 Observações Espacialmente Dependentes

A independência espacial das observações pode não ser desejável para algumas aplicações, dado que muitas vezes a estrutura de correlação espacial dos dados fornece informação importante para a análise dos mesmos. Para certos tipos de dependência espacial, especificada através da função de correlação espacial, ver (139); para outros modelos de dependência espacial, ver (49).

Outro tipo de especificação de dependência estocástica, que não emprega a formulação através da função de correlação espacial, é apresentado em (36,80). Uma implementação para se simularem ocorrências deste modelo é apresentada no Apêndice B, Seção B.1.

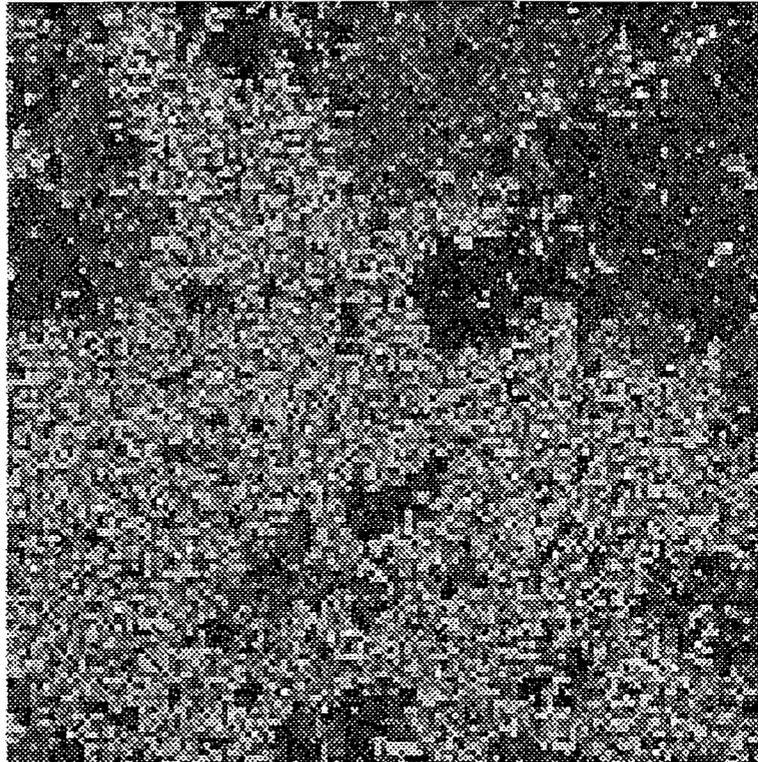


Fig. 3.12 - Simulação de uma imagem observada de três visadas.

Este modelo emprega variáveis aleatórias com contradomínio o campo dos complexos, que será denotado \mathbf{C} , isto é, $\mathbf{C} = \mathbb{R} \times j\mathbb{R}$. Seja $S = \{1, \dots, m\} \times \{1, \dots, n\}$ o suporte do processo e $\mathbf{Z}: \Omega \rightarrow \mathbf{C}$ o campo de amplitudes (complexas) que caracterizam um tipo de região.

Supõe-se que \mathbf{Z} possui as propriedades de ser:

- 1) circularmente simétrico,
- 2) conjuntamente Gaussiano, e
- 3) de média nula.

Por definição, a condição 1 quer dizer que, para todo $\theta \in (0, 2\pi]$ os campos \mathbf{Z} e $\mathbf{Z} \exp(j\theta)$ têm a mesma distribuição. Isto implica que são satisfeitas as seguintes condições:

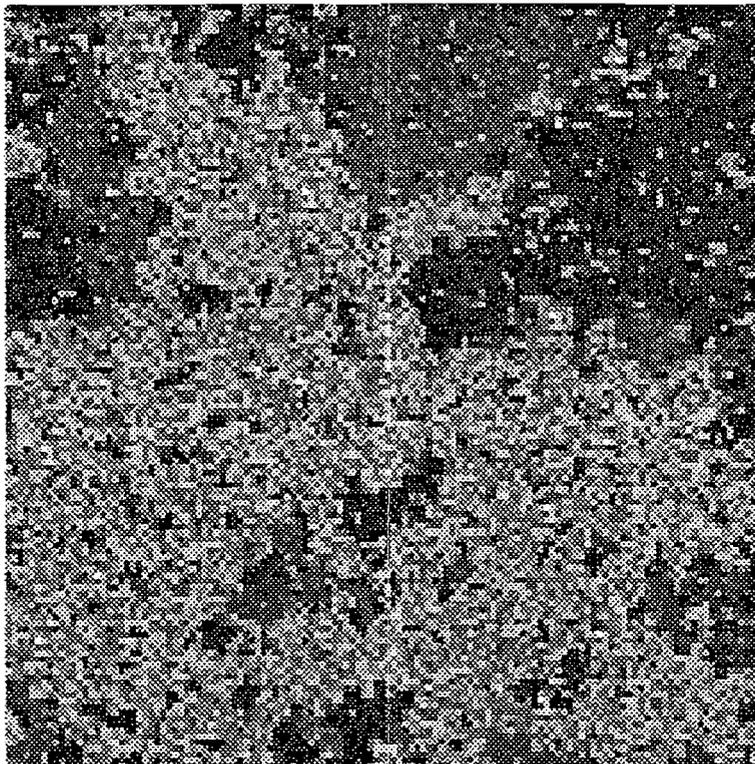


Fig. 3.13 - Simulação de uma imagem observada de seis visadas.

$$1) \mathbb{E}(\Re(Z_{i,j})\Re(Z_{k,\ell})) = \mathbb{E}(\Im(Z_{i,j})\Im(Z_{k,\ell})), \text{ e}$$

$$2) \mathbb{E}(\Re(Z_{i,j})\Im(Z_{k,\ell})) = -\mathbb{E}(\Im(Z_{i,j})\Re(Z_{k,\ell})),$$

para todo par de duplas de inteiros $((i, j), (k, \ell)) \in S \times S$, e, por sua vez, estas condições implicam que a distribuição conjunta de $\Re(\mathbf{Z})$ e $\Im(\mathbf{Z})$ pode ser escrita como:

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{\pi^N |R|} \exp(-\mathbf{z}^* R^{-1} \mathbf{z}) \quad \forall \mathbf{z} \in \mathbb{C},$$

onde R denota a matriz de covariância de \mathbf{Z} e \mathbf{z}^* denota o complexo conjugado de \mathbf{z} , para todo $\mathbf{z} \in \mathbb{C}$.

Para se especificar $f_{\mathbf{Z}}$ é necessário descrever R . Um modelo útil é o de autocorrelação separável. Para isto, supõe-se que vale:

$$\begin{aligned} \mathbb{E}(Z_{i,j} Z_{k,\ell}^*) &= 2\sigma^2 r_h(j, \ell) r_v(i, k), \text{ onde:} \\ r_h(j, \ell) &= r_h^*(\ell, j) = \rho_h^{(j-\ell)} \text{ para } j \geq \ell, \text{ e} \\ r_v(i, k) &= r_v^*(k, i) = \rho_v^{(i-k)} \text{ para } i \geq k, \end{aligned}$$

e $\rho_h, \rho_v \in \mathbb{C}$ são tais que $|\rho_h| < 1$ e $|\rho_v| < 1$.

Esta caracterização é equivalentemente definida pela autocorrelação especificada, para as coordenadas $\{(i, j) \in S: 1 < i < m, 1 < j < n\}$, por:

$$Z_{i,j} = \rho_h Z_{i,j-1} + \rho_v Z_{i-1,j} - \rho_h \rho_v Z_{i-1,j-1} + \varepsilon_{i,j}, \quad (3.11)$$

onde $\varepsilon_{i,j}: \Omega \rightarrow \mathbb{C}$ são variáveis aleatórias Gaussianas complexas independentes identicamente distribuídas com simetria circular, média nula e $\mathbb{E}(|\varepsilon_{i,j}|^2) = 2\sigma^2(1 - |\rho_h|^2)(1 - |\rho_v|^2)$. Daí pode provar-se que vale a seguinte relação:

$$f_{Z_s}(z \mid \mathbf{Z}_{S \setminus \{s\}} = \mathbf{z}_{S \setminus \{s\}}) = f_{Z_s}(z \mid \mathbf{Z}_{\partial_s^{(2)}} = \mathbf{z}_{\partial_s^{(2)}}), \quad (3.12)$$

onde $\partial_s^{(2)} = \{t \in S \setminus \{s\}: \|s - t\| \leq \sqrt{2}\}$.

A relação (3.11) é útil, além de caracterizar o processo, para se gerarem ocorrências de \mathbf{Z} por simulação. A relação (3.12) é importante pois permite caracterizar o processo \mathbf{Z} como um processo Markoviano com respeito à vizinhança de segunda ordem, denotada $\partial_s^{(2)}$. Esta última caracterização permitirá, junto com o Teorema 3.6 (página 90), afirmar algumas propriedades do processo *a posteriori*.

Se a imagem observada é de amplitude, na coordenada s observa-se a ocorrência de $|Z_s|$. Caso o processamento seja de intensidade, observa-se a ocorrência de $|Z_s|^2$. Note-se que, seja qual for o processamento, a informação de fase é perdida.

A Figura 3.14 mostra um evento simulado do processo proposto em (36,80). Nele tem-se que $S = \{1, \dots, 128\} \times \{1, \dots, 128\}$, $\rho_h = 0.29 + i0.31$, $\rho_v = 0.31 + i0.29$ e $\sigma = 1$, e processamento (dados) de amplitude.

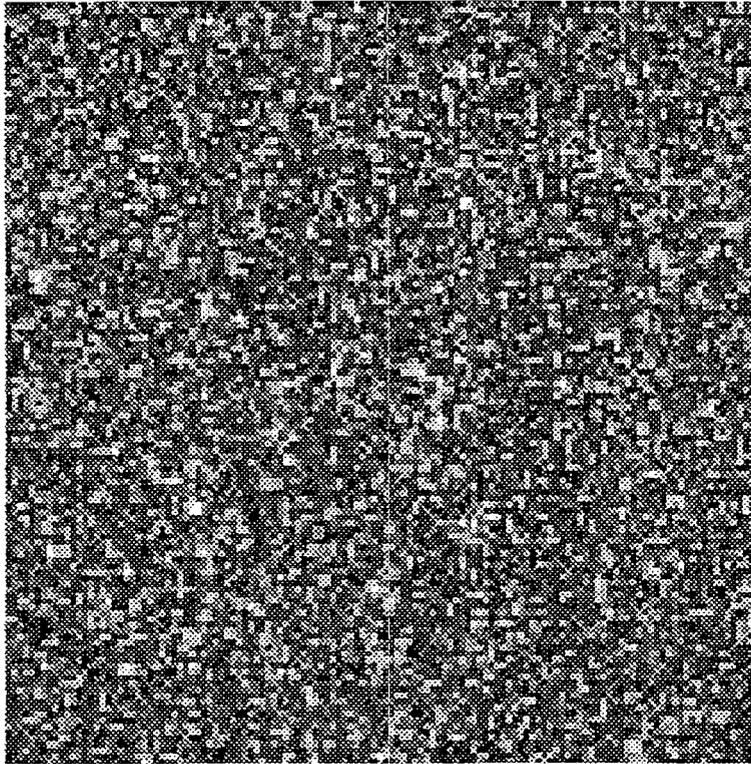


Fig. 3.14 - Simulação de um campo de *speckle* correlacionado.

3.6 Simulação de Imagens Ópticas

Uma das diferenças fundamentais entre as imagens ópticas e as de radar de abertura sintética é que, nas primeiras, pode-se supor que as observações obedecem, marginalmente, a uma distribuição normal. Ainda mais, para cada *pixel* as k bandas que formam uma imagem multiespectral podem modelar-se como ocorrências de uma distribuição Normal k -variada. Portanto, para se simular imagens multiespectrais é necessário a obtenção de ocorrências provindas da distribuição $\mathcal{N}k(\boldsymbol{\mu}, \Sigma)$, dada na equação (2.2).

Suponha-se disponível um método para gerar ocorrências de $Y \sim \mathcal{N}(0,1)$. A equação (2.2) fornece uma sugestão para se obter ocorrências de $\mathbf{X} \sim \mathcal{N}k(\boldsymbol{\mu}, \Sigma)$: basta gerar uma ocorrência do vetor $\mathbf{X} = (X_1, \dots, X_k)^T$ de variáveis independentes identicamente distribuídas como $\mathcal{N}(0,1)$ e definir (78):

$$\mathbf{Y}(\omega) = \mathbf{y} = \boldsymbol{\mu} + \mathbf{x}B,$$

com $BB^T = \Sigma$. Uma rotina baseada nesta ideia foi implementada na biblioteca de simulação `ranlib`. Para detalhes, ver o Apêndice A e as referências (2,15).

Na Figura 3.15 apresenta-se uma imagem simulada com características de imagem óptica. A distribuição *a priori* é o modelo de Potts-Strauss com $K = 2$ classes e $\beta = 0.99$. Para cada pixel x_s , observou-se a ocorrência de uma variável $\mathbf{Y}_s = (Y_s(1), Y_s(2), Y_s(3))$ aleatória com distribuição \mathcal{N}^3 e parâmetros diferentes para os dois valores possíveis de x_s . Os valores de $y_s(i)$, após truncamento e discretização, foram atribuídos aos níveis de vermelho, verde e azul, para $i = 1, 2, 3$ respectivamente, para formar a imagem colorida mostrada.

3.7 Uma Extensão do Teorema de Geman e Geman

No Capítulo 2 foram mencionados o Teorema de Geman e Geman (59) e a sua extensão (22). Esse teorema diz que, sob condições que serão apresentadas nesta Seção, a distribuição *a posteriori* de um campo Markoviano dada a sua versão degradada ainda é um campo Markoviano. Desta maneira, assegura-se a invariância destes modelos perante uma larga classe de operações.

A extensão apresentada neste trabalho generaliza o teorema original em duas direções:

- 1) substitui-se a hipótese de ruído branco normalmente distribuído pela hipótese de Markovianidade; desta maneira podem ser consideradas outras distribuições com correlação espacial;
- 2) substitui-se a hipótese do ruído ser aditivo ao sinal pela hipótese do ruído ser *misturado* ao sinal através de funções inversíveis; desta maneira, pode ser considerado ruído multiplicativo, entre outros.

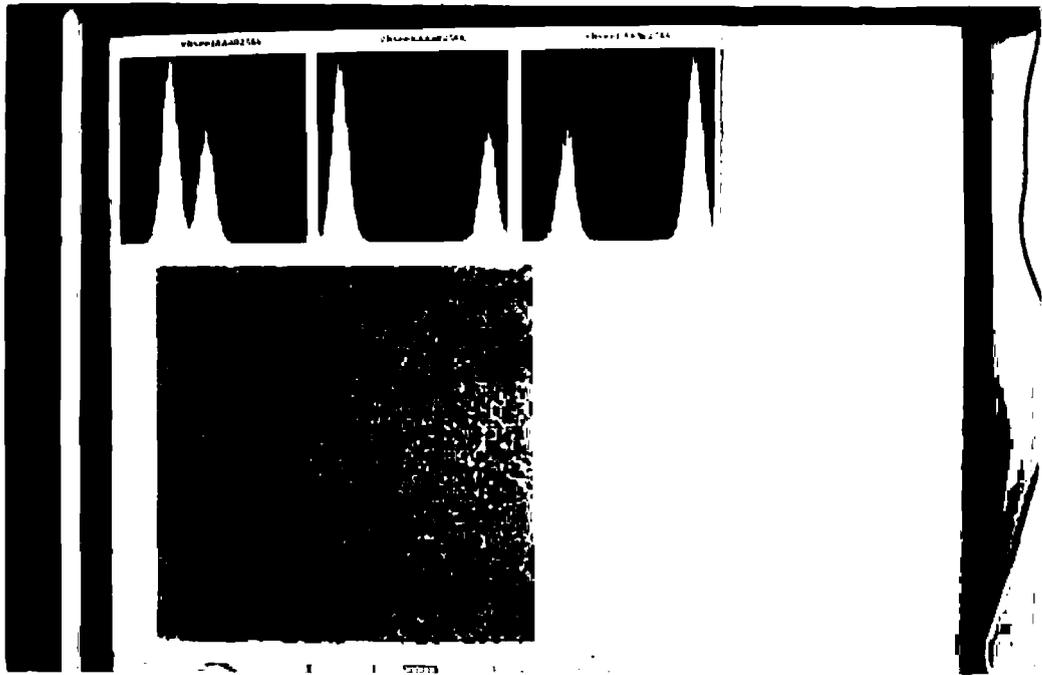


Fig. 3.15 - Simulação de uma imagem com duas classes, três bandas e distribuição $\mathcal{N}3$.

Utilizando os elementos definidos na Seção 3.2, e usando a mesma notação, esse Teorema será enunciado nesta Seção, e a sua prova será fornecida. Considera-se a possibilidade dos conjuntos Ξ_s serem diferentes para cada $s \in S$, isto é, o Teorema é válido para campos Markovianos em geral e não apenas para o modelo de Potss-Strauss.

Sejam os conjuntos S, Ξ_s com $s \in S$ e Ξ como já vistos. Sejam $\mathcal{V} = \{V_s: s \in S\}$ um sistema de vizinhanças de S , $\mathcal{G} = (S, \mathcal{V})$ um grafo, $(\Omega, \mathcal{A}, \mathbb{P}r)$ um espaço de probabilidade, e \mathbf{X} um \mathcal{G} -CAM.

Para cada $s \in S$ sejam τ_s e Γ_s conjuntos finitos que representam os possíveis valores das distorções e o ruído, respectivamente, que poderiam afetar uma imagem modelada pela distribuição de \mathbf{X} . Seja $\mathcal{B} = \{B_s: s \in S\}$ um sistema de vizinhanças de S e, para cada $s \in S$, seja $B_s^* = B_s \cup \{s\}$.

Considerem-se as seguintes funções:

$$\begin{aligned} H_s &: \Xi_{B_s^*} \longrightarrow \tau_s \text{ sobrejetora} \\ \phi_s &: \tau_s \longrightarrow \Gamma_s \text{ sobrejetora} \\ \odot_s &: \Gamma_s \times \Gamma_s \longrightarrow \Gamma_s \text{ sobrejetora e "inversível"}, \end{aligned}$$

isto é, existe uma função $\Phi_s: \Gamma_s \times \Gamma_s \longrightarrow \Gamma_s$ sobrejetora tal que:

$$\begin{aligned} \Phi_s(a, \odot_s(a, b)) &= b \\ \text{e } \odot_s(a, \Phi_s(a, b)) &= b \quad \forall (a, b) \in \Gamma_s \times \Gamma_s. \end{aligned}$$

Para cada $\mathbf{x} \in \Xi$ considera-se $\tilde{\mathbf{x}} \in \prod_{s \in S} \Xi_{B_s^*}$ dada por $\tilde{\mathbf{x}}_s = \mathbf{x}_{B_s^*}$. Definem-se as funções $\mathbf{H}: \prod_{s \in S} \Xi_{B_s^*} \longrightarrow \prod_{s \in S} \tau_s$ por $\mathbf{H}(\tilde{\mathbf{x}})(s) = H_s(\tilde{\mathbf{x}}_s)$. Também, definem-se as funções $\phi: \prod_{s \in S} \tau_s \longrightarrow \prod_{s \in S} \Gamma_s$ por $\phi(\tau)(s) = \phi_s(\mathbf{h}(s))$ para todo $\mathbf{h} \in \prod_{s \in S} \tau_s$. Definem-se as funções \odot sobre $\prod_{s \in S} (\Gamma_s \times \Gamma_s) = \prod_{s \in S} \Gamma_s \times \prod_{s \in S} \Gamma_s$ por $\odot(\mathbf{a}, \mathbf{b})(s) = \odot_s(\mathbf{a}(s), \mathbf{b}(s))$ para todo $(\mathbf{a}, \mathbf{b}) \in \prod_{s \in S} \Gamma_s \times \prod_{s \in S} \Gamma_s$.

Seja $\varepsilon: \Omega \longrightarrow \prod_{s \in S} \Gamma_s$ um campo Markoviano em relação ao grafo (S, \mathcal{B}) tal que \mathbf{X} e ε são independentes. Este campo representa o ruído.

Define-se o vetor aleatório representativo do campo realmente observado $\mathbf{Y}: \Omega \longrightarrow \Gamma = \prod_{s \in S} \Gamma_s$, por:

$$\mathbf{Y}(\omega) = \odot \left(\phi \left(H(\widetilde{\mathbf{X}}(\omega)) \right), \varepsilon(\omega) \right),$$

onde, para cada $\omega \in \Omega$, $\widetilde{\mathbf{X}}(\omega)$ é o elemento de $\prod_{s \in S} \Xi_{B_s^*}$ dado por:

$$\widetilde{\mathbf{X}}_s(\omega) = (\widetilde{\mathbf{X}}(\omega))_s = \mathbf{X}_{B_s^*}(\omega).$$

Proposição 3.6 Para cada $s \in S$ definem-se os conjuntos:

$$\begin{aligned} \mathcal{H}_s &= \{t \in S \setminus \{s\} : B_t \cap B_s \neq \emptyset\} \cup B_s \\ \mathcal{H}_s^p &= \left(\partial_s \cup \mathcal{H}_s \cup \bigcup_{r \in \mathcal{H}_s} \mathcal{H}_r \right) \setminus \{s\}, \end{aligned}$$

então $\{\mathcal{H}_s^p : s \in S\}$ é um sistema de vizinhanças em S .

Teorema 3.6 Considere $\mathbf{y} \in \Gamma$ tal que $\mathbb{P}\text{r}(\Omega^{(\mathbf{y})}) > 0$, onde $\Omega^{(\mathbf{y})} = \{\omega \in \Omega : \mathbf{Y}(\omega) = \mathbf{y}\}$. Seja $\mathbf{X}^{(\mathbf{y})}: \Omega^{(\mathbf{y})} \longrightarrow \Xi$ definido como $\mathbf{X}^{(\mathbf{y})}(\omega) = \mathbf{X}(\omega)$, então $\mathbb{P}\text{r}(\mathbf{X}^{(\mathbf{y})} = \mathbf{x}) = \mathbb{P}\text{r}(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y})$. Então $\mathbf{X}^{(\mathbf{y})}$ é um campo Markoviano em relação ao sistema de vizinhança $\{\mathcal{H}_s^p : s \in S\}$.

Para provar este resultado é suficiente provar que, para cada $s \in S$, a função $\mathbf{x} \mapsto \mathbb{P}\text{r}(X_s^{(\mathbf{y})} = x(s) \mid \mathbf{X}_{S \setminus \{s\}}^{(\mathbf{y})} = \mathbf{x}_{S \setminus \{s\}})$ depende somente dos valores x_s e $\mathbf{x}_{\mathcal{H}_s^p}$.

Sejam \mathcal{C} e \mathcal{D} os conjuntos:

$$\begin{aligned} \mathcal{C} &= \{C \in S : C \text{ é completo em relação a } \{\partial_s : s \in S\}\} \\ \mathcal{D} &= \{D \in S : D \text{ é completo em relação a } \{B_s : s \in S\}\}. \end{aligned}$$

Usando o Teorema 3.1 é possível ver que para cada $C \in \mathcal{C}$ e para cada $D \in \mathcal{D}$ existem funções $V_C: \Xi_C \longrightarrow \mathbb{R}$ e $W_D: \Gamma_D \longrightarrow \mathbb{R}$, e constantes $Z_{\mathbf{X}}$ e Z_{ε} tais que

$$\begin{aligned} \mathbb{P}\text{r}(\mathbf{X} = \mathbf{w}) &= Z_{\mathbf{X}}^{-1} \exp \left\{ - \sum_{C \in \mathcal{C}} V_C(\mathbf{w}_C) \right\} \quad \forall \mathbf{w} \in \Xi \\ \mathbb{P}\text{r}(\varepsilon = \boldsymbol{\eta}) &= Z_{\varepsilon}^{-1} \exp \left\{ - \sum_{D \in \mathcal{D}} W_D(\boldsymbol{\eta}_D) \right\} \quad \forall \boldsymbol{\eta} \in \Gamma. \end{aligned} \quad (3.13)$$

Considerem-se $\mathbf{x} \in \Xi$ e $s \in S$. Seja o elemento $\mathbf{e} \in \Gamma$ definido por:

$$\mathbf{e}_t = \Phi_t\left(\phi_t\left(H_t(\mathbf{x}_{B_t^*})\right), y_t\right) \quad \forall t \in S.$$

Para cada $z \in \Xi_s$, $\mathbf{x}[z]$ denotará o elemento de Ξ definido por

$$\mathbf{x}[z]_t = \begin{cases} z & \text{se } t = s \\ x_t & \text{se } t \neq s; \end{cases}$$

então, para cada $z \in \Xi_s$, $\mathbf{e}[z]$ denotará o elemento de Γ definido por:

$$\mathbf{e}[z]_t = \Phi_t\left(\phi_t\left(H_t(\mathbf{x}[z]_{B_t^*})\right), y_t\right) \quad \forall t \in S.$$

Suponha-se que os três seguintes Lemas valem:

Lema 3.1

$$\begin{aligned} \mathbb{P}_\Gamma(X_s^{(\mathbf{y})} = x_s \mid \mathbf{X}_{S \setminus \{s\}}^{(\mathbf{y})} = \mathbf{x}_{S \setminus \{s\}}) &= \\ &= \frac{\exp\left(-\sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C)\right) \exp\left(-\sum_{D \in \mathcal{D}} W_D(\mathbf{e}_D)\right)}{\sum_{z \in \Xi_s} \exp\left(-\sum_{C \in \mathcal{C}} V_C(\mathbf{x}[z]_C)\right) \cdot \exp\left(-\sum_{D \in \mathcal{D}} W_D(\mathbf{e}[z]_D)\right)}. \end{aligned}$$

Lema 3.2 Se

$$\begin{aligned} \mathcal{C}_1 &= \{C \in \mathcal{C} : s \notin C\}, \quad \mathcal{C}_2 = \mathcal{C} \setminus \mathcal{C}_1; \\ \mathcal{D}_1 &= \{D \in \mathcal{D} : D \cap B_s^* = \emptyset\}, \quad \mathcal{D}_2 = \mathcal{D} \setminus \mathcal{D}_1, \end{aligned}$$

então para cada $z \in \Xi_s$, vale que:

- 1) $\sum_{C \in \mathcal{C}_1} V_C(\mathbf{x}[z]_C) = \sum_{C \in \mathcal{C}_1} V_C(\mathbf{x}_C)$
- 2) $\sum_{D \in \mathcal{D}_1} W_D(\mathbf{e}[z]_D) = \sum_{D \in \mathcal{D}_1} W_D(\mathbf{e}_D)$
- 3) $\sum_{C \in \mathcal{C}_2} V_C(\mathbf{x}[z]_C) = f(z, \mathbf{x}_{\mathcal{H}_s^*})$
- 4) $\sum_{D \in \mathcal{D}_2} W_D(\mathbf{e}[z]_D) = g(z, \mathbf{x}_{\mathcal{H}_s^*})$,

onde as funções f e g definem-se no seguinte:

Lema 3.3 Sejam F e G funções reais definidas sobre Ξ por

$$F(\mathbf{z}) = \sum_{C \in \mathcal{C}_2} V_C(\mathbf{z}_C) \text{ e } G(\mathbf{z}) = \sum_{D \in \mathcal{D}_2} W_D(\boldsymbol{\eta}(\mathbf{z})_D),$$

onde $\boldsymbol{\eta}(\mathbf{z})$ é o elemento de Γ dado por:

$$\boldsymbol{\eta}(\mathbf{z})_t = \Phi_t\left(\phi_t\left(H_t(\mathbf{z}_{B_t^*})\right), y_t\right) \quad \forall t \in S.$$

Se \mathbf{z} e \mathbf{z}^* são elementos de Ξ tais que $\mathbf{z}_{\mathcal{H}_s^{p+}} = \mathbf{z}_{\mathcal{H}_s^{p+}}^*$, onde $\mathcal{H}_s^{p+} = \mathcal{H}_s^p \cup \{s\}$, então:

$$1) F(\mathbf{z}) = F(\mathbf{z}^*)$$

$$2) G(\mathbf{z}) = G(\mathbf{z}^*)$$

Do Lema 3.3 decorre que:

$$F(\mathbf{z}) = f(\mathbf{z}_{\mathcal{H}_s^{p+}}) \quad \forall \mathbf{z} \in \Xi$$

$$G(\mathbf{z}) = g(\mathbf{z}_{\mathcal{H}_s^{p+}}) \quad \forall \mathbf{z} \in \Xi, \text{ onde}$$

$$f(\mathbf{z}_{\mathcal{H}_s^{p+}}) = F(\mathbf{z}_{\mathcal{H}_s^{p+}}, \mathbf{z}_0) \text{ e}$$

$$g(\mathbf{z}_{\mathcal{H}_s^{p+}}) = G(\mathbf{z}_{\mathcal{H}_s^{p+}}, \mathbf{z}_0)$$

sendo \mathbf{z}_0 um elemento (fixado) arbitrário do conjunto $\Xi_{S \setminus \mathcal{H}_s^{p+}}$.

Finalmente, aplicando os três Lemas anteriores, o Teorema segue pois:

$$\begin{aligned} \mathbb{P}_r(X_s^{(y)} = x_s \mid \mathbf{X}_{S \setminus \{s\}}^{(y)} = \mathbf{x}_{S \setminus \{s\}}) &= \\ \frac{\exp\left(-\sum_{C \in \mathcal{C}_1} V_C(\mathbf{x}_C)\right) \exp(-f(x_s, \mathbf{x}_{\mathcal{H}_s^p})) \exp\left(-\sum_{D \in \mathcal{D}_1} W_D(\mathbf{e}_D)\right) \exp(-g(x_s, \mathbf{x}_{\mathcal{H}_s^p}))}{\exp\left(-\sum_{C \in \mathcal{C}_1} V_C(\mathbf{x}_C)\right) \exp\left(-\sum_{D \in \mathcal{D}_1} W_D(\mathbf{e}_D)\right) \sum_{z \in \Xi_s} \exp(-f(z, \mathbf{x}_{\mathcal{H}_s^p})) \exp(-g(z, \mathbf{x}_{\mathcal{H}_s^p}))} &= \\ = \frac{\exp(-f(x_s, \mathbf{x}_{\mathcal{H}_s^p}) - g(x_s, \mathbf{x}_{\mathcal{H}_s^p}))}{\sum_{z \in \Xi_s} \exp(-f(z, \mathbf{x}_{\mathcal{H}_s^p}) - g(z, \mathbf{x}_{\mathcal{H}_s^p}))}. \end{aligned}$$

Prova do Lema 3.1:

$$\begin{aligned} \mathbb{P}_r(X_s^{(y)} = x_s \mid \mathbf{X}_{S \setminus \{s\}}^{(y)} = \mathbf{x}_{S \setminus \{s\}}) &= \frac{\mathbb{P}_r(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{\mathbb{P}_r(\mathbf{Y} = \mathbf{y})} \\ &= \frac{\mathbb{P}_r(\mathbf{X} = \mathbf{x}, \boldsymbol{\varepsilon} = \mathbf{e})}{\mathbb{P}_r(\mathbf{Y} = \mathbf{y})} \\ &= \frac{\mathbb{P}_r(\mathbf{X} = \mathbf{x}, \boldsymbol{\varepsilon} = \mathbf{e})}{\sum_{z \in \Xi_s} \mathbb{P}_r(X_s = z, \mathbf{X}_{S \setminus \{s\}} = \mathbf{x}_{S \setminus \{s\}}, \boldsymbol{\varepsilon} = \mathbf{e}[z])} = \frac{\mathbb{P}_r(\mathbf{X} = \mathbf{x}, \boldsymbol{\varepsilon} = \mathbf{e})}{\sum_{z \in \Xi_s} \mathbb{P}_r(\mathbf{X} = \mathbf{x}[z], \boldsymbol{\varepsilon} = \mathbf{e}[z])}; \end{aligned}$$

agora, para todo $z \in \Xi_s$, e usando a independência entre \mathbf{X} e $\boldsymbol{\varepsilon}$, tem-se que

$$\mathbb{P}\text{r}(\mathbf{X} = \mathbf{x}[z], \boldsymbol{\varepsilon} = \mathbf{e}[z]) = \mathbb{P}\text{r}(\mathbf{X} = \mathbf{x}[z]) \cdot \mathbb{P}\text{r}(\boldsymbol{\varepsilon} = \mathbf{e}[z]).$$

Logo, este Lema é consequência direta da relação (3.13), provinda do Teorema 3.1.

Prova do Lema 3.2:

- 1) $C \in \mathcal{C}_1 \Rightarrow s \notin C \Rightarrow x[z]_C = x_C$ para todo $z \in \Xi_s$.
- 2) $D \in \mathcal{D}_1 \Rightarrow D \cap B_s^* = \emptyset, t \in D \Rightarrow t \neq s$ e $s \notin B_t$, dado que se $s \in B_t$ então $t \in B_s$, logo $D \cap B_s^* \neq \emptyset$. Portanto, $\mathbf{x}[z]_{B_t^*} = \mathbf{x}_{B_t^*}$ para todo $t \in D$ e para todo $z \in \Xi_s$, logo $e[z]_t = \Phi_t(\phi_t(H_t(\mathbf{x}[z]_{B_t^*})), y_t) = \Phi_t(\phi_t(H_t(\mathbf{x}_{B_t^*})), y_t) = e_t$ para todo $t \in D$ e para todo $z \in \Xi_s$.
- 3) É uma consequência imediata do Lema 3.3.
- 4) É uma consequência imediata do Lema 3.3.

Prova do Lema 3.3

- 1) Se $C \in \mathcal{C}_2$, então $s \in C$, logo $C \subset \partial_s \cup \{s\}$ dado que C é um clique para $\{\partial_t: t \in S\}$; portanto $C \subset \mathcal{H}_s^{p+}$. Assim, $F(\mathbf{z})$ depende somente das coordenadas de \mathbf{z} que pertencem a \mathcal{H}_s^{p+} .
- 2) Supondo que $D \in \mathcal{D}_2$; então $D \cap B_s^* \neq \emptyset$. Tem-se que provar que $t \in D \Rightarrow B_t^* \subset \mathcal{H}_s^{p+}$. Seja t_0 um elemento qualquer de $D \cap B_s^*$. Primeiro, considere-se o caso $t_0 = s$ e seja t em D , $t \neq t_0 \Rightarrow t \neq s$, mais t e s estão em D , e D é um clique de $\{B_\xi: \xi \in S\}$, logo $t \in B_s$. Se $u \in B_t$ então $t \in B_u$ logo $B_s \cap B_u \neq \emptyset$, portanto $u \in \mathcal{H}_s \cup \{s\}$; assim, $B_t^* \subset B_s \cup \mathcal{H}_s \cup \{s\} = \mathcal{H}_s \cup \{s\} \subset \mathcal{H}_s^{p+}$. Depois, considere-se o caso $t_0 \neq s$; dado que $t_0 \in B_s^*$, t_0 tem que pertencer a B_s ; o fato de $B_t^* \subset \mathcal{H}_s^{p+}$ prova-se da mesma maneira que o anterior. Agora, seja t um elemento de D tal que $t \neq t_0$ (e $t \neq s$), isto implica que $t_0 \in B_t$ dado que D é um clique de $\{B_\xi: \xi \in S\}$, logo $B_t \cap B_s \neq \emptyset$ e $t \in \mathcal{H}_s$. Se $u \in B_t$ e $u \neq s$, então $u \in \mathcal{H}_t$ e $u \in \bigcup_{r \in \mathcal{H}_s} \mathcal{H}_r \subset \mathcal{H}_s^{p+}$. Logo, $B_t^* \subset \mathcal{H}_s^{p+}$.

3.7.1 Nota sobre a Extensão

A versão original deste Teorema (59) parece mais geral do que a extensão aqui apresentada em um aspecto: a primeira é provada para quando o campo de ruído é ruído Gaussiano branco (uma coleção de variáveis aleatórias independentes identicamente distribuídas como $\mathcal{N}(\cdot, \cdot)$). Desta maneira, permitiria-se, para o ruído, \mathbb{R}^S como contradomínio sendo que a extensão exige que esse contradomínio seja de cardinalidade finita.

Porém, a prova original é feita para variáveis aleatórias discretas, tomando-se um truncamento e discretização de variáveis aleatórias normais. Assim sendo, o teorema original e a sua extensão possuem a mesma validade no que diz respeito ao contradomínio do ruído que, para ambos, deve ser no máximo enumerável.

3.7.2 Exemplo de Aplicação a Imagens SAR

Considere-se o modelo de Potts-Strauss com vizinhança de primeira ordem e borrimento de segunda ordem. Assim sendo, seja $S = \{1, \dots, 128\} \times \{1, \dots, 128\}$ o suporte para todos os campos e, para todo $s \in S$, seja $\Xi_s = \{6, 8, 10\}$ o contradomínio da *verdade terrestre*. Desta maneira, cada *pixel* poderá pertencer a uma das três classes de retroespalhamento: “escura”, “média” e “clara”. As vizinhanças de \mathbf{X} são as de primeira ordem: $\mathcal{V} = \{\partial_s : s \in S\} = \{t \in S : \|s - t\| = 1\}$, e o borrimento é definido sobre as janelas (que são vizinhanças, de segunda ordem) $\mathcal{H} = \{H_s : s \in S\} = \{t \in S \setminus \{s\} : \|s - t\| \leq \sqrt{2}\}$. Dado que os pixels das bordas e esquinas do suporte têm menos vizinhos do que os restantes, este modelo costuma ser chamado de *modelo com condições de fronteira livres*. Uma outra possibilidade seria definir vizinhanças com simetria toroidal, isto é, cada esquina tem por vizinhos as outras três esquinas, etc., mas este modelo pode ser pouco realista nas aplicações. A Figura 3.10 mostra um evento do modelo de Potts-Strauss, com vizinhança de primeira ordem e condições de fronteira livres, três classes e $\beta = 0.99$.

Por simplicidade notacional são definidos os conjuntos $C = \{c = (i, j) \in S : 1 < i < 128, 1 < j < 128\}$ (as posições que têm quatro vizinhos), $E = \{(1, 1), (1, 128), (128, 1), (128, 128)\}$ (as esquinas) e (as bordas) $L = S \setminus (C \cup E)$.

É imediato verificar que vale a relação:

$$\#B_s^* = \begin{cases} 4 & \text{se } s \in E \\ 6 & \text{se } s \in L \\ 9 & \text{se } s \in C. \end{cases}$$

Definindo os borramentos como $H_s(\tilde{\mathbf{x}}) = \frac{1}{\#B_s^*} \sum_{t \in B_s^*} x_t$ para todo $s \in S$, isto é, a média na janela considerada B_s^* , tem-se que os domínios do borramento estão dados por:

$$\begin{array}{ccc} \circ & \circ & \circ \\ \circ & \bullet & \circ \end{array} \quad \text{para posições } \bullet \in C;$$

$$\begin{array}{ccc} \circ & \circ & \circ \\ \bullet & \circ & \text{ou} & \circ & \bullet & \text{ou} & \circ & \bullet & \circ & \text{ou} & \circ & \circ & \circ \\ \circ & \circ & & \circ & \circ & \circ & & \circ & \bullet & \circ \end{array} \quad \text{para posições } \bullet \in L;$$

$$\begin{array}{ccc} \bullet & \circ & \text{ou} & \circ & \circ & \text{ou} & \circ & \bullet & \text{ou} & \circ & \circ \\ \circ & \circ & & \bullet & \circ & & \circ & \circ & & \circ & \bullet \end{array} \quad \text{para posições } \bullet \in E,$$

e, substituindo x_t por permutações dos valores em Ξ_s , pode-se calcular os conjuntos τ_s : os contradomínios da imagem borrada. Contudo, é importante frisar que a extensão do teorema antes apresentada permite considerar outros tipos de borramentos, mais gerais do que simples médias: poderia se pensar em médias ponderadas ou, mais geralmente, borramentos que dependam da coordenada.

Considerar-se-ão degradações não lineares, quadráticas, dadas, para todo $s \in S$, por $\phi_s(y) = y^2$ para todo $y \in \tau_s$. As misturas resultarão do produto

da imagem borrada e degradada vezes uma imagem de ruído. A imagem de ruído y é a que resulta de simular uma ocorrência do processo comentado na Seção 3.5.2, e é como a mostrada na Figura 3.14. O contradomínio da imagem de ruído é $\Gamma_s = \mathbb{R}$ para todo $s \in S$.

Para fins de apresentação dos resultados são efetuados dois truncamentos e uma discretização em 256 níveis. Se z é o valor matematicamente verdadeiro, observa-se o valor dado por $\min\{\max\{\lfloor z + \frac{1}{2} \rfloor\}, 0\}, 255\}$. Estes truncamentos e discretização, por não serem funções inversíveis, não estão previstos nem no teorema nem na extensão apresentada.

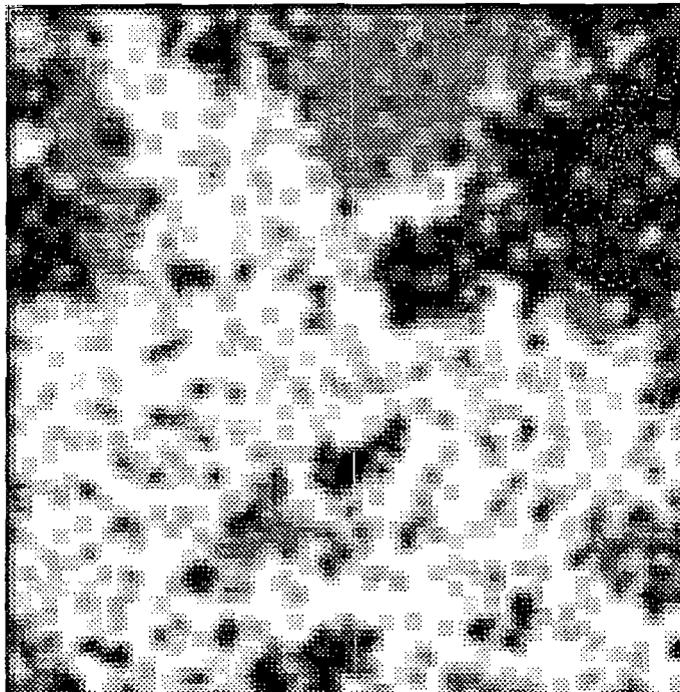


Fig. 3.16 - Ocorrência da distribuição *a priori* borrada em vizinhanças de segunda ordem.

Na Figura 3.16 é apresentada a versão borrada, em relação a vizinhanças de ordem dois, da ocorrência da distribuição *a priori* mostrada na Figura 3.10.

A Figura 3.17 ilustra o efeito que a transformação quadrática tem sobre a imagem mostrada na Figura 3.16. Note-se que as diferenças de tons são aumentadas.

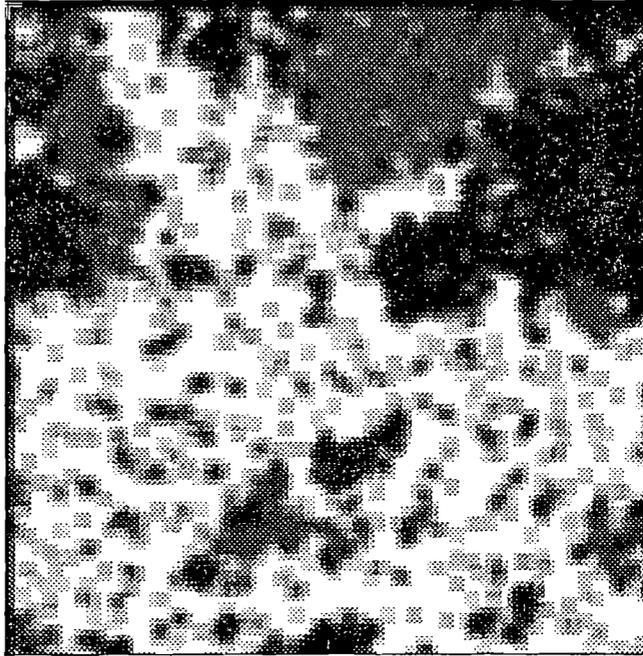


Fig. 3.17 - Ocorrência da distribuição *a priori* borrada em vizinhanças de segunda ordem e elevada ao quadrado.

A Figura 3.18 mostra uma ocorrência do processo observado (imagem degradada, os dados) comentado nesta Seção: resulta de multiplicar a imagem mostrada na Figura 3.17 pela imagem mostrada na Figura 3.14.

Sob as condições especificadas acima, a distribuição *a posteriori*, dada a observação dos dados, é Markoviana com respeito a uma nova vizinhança. As vizinhanças envolvidas são ilustradas na Figura 3.19.

Assim sendo, se \mathbf{X} (a distribuição *a priori*), é um campo Markoviano em relação à vizinhança de primeira ordem ($\partial_s = \{t \in S: \|s - t\| = 1\}$), é submetido a borramento em relação à vizinhança de segunda ordem ($\partial_s^{(2)} = \{t \in S \setminus \{s\}: \|s - t\| \leq \sqrt{2}\}$) e misturado a um campo, Markoviano em relação a $\partial_s^{(2)}$, então a distribuição de $\mathbf{X} \mid \mathbf{Y} = \mathbf{y}$ é Markoviana em relação à vizinhança $\partial^{(5)} = \{t \in S \setminus \{s\}: \text{máx}\{|s_1 - t_1| + |s_2 - t_2|\} \leq 5\}$.

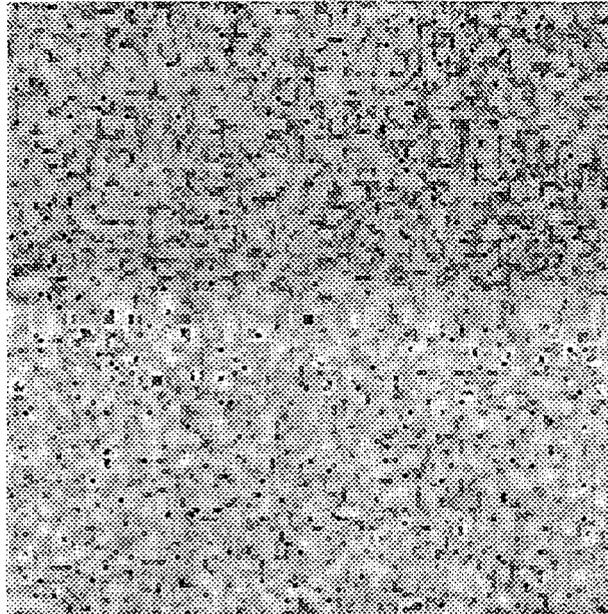


Fig. 4.8 - Imagem SAR-580, logaritmo da banda L.

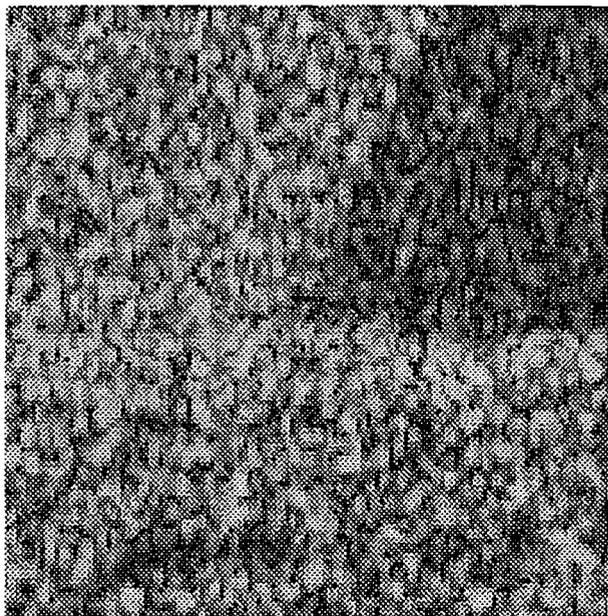


Fig. 4.9 - Imagem SAR-580, logaritmo da banda X.

CAPÍTULO 4

IMPLEMENTAÇÕES DO ALGORITMO ICM

4.1 Técnicas Markovianas de Segmentação

No Capítulo 2 foram introduzidas algumas técnicas para segmentação de imagens, sendo que a mais popular é a de segmentação por máxima verossimilhança pontual.

Mais modernamente, e sempre dentro do contexto estatístico, têm sido propostos alguns modelos que incorporam a noção de dependência espacial entre classes e/ou observações, dadas as classes (27,36,80). São os campos aleatórios Markovianos popularizados dentro da comunidade de processamento de imagens, entre outros, por Besag (9,10) e pelos irmãos Geman (58,59,60).

O uso destes modelos permite uma maior flexibilidade (a segmentação por máxima verossimilhança é um caso particular deles) ao custo, entretanto, de maior complexidade computacional. Este custo é justificável pela obtenção de melhores resultados, em geral. Outra vantagem da modelagem Markoviana é a *invariância* dos modelos envolvidos, tal como foi visto no Capítulo 3. Para aplicações específicas de técnicas Markovianas a problemas de processamento de imagens e visão computacional podem-se examinar os trabalhos (29,133), e as referências nelas indicadas. É importante frisar que as técnicas Markovianas não formam uma *ilha*, isto é, não estão isoladas de outros métodos; em (141) mostra-se como algumas operações da Morfologia Matemática se relacionam com técnicas baseadas na modelagem Markoviana.

Tal como foi dito no Capítulo 2, os principais estimadores relacionados às técnicas Markovianas são o ICM (*Iterated Conditional Modes*: modas condicionais iterativas), o MAP (*Maximum A Posteriori*: máximo *a posteriori*), o MPM (*Marginal Posterior Mode*: moda da marginal *a posteriori*) e o ABC (*Adaptive*

Bayesian Classification: classificação Bayesiana adaptativa).

A solução (estimador) MAP, tal como foi visto no Capítulo 2, é dada por $\hat{\mathbf{x}}_{\text{MAP}} = \text{máx}_{\mathbf{x}}^{-1} \mathbb{P}\text{r}(\mathbf{X} = \mathbf{x} \mid \mathbf{y})$, e a complexidade da obtenção depende do modelo empregado.

Quando o modelo para \mathbf{X} é o Markoviano (isto é, há interação entre posições vizinhas) não é possível, em geral, obter $\hat{\mathbf{x}}_{\text{MAP}}$ em forma direta pois, para tanto, deveria ser feita uma procura exaustiva no conjunto de todas as soluções possíveis, o que é computacionalmente inviável.

Uma forma de se obter a solução MAP, embora aproximadamente, é usando o algoritmo de *simulated annealing*. Ele se baseia numa analogia entre o problema em questão e técnicas de mecânica estatística para a localização de mínimos de funções com muitas variáveis.

O método de *simulated annealing* consiste em trabalhar com uma reparametrização de $\mathbb{P}\text{r}(\mathbf{x} \mid \mathbf{y})$ introduzindo T , um parâmetro não negativo chamado *temperatura*. Esta reparametrização possui as seguintes propriedades:

$$\mathbb{P}\text{r}_T(\mathbf{x} \mid \mathbf{y}) = \begin{cases} \pi(\mathbf{x}) & \text{se } T = 0 \\ \mathbb{P}\text{r}(\mathbf{x} \mid \mathbf{y}) & \text{se } T = 1 \\ \mu(\mathbf{x}) & \text{se } T = \infty, \end{cases} \quad (4.1)$$

onde $\mu(\mathbf{x})$ é a distribuição Uniforme sobre todas as soluções possíveis, e $\pi(\mathbf{x})$ é a distribuição Uniforme sobre todas as configurações \mathbf{x}^* tais que a probabilidade $\mathbb{P}\text{r}(\mathbf{x}^* \mid \mathbf{y})$ é máxima.

Uma vez estabelecida $\mathbb{P}\text{r}_T(\mathbf{x} \mid \mathbf{y})$, o método consiste em simular uma sequência de soluções (configurações) $\mathbf{x}(0), \mathbf{x}(1), \dots$, cada uma delas provindo dessa distribuição. Para que seja atingida uma configuração que maximize $\mathbb{P}\text{r}(\mathbf{x} \mid \mathbf{y})$ é necessário que tais ocorrências provenham da distribuição dada na equação (4.1) às temperaturas $\infty = T_0 \geq T_1 \geq \dots$, respectivamente. Sob condições bastante gerais (ver (25,59)) essa sequência de simulações convergirá a um dos eventos \mathbf{x}^* .

As principais desvantagens deste método são a dificuldade de implementação e o tempo que se leva para obter uma aproximação aceitável da solução desejada (ver Tabela 4.1).

TABELA 4.1 - PRINCIPAIS ESTIMADORES BAYESIANOS E AS SUAS CARACTERÍSTICAS

Estimador	Definição	Domina	Tempo	Par. Desc.
MAP Pontual	$[\max_{x_s}^{-1} \text{Pr}(x_s y_s)]_{s \in S}$	Pontual	1 segundo	0
MAP Global	$\max_{\mathbf{x}}^{-1} \text{Pr}(\mathbf{x} \mathbf{y})$	Global	10 horas	≥ 5
MPM	$[\max_{x_s}^{-1} \{\text{Pr}(x_s \mathbf{y})\}]_{s \in S}$	Local	10 minutos	≥ 2
ICM	Algoritmo Iterativo	Mistura	10 segundos	0

Define-se o estimador MPM como a configuração que, coordenada a coordenada, maximiza a distribuição marginal *a posteriori* (daí o nome *Marginal Posterior Mode*: moda —valor mais provável— da distribuição marginal *a posteriori*). Formalmente, este estimador define-se e denota-se como: $\hat{\mathbf{x}}_{\text{MPM}} = [\tilde{x}_s]_{s \in S}$, onde $\tilde{x}_s = \max_{x_s}^{-1} \{\text{Pr}(x_s | \mathbf{y})\}$.

Também, quando a modelagem de \mathbf{x} é Markoviana, não se conhecem técnicas eficientes que forneçam tal solução em forma direta. Para se obter uma solução aproximada, simula-se uma sequência de eventos $\mathbf{x}(0), \mathbf{x}(1), \dots$ da distribuição $\text{Pr}(\mathbf{x} | \mathbf{y})$ e retém-se, em cada coordenada, o valor mais frequentemente observado. Se o tamanho da amostra for suficientemente grande, a aproximação ao $\hat{\mathbf{x}}_{\text{MPM}}$ será boa.

Embora o tempo de execução de um programa para aproximar o estimador MPM seja bem menor do que aquele para aproximar o estimador MAP Global (ver comentários a esse respeito em (39) e a Tabela 4.1), o primeiro ainda apresenta dificuldades para a sua aplicação prática. Os tempos apresentados na Tabela 4.1 são meramente comparativos, e se referem a uma mesma máquina processando uma mesma imagem pelos algoritmos indicados. A terceira coluna indica qual é a característica que predomina na solução. A última coluna indica o número mínimo de parâmetros desconhecidos que devem ser ajustados (por tentativa e erro,

já que não se dispõe, em geral, de resultados a esse respeito) para se obter uma solução satisfatória.

O algoritmo ABC é um procedimento de segmentação contextual que, começando com uma segmentação desse tipo, a refina, modificando passo a passo a distribuição *a priori*. Este método condiciona o processo aos dados de treinamento, e modela as probabilidades *a priori* de tal maneira que dependem da localização espacial (81,82,83).

O uso específico de técnicas Markovianas em processamento de imagens SAR já aparece no trabalho de Besag (9) e, nos últimos anos, tem ganhado muito espaço na literatura (36,47,60,80,123,129).

4.2 O algoritmo ICM

O uso do algoritmo ICM para segmentação de imagens apresenta, como já visto, algumas vantagens sobre os estimadores MAP e MPM. Talvez a vantagem mais interessante, dada a tônica desta tese, seja o fato dele poder ser implementado de tal maneira que o usuário não precisa, para aplicá-lo, de mais conhecimentos do que os requeridos para utilizar o método de segmentação por máxima verossimilhança pontual, que é de uso corrente. Este algoritmo foi originalmente proposto como uma forma de se obter uma solução MAP aproximada e, mais tarde, ganhou o *status* de algoritmo com interesse próprio.

Tal como foi descrito, o algoritmo é do tipo *spin-flip*, mas pode ser implementado, também, em forma paralela; porém, a propriedade de assegurar uma seqüência de estimações com a característica descrita acima, só está provada para implementações sequenciais (veja o uso da imagem intermediária \mathbf{w} no algoritmo).

Para detalhes de implementação (a respeito da seqüência de *pixels* visitados $\{s, t, \dots\}$, da estimativa inicial $\hat{\mathbf{x}}(0)$ e dos possíveis critérios de parada), para a conexão do ICM com técnicas de relaxação estocástica e para resultados

experimentais ver as referências (9,10). O algoritmo HCF (*Highest Confidence First*) consiste, basicamente, do algoritmo ICM com uma especificação da sequência de *pixels* visitados baseada na mudança de energia que a troca produz (30,31,33). Ainda mais, existem trabalhos relacionando o ICM com técnicas de ondeletas (*wavelets*) e processadores piramidais (48).

Vários detalhes não estão completamente especificados no que diz respeito a este algoritmo; por exemplo, que parâmetros usar em cada iteração? Isto é, os parâmetros da distribuição *a priori* (por exemplo, θ) e da degradação (por exemplo, ξ) devem ser atualizados em cada iteração, ou não?

A seguir mostra-se uma implementação feita para trabalhar com seis opções diferentes:

- 1) Sem atualizar nem ξ nem θ .
- 2) Atualizando ξ mas não θ .
- 3) Sem atualizar ξ e atualizando θ .
- 4) Sem atualizar ξ e atualizando θ , mas fixando limitantes superiores para algum(ns) parâmetro(s).
- 5) Atualizando ambos ξ e θ .
- 6) Atualizando ambos ξ e θ , mas fixando limitantes superiores para algum(ns) parâmetro(s) do último.

A possibilidade de fixar limitantes superiores para algum(ns) parâmetro(s) de θ está relacionada ao aparecimento de *transição de fase*, assunto que foi comentado no Capítulo 3.

4.2.1 Implementação do ICM para Segmentação de Imagens SAR de uma Visada e Detecção Linear

Prosseguindo com o exemplo de imagens SAR monoespectrais (vide Exemplo 2.8), projeta-se o algoritmo ICM específico para este caso. Para tanto,

usar-se-á como distribuição *a priori* o modelo de Ising, isto é, o modelo de Potts-Strauss com $\mathbf{X}: \Omega \rightarrow \{-1, +1\}$ com suporte finito $S = \{1, \dots, 512\}^2$ e distribuição conjunta dada por $\mathbb{P}\text{r}(\mathbf{X} = \mathbf{x}) = Z_\beta^{-1} \exp\left(\beta \sum_{s,t: \|s-t\|=1} \mathbb{U}_{x_s}(x_t)\right)$.

Supondo que a *verdade terrestre* é uma coleção de posições, cada uma com um valor ξ_{-1} ou ξ_{+1} (com $0 < \xi_{-1} < \xi_{+1} < +\infty$) de “backscatter”, o sistema imageador retorna, para essa posição, a observação de uma variável Rayleigh com o parâmetro correspondente. É o *modelo multiplicativo* proposto para este tipo de imageamento (36,80).

Este modelo de degradação é um caso particular do proposto no trabalho (59), e modelar-se-á a imagem observada \mathbf{Y} segundo a seguinte distribuição conjunta:

$$f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \prod_{s \in S} f_{Y_s|x_s}(y_s) = \prod_{s \in S} \frac{y_s}{\xi_{x_s}} \exp\left(-\frac{1}{2}(y_s/\xi_{x_s})^2\right) \mathbb{U}_{(0,+\infty)}(y_s),$$

com $0 < \xi_{-1} < \xi_{+1} < +\infty$. O modelo empregado é o de variáveis aleatórias independentes, daí que a sua distribuição conjunta seja dada pelo produto das densidades.

As distribuições condicionais envolvidas no ICM, obtidas pelo uso do modelo de Potts-Strauss com $\Xi_s = \{-1, +1\}$, isto é, o modelo de Ising, são:

$$\begin{aligned} \mathbb{P}\text{r}(x_s|\mathbf{y}, \mathbf{x}_{S \setminus \{s\}}) &\propto f_{Y_s|x_s}(y_s) \mathbb{P}\text{r}(x_s|\mathbf{x}_{\partial_s}), \\ \text{com } \mathbb{P}\text{r}(x_s|\mathbf{x}_{\partial_s}) &= \frac{\exp\left(\frac{\beta}{2}(1+x_s)v(s)\right)}{1 + \exp(\beta v(s))} \text{ e } v(s) = \sum_{t \in \partial_s} x_t. \end{aligned}$$

Para se obter esta forma, partindo de $\mathbb{P}\text{r}(x_s|\mathbf{y}, \mathbf{x}_{S \setminus \{s\}})$, primeiro se usa a definição do modelo de Potts-Strauss; assim sendo, esta expressão pode ser simplificada a $\mathbb{P}\text{r}(x_s|\mathbf{y}, \mathbf{x}_{\partial_s})$. Feito isto, emprega-se o fato do ruído, dadas as classes, ser formado por variáveis independentes; logo, simplifica-se a probabilidade anterior obtendo $f_{Y_s|x_s}(y_s) \mathbb{P}\text{r}(x_s|\mathbf{x}_{\partial_s})$.

A configuração inicial $\hat{\mathbf{x}}(0)$ será, por escolha, a segmentação de máxima verossimilhança pontual da imagem \mathbf{y} (passo 4 do algoritmo).

De posse dessas equações pode-se escrever o seguinte algoritmo:

- 1) Atribuir $k = 0$, $I = N$ e $\ell = N$.
- 2) Ler $o \in \{0, 1, 2, 3\}$.
 - (a) Se $o = 0$ ou $o = 2$ então ler $\hat{\beta}(k)$; seguir em 3). Se não ler $\ell \in \{S, N\}$.
 - (b) Se $\ell = S$ ler β_M . Se não atribuir $\beta_M = +\infty$.
- 3) Ler $\hat{\xi}_{-1}(k)$ e $\hat{\xi}_{+1}(k)$.
- 4) Para todo $s \in S$ atribuir:

$$\hat{x}_s(k) = \begin{cases} -1 & \text{se } y_s \leq \sqrt{\frac{4 \ln \left(\frac{\hat{\xi}_{+1}(k)}{\hat{\xi}_{-1}(k)} \right)}{(\hat{\xi}_{-1}(k))^{-2} - (\hat{\xi}_{+1}(k))^{-2}}} \\ +1 & \text{caso contrário} \end{cases} \quad (4.2)$$

- 5) Fazer,
 - (a) Atribuir $k = k + 1$.
 - (b) Atualizar as estimações dos parâmetros das distribuições Rayleigh: $\hat{\xi}_c(k) = f_{c,o}(\hat{\mathbf{x}}(k-1))$ para $c = -1, +1$, e atualizar a estimacão do parâmetro da distribuição *a priori*: $\hat{\beta}(k) = f_{2,o,\ell}(\hat{\mathbf{x}}(k-1))$.
 - (c) Atribuir $\mathbf{w} = \hat{\mathbf{x}}(k-1)$.
 - (d) Para cada $p = 1, 2, 3, 4$ e para todo $s \in S_p$ fazer
 - i) Atribuir

$$w_s = \begin{cases} -1 & \text{se } y_s \leq \sqrt{\frac{4 \ln \left(\frac{\hat{\xi}_{+1}(k)}{\hat{\xi}_{-1}(k)} \right) - 2\hat{\beta}(k)v(s)}{(\hat{\xi}_{-1}(k))^{-2} - (\hat{\xi}_{+1}(k))^{-2}}} \\ +1 & \text{caso contrário.} \end{cases}$$

- ii) Se $w_s \neq \hat{x}_s(k-1)$ atribuir $I = S$.

- (e) Atribuir $\hat{\mathbf{x}}(k) = \mathbf{w}$

enquanto $I = S$.

- 6) Retornar $(\hat{\mathbf{x}}(k), k)$.
- 7) Fim.

No algoritmo acima emprega-se a notação que se descreve a seguir. O conjunto de suporte $S = \{1, \dots, 512\} \times \{1, \dots, 512\}$ foi particionado nos quatro subconjuntos descritos abaixo (o subíndice “ p ” acima é por *partição*):

$$S_1 = \{(i, j) \in S : i = 1(2)511, j = 1(2)511\},$$

$$S_2 = \{(i, j) \in S : i = 2(2)512, j = 2(2)512\},$$

$$S_3 = \{(i, j) \in S : i = 2(2)512, j = 1(2)511\},$$

$$S_4 = \{(i, j) \in S : i = 1(2)511, j = 2(2)512\},$$

e é claro que é uma partição, pois $\bigcup_{p=1}^4 S_p = S$ e $S_i \cap S_j = \emptyset$ sempre que $i \neq j$. Utiliza-se esta partição para evitar os efeitos indesejáveis que varreduras sistemáticas podem acarretar: o efeito diagonal citado por alguns autores. Este procedimento é ilustrado na Figura 4.1, onde cada número representa a ordem na qual a correspondente coordenada é visitada pelo algoritmo.

1	10	2	11	3	12
19	28	20	29	21	30
4	13	5	14	6	15
22	31	23	32	24	33
7	16	8	17	9	18
25	34	26	35	27	36

Fig. 4.1 - Seqüência de visitas ao suporte $S = \{1, \dots, 6\} \times \{1, \dots, 6\}$.

O produto $\hat{\beta}(k)v(s)$ contem a informação contextual, diferenciando assim o algoritmo ICM (desta implementação), de uma segmentação por máxima verossimilhança pontual. Este valor fará com que o *pixel* s seja atribuído à classe -1 ou à classe $+1$ em função, tanto do valor observado y_s , quanto das classes às quais foram atribuídos os seus vizinhos.

A estimação dos parâmetros da distribuição de Rayleigh foi feita usando o estimador de máxima verossimilhança já descrito na equação (2.8). Esta estimação é ativada se a variável o (por *opção*) valer 2 ou 3, segundo a fórmula

$$f_{c,o}(\hat{\mathbf{x}}(k)) = \begin{cases} \hat{\xi}_c(0) & \text{se } o = 0 \text{ ou } o = 1 \\ \sqrt{\frac{1}{2\#\{s \in S: \hat{x}_s(k)=c\}} \prod_{s \in S: \hat{x}_s(k)=c} y_s^2} & \text{caso contrário,} \end{cases}$$

com $c = -1, +1$.

Na Figura 4.2 mostra-se um exemplo de aplicação desta técnica a uma imagem SAR aerotransportado obtida pelo sensor SAR-580 (Banda L), sobre a região de Freiburg, Alemanha, com uma visada e detecção linear.

Na Figura 4.3 apresenta-se a segmentação ponto a ponto por máxima verossimilhança da imagem anterior (após uma estimação, também por máxima verossimilhança, dos parâmetros); é evidente o efeito de pontilhado causado pelo ruído *speckle*.



Fig. 4.2 - Imagem original: banda L, 512 *pixels* de lado.

Na Figura 4.4 tem-se o resultado de aplicar o algoritmo descrito acima, obtendo um resultado bem mais satisfatório, devido à utilização do contexto

espacial das classes, expresso pelo campo aleatório Markoviano.



Fig. 4.3 - Segmentação da imagem original em duas classes pelo algoritmo de máxima verossimilhança.



Fig. 4.4 - Segmentação da imagem original em duas classes pelo algoritmo ICM.

O parâmetro de atratividade, β , foi estimado utilizando-se a solução da equação de máxima pseudoverossimilhança (3.5).

Os resultados aqui apresentados foram satisfatórios o suficiente para motivar implementações desta técnica estendidas a outros casos. A primeira extensão foi, da mesma maneira que a anteriormente apresentada, uma implementação experimental concebida para distribuições Rayleigh e mais de duas classes.

Implementaram-se, então, os algoritmos de segmentação por máxima verossimilhança e ICM para distribuições Rayleigh e um número arbitrário de classes. Para a estimação do parâmetro de atratividade, β , nestas implementações foram empregadas as soluções das correspondentes equações de máxima pseudoverossimilhança: (3.6) ou (3.7). Não é oferecida a possibilidade de se reestimarem os parâmetros das observações por classe, nestas últimas: fato sugerido pelo uso do algoritmo para duas classes.

Aplicando-se estes algoritmos à imagem 4.2, porém, para três classes, obtiveram-se os resultados mostrados nas Figuras 4.5 e 4.6.



Fig. 4.5 - Segmentação da imagem original em três classes pelo algoritmo de máxima verossimilhança.



Fig. 4.6 - Segmentação da imagem original em três classes pelo algoritmo ICM.

Contudo, e apesar dos bons resultados obtidos, é importante ressaltar que estas implementações apresentam algumas limitações importantes:

- 1) Os algoritmos são específicos para a distribuição Rayleigh e uma única banda.
- 2) Por serem implementações experimentais, elas não possuem uma interface amigável com o usuário.

Dado que na Divisão de Processamento de Imagens do INPE, onde esta tese foi desenvolvida, está sendo desenvolvido um sistema de processamento de dados georeferenciados, o SPRING (26), resolveu-se aproveitar a estrutura modular do mesmo e o fato de já estar pronta a segmentação por máxima verossimilhança para incorporar o algoritmo ICM.

Foi avaliado o uso do algoritmo ICM para a filtragem de imagens SAR. Para tanto, foram considerados os casos $\Xi_s = \{0, \dots, \frac{256}{k} - 1\}$, com $k = 1, 2, 4$, para todo $s \in S$. As configurações iniciais foram obtidas de tal maneira que a observação y_s era atribuída, pelo algoritmo de máxima verossimilhança, à classe n se $y_s \in [nk, (n+1)k - 1]$, para cada $n = 0, 1, \dots, \frac{256}{k} - 1$. A conclusão é que, embora

seja possível o uso do algoritmo ICM para filtrar imagens, o custo computacional é muito alto (para $k = 1$, por exemplo, foram requeridas quatro horas para filtrar uma imagem de 512×512 *pixels* em uma SPARCstation 10) e o resultado obtido não justifica o tempo de processamento. Resultados análogos podem ser obtido pelo uso do filtro de Nagao-Matsuyama (113).

4.2.2 Implementações do ICM para Segmentação de Imagens Ópticas Multiespectrais no Ambiente Spring

O SPRING é um sistema que incorpora ferramentas de processamento de imagens, sistema de informações geográficas e banco de dados. Está sendo desenvolvido no INPE para estações de trabalho, sistema operacional UNIX, na linguagem de programação C++.

A abordagem utilizada para o desenvolvimento do sistema foi a de programação orientada a objetos. O sistema constitui-se basicamente de um conjunto de classes (Class em C++) que modelam objetos como, por exemplo, imagens, polígonos, linhas e texto. Além de objetos, processos também são modelados utilizando o conceito de classe. Por exemplo, o processo de geração de uma composição colorida a partir de três imagens digitais é modelado por uma classe.

O conceito de herança é utilizado para modelar relacionamentos entre classes, que são do tipo "É-UM". O módulo de classificação de imagens do SPRING é composto, basicamente, de uma hierarquia de classes. Na raiz da árvore encontra-se a classe *GenericClassifier*. Dela derivam-se duas classes: *SupervisedClassifier* e *Isodata*. A primeira destas duas classes se especializa nas classes *MaxverPixel* e *EuclidianDistancePixel*, que modelam os classificadores por máxima verossimilhança e por distância euclidiana, respectivamente.

A integração do classificador ICM foi feita pela criação da classe *MaxverPixelICM*, como especialização da classe *MaxverPixel*. Desta forma, a implementação dos ancestrais (dados e métodos) é reutilizada diretamente para definir esta nova classe. Como exemplo de reutilização, pode-se citar o método *Classi-*

`fySampleTest()`, pertencente à classe `SupervisedClassifier`, que classifica uma amostra de teste. Já os novos métodos, tais como o método `GetBeta()` que calcula o valor do parâmetro β em cada iteração e o método `ClassifyPixels()` que classifica um conjunto de dados, foram integrados diretamente a partir das implementações experimentais vistas na Seção 4.2.1. O esforço de modificação na interface de comunicação com o usuário foi mínimo pois a premissa para a integração do ICM ao SPRING foi a de não requerer, do usuário, maiores conhecimentos para o seu uso do que os requeridos pelo MAXVER, já implementado. Desta maneira, com o mesmo esforço por parte do usuário (e com um pouco mais de tempo de processamento), se obtêm segmentações menos ruidosas do que as obtidas pelo MAXVER.

A distribuição considerada por esta implementação é a normal k -variada (equação 2.2). Esta distribuição resulta apropriada para dados obtidos na faixa óptica do espectro. Contudo, a aplicabilidade desta implementação do algoritmo ICM para imagens SAR de uma visada e detecção linear fica garantida pelo uso de, entre outras, duas técnicas que *normalizam* as observações: transformações pontuais e filtragem de dados.

A técnica de normalização de dados de radar por filtragem é abordada, como consequência da avaliação dos filtros propostos no Capítulo 5.

A transformação pontual que será considerada para a normalização dos dados SAR é a logarítmica, descrita pela equação:

$$y' = a \left[\left(\ln(y + 1) + \frac{1}{2} \right) \right], \quad (4.3)$$

onde a constante a faz com que o valor máximo de tom de cinza após a transformação seja 255, isto é, $a = 255 \left(\left[\ln(255 + 1) + \frac{1}{2} \right] \right)^{-1}$. Esta transformação faz com que observações provindas da distribuição Rayleigh (um modelo razoável para as bandas originais) sejam mais parecidas com observações provindas de uma distribuição Gaussiana.

Seja $Y: \Omega \rightarrow \mathbb{R}_+$ uma variável aleatória com distribuição Rayleigh e parâmetro ξ . Seja a variável aleatória $W: \Omega \rightarrow \mathbb{R}$ dada por $W = \ln(Y)$. Então, aplicando o método do Jacobiano (76) e a propriedade de escala da dis-

tribuição Rayleigh, é imediato se obter a densidade de W dada, para todo $y \in \mathbb{R}$ por:

$$f_W(y) = \exp\left(2(y - \ln(\xi)) - \frac{1}{2} \exp(2(y - \ln(\xi)))\right).$$

Na Tabela 4.2 comparam-se os valores das esperança, variância, assimetria e kurtose de variáveis aleatórias Rayleigh padrão e o seu logaritmo. Estes valores, para o logaritmo da distribuição Rayleigh, foram obtidos numericamente e são, portanto, aproximados. Desta Tabela, pode concluir-se que a transformação logarítmica é efetiva na redução tanto da assimetria como da kurtose de variáveis aleatórias Rayleigh; portanto, esta transformação possui um efeito *normalizador* sobre as mesmas.

TABELA 4.2 - VALORES (APROXIMADOS) DA ESPERANÇA, VARIÂNCIA, ASSIMETRIA E KURTOSE DAS DISTRIBUIÇÕES RAYLEIGH PADRÃO E O SEU LOGARITMO

—	$X \sim \mathcal{R}(1)$	$\ln(X)$
IE	1.2533	0.0580
Var	0.4293	0.4112
γ_1	13.3670	3.2301
γ_2	40.4078	2.0370

Na Figura 4.7 observa-se o resultado de aplicar a transformação logarítmica a três densidades Rayleigh com diferentes parâmetros.

Nas Figuras 4.8 e 4.9 observa-se o resultado de aplicar a transformação (4.3) às bandas originais. O algoritmo MAXVER aplica-se a estas duas novas imagens combinadas, estimando-se os parâmetros relevantes para segmentar em três regiões distintas; o resultado deste algoritmo mostra-se na Figura 4.10.

Empregando a segmentação MAXVER como solução inicial, o resultado do algoritmo ICM é mostrado na Figura 4.11. O critério de parada para obtê-la foi: continuar iterando enquanto mudarem 0.5% (1311) ou mais do total de *pixels* de uma iteração para a seguinte. Os valores do estimador do parâmetro de atratividade e do número de *pixels* que mudam de iteração para iteração são

mostrados na Tabela 4.3.

TABELA 4.3 - VALORES DO ESTIMADOR DO PARÂMETRO DE ATRATIVIDADE E DO NÚMERO DE *PIXELS* TROCADOS POR ITERAÇÃO NO ALGORITMO ICM

Iteração	1	2	3	4
$\hat{\beta}$	1.09	1.61	1.91	2.09
# trocados	18761	5229	1851	745

O algoritmo ICM, usando a segmentação MAXVER como configuração inicial, precisou de quatro iterações para gerar uma solução. Esta solução (que incorpora contexto espacial), quando comparada com a obtida pelo algoritmo de máxima verossimilhança, mostra-se bastante mais *suave* facilitando, assim, a sua aplicação à interpretação visual da área sob estudo.

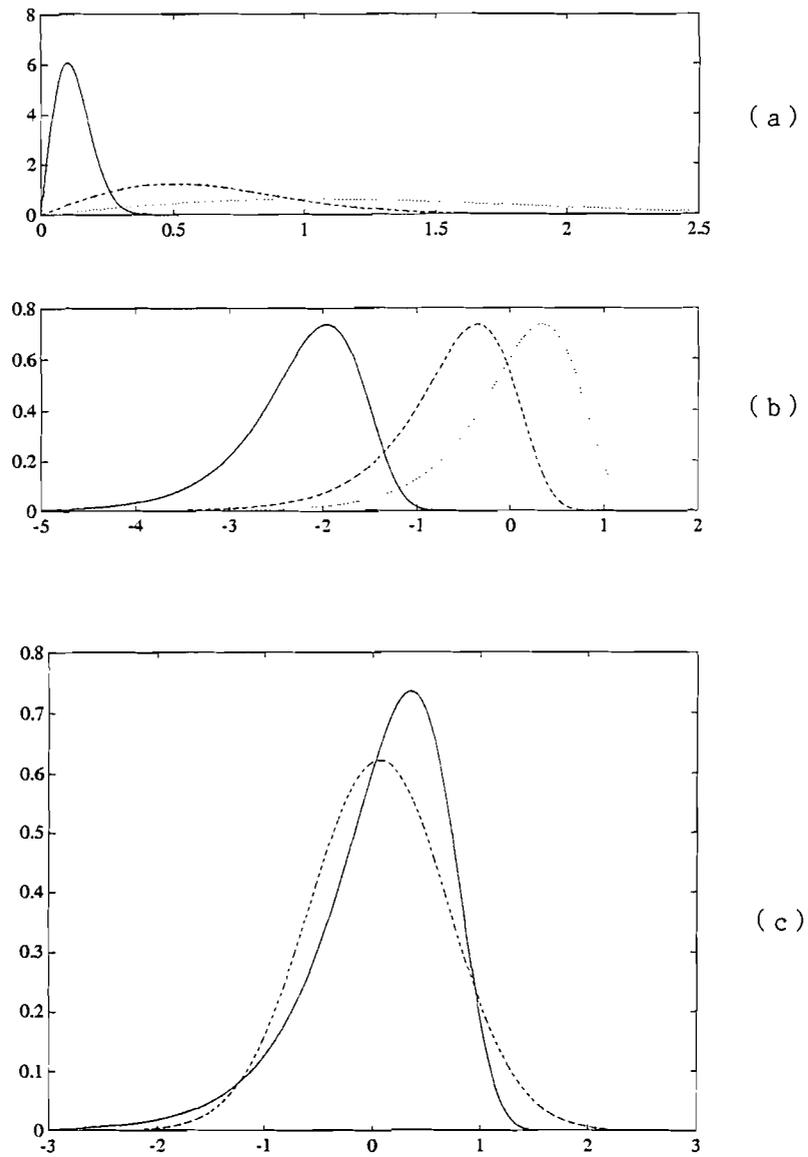


Fig. 4.7 - Transformação logarítmica da distribuição Rayleigh: (a) três densidades Rayleigh; (b) as densidades dos respectivos logaritmos; (c) densidade do logaritmo e densidade da normal com mesmas média e variância superpostas.

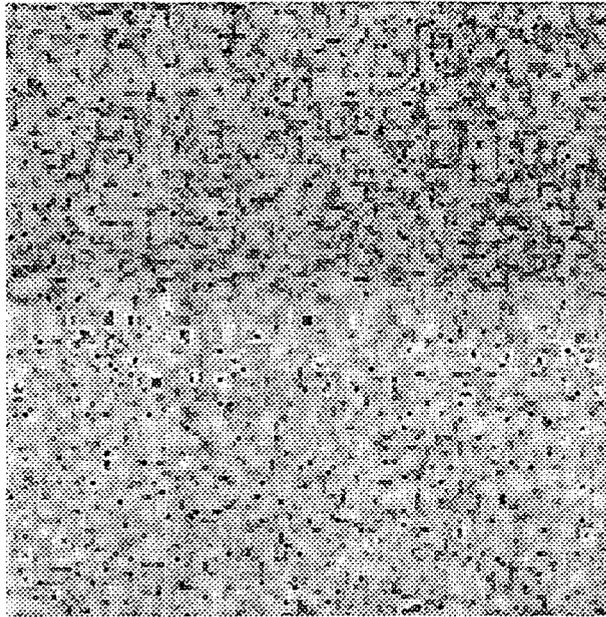


Fig. 4.8 - Imagem SAR-580, logaritmo da banda L.



Fig. 4.9 - Imagem SAR-580, logaritmo da banda X.



Fig. 4.10 - Segmentação por máxima verossimilhança dos logaritmos das bandas L e X.

FONTE: (46), p. 36.

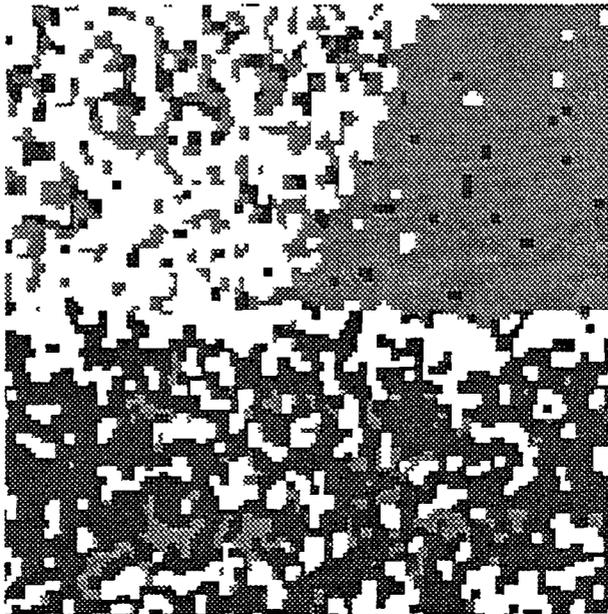


Fig. 4.11 - Segmentação pelo algoritmo ICM dos logaritmos das bandas L e X.

FONTE: (46), p. 36.

CAPÍTULO 5

A DISTRIBUIÇÃO RAYLEIGH E FILTROS ROBUSTOS
PARA DETEÇÃO LINEAR E UMA VISADA

5.1 Introdução

No decorrer deste trabalho observou-se que tanto para a restauração, a análise e a simulação de dados SAR faz-se necessária a suposição de algumas distribuições para os dados observados. Neste Capítulo discutem-se alguns modelos para os dados de radar de abertura sintética (SAR). Apresentam-se também alguns aspectos estatísticos acerca desses dados, dando ênfase ao comportamento distribucional dos mesmos (43,139).

O objetivo deste Capítulo é apresentar alguns aspectos relativos a essas propriedades, enfatizando modelos para alvos homogêneos extensos, característicos de muitas áreas florestais.

Sejam as funções reais g_1 e g_2 definidas sobre $S = \{0, \dots, n_1 - 1\} \times \{0, \dots, n_2 - 1\}$, isto é, $g_1, g_2: S \rightarrow \mathbb{R}$. O produto de convolução (circular discreta (7,75,97)) de g_1 por g_2 é a nova função $g: S \rightarrow \mathbb{R}$ dada por:

$$g(u) = (g_1 * g_2)(u) = \sum_{s \in S} g_1(s)g_2(u - s) \forall u \in S,$$

onde as operações são definidas “módulo(n_1, n_2)”.

O modelo admitido aqui para o SAR é uma versão simplificada daquele proposto em (16), pois não leva em consideração o ruído intruduzido pelo sistema radar. O modelo é ilustrado na Figura 5.1, onde:

- $s = (x, y)$ é um vetor bidimensional, x representa azimute (direção do vôo) e y o *range* (direção perpendicular ao vôo);



Fig. 5.1 - Modelo para o SAR.

- $n(s)$ denota a reflectividade complexa na posição s (um número complexo que fornece as mudanças na amplitude e fase causadas pela interação das ondas transmitidas com a superfície);
- $k(s)$ e $l(s)$ são as respostas à função impulso de dois filtros, sendo que $k(s)$ leva em consideração o padrão da antena e deslocamento da fase, e $l(s)$ é a resposta da função impulso do filtro utilizado no processamento dos dados; e
- $z(s)$ é o sinal de saída.

Portanto,

$$z(s) = n(s) * k(s) * l(s) = n(s) * h(s) \quad (5.1)$$

onde “*” denota convolução e $h(s) = k(s) * l(s)$. A função $h(s)$ é denominada função de espalhamento pontual do sistema, visto que descreve a resposta do sistema a um único espalhador. Implícita na equação (5.1) está a diferença fundamental entre o radar e os sensores ópticos, uma vez que $z(s)$, $n(s)$ e $h(s)$ podem ser todos números complexos. Isto implica que a reflectividade complexa do radar inclui alterações em amplitude e fase da onda incidente. $z(s)$ é denominada imagem complexa e é na verdade um conjunto de duas imagens, $a(s)$ e $b(s)$, cada qual representando a projeção da amplitude em dois planos perpendiculares: $a(s) = A(s) \cos \phi(s)$ e $b(s) = A(s) \sin \phi(s)$, onde $A(s)$ e $\phi(s)$ são respectivamente a amplitude e a fase do sinal de retorno na posição s .

Para se formar uma imagem do terreno, forma-se a imagem de amplitude $A(s) = |z(s)|$, ou a imagem de intensidade $I(s) = |z(s)|^2 = A^2(s)$. Ressalta-se que tanto na imagem de amplitude quanto na de intensidade, a informação de fase é perdida.

Obviamente, as propriedades estatísticas apresentadas por $z(s)$, $A(s)$ e $I(s)$ dependem do modelo adotado para a reflectividade complexa $n(s)$. O modelo mais comum é o multiplicativo (105), isto é:

$$n(s) = \nu(s)m(s), \quad (5.2)$$

onde $\nu(s)$ é a amplitude do sinal de retorno (relacionada à seção transversal do radar $\sigma(s)$ por $\nu(s) = \sqrt{\sigma(s)}$) e $m(s)$ é um ruído multiplicativo, denominado *speckle*. Em geral, admite-se que as partes real e imaginária do *speckle* têm distribuição Gaussiana, como consequência natural do teorema central do limite, uma vez que o *speckle* resulta do fato da célula de resolução possuir muitos espalhadores.

Para regiões homogêneas extensas, isto é, regiões onde a média da amplitude do sinal de retorno, ou a média da seção transversal do radar é constante para todo *pixel*, sob certas suposições, $n(s)$ é um ruído branco com distribuição Gaussiana. Embora este seja um modelo bem simples para a reflectividade complexa, ele é considerado válido para muitos tipos de alvos, como por exemplo campos agrícolas (120). Com essas suposições é possível provar-se que as partes real ($a(s)$) e imaginária ($b(s)$) da imagem complexa $z(s)$ são normalmente distribuídas com média zero e mesmo desvio padrão. Com as suposições acima, as imagens de amplitude e de intensidade têm funções densidade de probabilidade Rayleigh (distribuição 2.8) e Exponencial (distribuição 2.9), respectivamente. É possível também provar-se que têm a mesma função de autocorrelação espacial, e que esta depende somente da função de espalhamento pontual $h(s)$ do sistema do radar (ver, por exemplo (139)). Se $h(s)$ for uma função par (o que é geralmente o caso para sistemas SAR), então a correlação cruzada entre $a(s)$ e $b(s)$ é igual a zero para qualquer s .

As funções densidade de probabilidade expressas nas equações (2.5) e (2.9) são mostradas na Figuras 5.2. Os parâmetros das densidades Rayleigh

são $\xi = 1$, $\xi = 0.5$ e $\xi = 0.1$ (linhas de pontos, tracejada e contínua, respectivamente. Os parâmetros das densidades Exponencial são $\theta = 1$, $\theta = 0.7$ e $\theta = 0.32$ (linhas de pontos, contínua e tracejada, respectivamente. Vê-se, pelo gráfico da Exponencial, que qualquer que seja a média da intensidade, o valor mais provável é o zero. Devido à longa cauda da distribuição, a probabilidade de ocorrência de grandes valores é relativamente alta (com respeito, por exemplo, à distribuição Normal), e tal probabilidade também aumenta com o parâmetro θ . As implicações disso são:

- a informação por *pixel* é pequena e não é uma medida confiável para a intensidade média;
- sinais mais fortes (maior *backscatter*) não levam a resultados mais confiáveis; o erro amostral aumenta linearmente com a média da intensidade.

Portanto, um dos modos de se obter resultados mais confiáveis através de dados do SAR é através da média dos valores dos *pixels*. Isto pode ser feito no processamento dos dados, denominado processamento de múltiplas visadas ou *multi-look*, sacrificando-se a resolução espacial. No caso de dados *multi-look* de regiões homogêneas extensas, a intensidade terá uma distribuição Gamma.

Como foi mencionado anteriormente, as suposições feitas acima para regiões homogêneas extensas, muito embora tornem simples o modelo para radar, podem ser consideradas válidas para vários tipos de alvos, em especial campos agrícolas. Desvios dessas suposições têm sido considerados por vários autores. Um dos desvios considerados é quando um dos espalhadores na célula de resolução domina os outros, levando a uma distribuição Ricean da amplitude (64,128). Em (73) considera-se o caso onde o número de espalhadores não é grande o suficiente para que a suposição de normalidade seja satisfeita, e supõe-se que o número de espalhadores dentro da célula de resolução flutua aleatoriamente ao redor de um valor médio, levando a um modelo de Poisson para o processo de espalhamento. Em (72) propõe-se uma distribuição K para o processo de espalhamento, através de uma suposição binomial negativa para o número de espalhadores. Em (74) considera-se que a seção transversal possui uma distribuição Gamma, levando a uma distribuição K da intensidade. Em (115,116) considera-se a correlação entre os espalhadores, o

que equivale a introduzir não-estacionaridade (não homogeneidade) na média da intensidade. Em (100) considera-se, também, propriedades de segunda ordem das imagens complexas e de intensidade, quando a média da seção transversal do radar não é constante na região considerada, supondo que as variâncias do processo $n(s)$ variam.

Outras suposições para as distribuições de dados de radar incluem as distribuições Beta, Log-normal, Weibull e Gamma, entre outras (ver, por exemplo (98,142)). Estas hipóteses serão avaliadas, pelo uso dos testes Kolmogorov-smirnov e χ^2 , em dados obtidos na região de Tapajós pela missão SAREX. Este é um dos assuntos tratados no Capítulo 6. Alguns ajustes de distribuições a amostras de floresta e de pastagem podem se ver nas Figuras 5.3 e 5.4.

O modelo de *speckle* multiplicativo tem sido usado por vários autores (ver, por exemplo, os trabalhos (3,57,62,87,91,92,99), entre outros), em geral com dados de intensidade, isto é, $I(s) = \sigma(s)S(s)$, onde $S(s)$ representa o *speckle* (intensidade). Normalmente, nenhuma suposição é feita para as propriedades da superfície, mas se a mesma não for homogênea (isto é, se a média da seção transversal do radar não for constante na região considerada) este modelo multiplicativo deve ser usado com cuidado. Como ressaltado em (100,134), o modelo (5.2) é somente uma aproximação, que certamente não será boa se a superfície imageada apresentar variações abruptas.

Na prática, entretanto, muitos tipos de problemas (mau funcionamento do sistema ou má manipulação dos dados) podem afetar os dados do SAR e, conseqüentemente, suas propriedades estatísticas (ver, por exemplo, (139)).

5.2 Propriedades da Distribuição Rayleigh

Serão denotadas, para todo $i \in \mathbb{N}$, com Z_i as variáveis aleatórias com distribuição normal padrão, i.e., aquelas cuja densidade é:

$$f_{Z_i}(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \text{ para todo } z \in \mathbb{R}.$$

Distribuição 5.1 A variável aleatória W tem uma distribuição χ com $\tau > 0$ graus de liberdade, denotada χ_τ , se a sua densidade é dada por:

$$f_W(w) = \frac{w^{\tau-1}}{2^{\frac{\tau}{2}} \Gamma\left(\frac{\tau}{2}\right)} \exp\left(-\frac{w^2}{2}\right) \mathbb{I}_{\mathbb{R}^+}(w).$$

O momento de ordem k de W é dado por:

$$\mathbb{E}(W^k) = 2^{\frac{k}{2}} \frac{\Gamma\left(\frac{\tau+k}{2}\right)}{\Gamma\left(\frac{\tau}{2}\right)}. \quad (5.3)$$

Em (86) pode se ver que a distribuição da variável aleatória V definida, para todo $n \in \mathbb{N}$, por

$$V = \sqrt{\sum_{1 \leq i \leq n} Z_i^2} \quad (5.4)$$

é χ_n se, para todo $i \neq j$, Z_i é independente de Z_j . Em aplicações estatísticas é comum o uso da distribuição da variável aleatória V^2 , chamada χ^2 com n graus de liberdade, e denotada χ_n^2 , para testes de hipóteses.

Neste trabalho o interesse central radica no uso da distribuição χ_2 , chamada *distribuição Rayleigh padrão*. Esta é a distribuição da distância de um ponto em \mathbb{R}^2 à origem que, começando na origem $(0,0)$, anda em cada eixo uma quantidade normal padrão independente.

Seja a constante positiva $\xi \in \mathbb{R}^+$; diz-se que a variável aleatória Y tem distribuição de Rayleigh com parâmetro ξ (e é denotada $Y \sim \mathcal{R}(\xi)$) se a sua densidade é

$$f_\xi(y) = \frac{y}{\xi^2} \exp\left(-\frac{1}{2} \left(\frac{y}{\xi}\right)^2\right) \mathbb{I}_{\mathbb{R}^+}(y). \quad (5.5)$$

A sua função de distribuição acumulada é obtida integrando em relação a y :

$$F_\xi(t) = \left[1 - \exp\left(-\frac{1}{2} \left(\frac{t}{\xi}\right)^2\right) \right] \mathbb{I}_{\mathbb{R}^+}(t). \quad (5.6)$$

É imediato verificar que se $X \sim \mathcal{R}(1)$, então $\xi X \sim \mathcal{R}(\xi)$, isto é, as variáveis aleatórias Rayleigh são uma família de distribuições de escala.

Seja $F: \mathbb{R} \rightarrow [0, 1]$ a função de distribuição acumulada de uma variável aleatória qualquer Y , e $0 < \alpha < 1$, então o α -quantil de Y define-se como: $\inf\{t \in \mathbb{R}: F(t) \geq \alpha\}$. Esta quantidade é denotada y_α .

Usando a equação (5.6) é imediato obter-se o α -quantil de uma variável aleatória com distribuição $\mathcal{R}(\xi)$:

$$y_\alpha = \xi \sqrt{2 \ln \frac{1}{1-\alpha}},$$

e, assim, obter-se as suas mediana e distância inter-quartil, dadas por:

$$\text{Med}(Y) = y_{\frac{1}{2}} = \xi \sqrt{2 \ln 2}, \text{ e} \quad (5.7)$$

$$\text{IQR}(Y) = y_{\frac{3}{4}} - y_{\frac{1}{4}} = \xi \left(\sqrt{2 \ln 4} - \sqrt{2 \ln \frac{4}{3}} \right). \quad (5.8)$$

Para mais detalhes a respeito de estatísticas de ordem de variáveis aleatórias Rayleigh, pode se ver o livro (6).

Se $Y \sim \mathcal{R}(\xi)$, então usando a equação (5.3) e a propriedade de escala, podem calcular-se a sua esperança e variância:

$$\mathbb{E}(Y) = \sqrt{\frac{\pi}{2}} \xi \text{ e } \text{Var}(Y) = \left(2 - \frac{\pi}{2}\right) \xi^2, \quad (5.9)$$

respectivamente. A densidade de Y , equação (5.5), é mostrada na Figura 5.5 para dois valores diferentes do parâmetro ξ . Os valores dos parâmetros para desenhar estas densidades foram obtidos aplicando o estimador de máxima verossimilhança (equação (5.17)) sobre subregiões homogêneas, selecionadas visualmente, da Figura 5.6.

Outras propriedades de interesse desta distribuição são:

$$\mathbb{E}(Y - \mathbb{E}(Y))^k = (\sqrt{2}\xi)^k \frac{k}{2} \Gamma\left(\frac{k}{2}\right) \quad (5.10)$$

$$\gamma_1 = \frac{2\sqrt{\pi}(\pi - 3)}{(4 - \pi)^{3/2}} \quad (5.11)$$

$$\gamma_2 = \frac{32 - 3\pi^2}{(4 - \pi)^2} \quad (5.12)$$

$$m(Y) = \xi, \quad (5.13)$$

onde as equações (5.10), (5.11), (5.12) e (5.13) são, respectivamente, o momento central de ordem k , a assimetria, a kurtose e a moda de $Y \sim \mathcal{R}(\xi)$. É importante notar que as quantidades (5.11) e (5.12) não dependem do parâmetro ξ .

5.3 Estimação

A parte inteira de todo número real positivo $x \in \mathbb{R}^+$ será denotada $\lfloor x \rfloor$, i.e., $\lfloor x \rfloor = \max\{k \in \mathbb{N}: k \leq x\}$. Por simplicidade notacional, os conjuntos dos números pares e ímpares serão denotados \mathcal{E} e \mathcal{O} , respectivamente

Seja agora o vetor n -dimensional de reais $\mathbf{a} = (a_1, \dots, a_n)$, e \mathbf{a}_A um subconjunto de valores de \mathbf{a} , onde $\mathbf{a}_A = \{a_j\}$ com $j \in A \subset \{1, \dots, n\}$. O vetor \mathbf{a} ordenado em forma crescente será denotado $\mathbf{a}_{(\bullet)}$, onde $\mathbf{a}_{(\bullet)} = (a_{n:1}, \dots, a_{n:n})$ (i.e. $a_{n:1} \leq a_{n:2} \leq \dots \leq a_{n:n}$).

A avaliação de alguma das seguintes quantidades será chamada de *análise da distribuição empírica* do vetor \mathbf{a} .

- * Média amostral de $\mathbf{a} = \bar{\mathbf{a}} = \frac{1}{N} \sum_{i=1}^N a_i$.
- * Variância amostral de $\mathbf{a} = \widehat{\text{Var}}(\mathbf{a}) = S^2(\mathbf{a}) = \frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{\mathbf{a}})^2$.
- * Desvio padrão amostral de $\mathbf{a} = S(\mathbf{a}) = \sqrt{S^2(\mathbf{a})}$.
- * Assimetria amostral de $\mathbf{a} = \widehat{\gamma}_1(\mathbf{a}) = \frac{\widehat{\mu}_3(\mathbf{a})}{S^3(\mathbf{a})}$, onde

$$\widehat{\mu}_3(\mathbf{a}) = \frac{N}{(N-1)(N-2)} \sum_{1 \leq i \leq N} (a_i - \bar{\mathbf{a}})^3.$$

- * Kurtose amostral de $\mathbf{a} = \widehat{\gamma}_2(\mathbf{a}) = \frac{\widehat{\mu}_4(\mathbf{a})}{S^4(\mathbf{a})}$, onde

$$\begin{aligned} \widehat{\mu}_4(\mathbf{a}) &= k_N \sum_{1 \leq i \leq N} (a_i - \bar{\mathbf{a}})^4 - \frac{3(S^2(\mathbf{a}))^2(N-1)(2N-3)}{N(N-2)(N-3)}, \\ k_N &= \frac{N^2 - 2N + 3}{(N-1)(N-2)(N-3)} \end{aligned}$$

- * Variância da média amostral de $\mathbf{a} = S^2(\bar{\mathbf{a}}) = \frac{1}{N} S^2(\mathbf{a})$.
- * Mediana amostral de \mathbf{a} :

$$Q_2(\mathbf{a}) = \begin{cases} a_{n:\lfloor n/2 \rfloor + 1} & \text{se } n \in \mathcal{O} \\ \frac{1}{2}(a_{n:n/2} + a_{n:n/2+1}) & \text{se } n \in \mathcal{E}. \end{cases}$$

★ Quartil amostral inferior do vetor \mathbf{a} :

$$Q_1(\mathbf{a}) = \begin{cases} a_{n:(\ell+1)/2} & \text{se } \ell \in \mathcal{O} \\ \frac{1}{2}(a_{n:\ell/2} + a_{n:\ell/2+1}) & \text{se } \ell \in \mathcal{E}. \end{cases} \quad (5.14)$$

★ Quartil amostral superior do vetor \mathbf{a} :

$$Q_3(\mathbf{a}) = \begin{cases} a_{n:(n+1-(\ell+1)/2)} & \text{se } \ell \in \mathcal{O} \\ \frac{1}{2}(a_{n:n+1-\ell/2} + a_{n:n-\ell/2}) & \text{se } \ell \in \mathcal{E}. \end{cases} \quad (5.15)$$

Nas equações (5.14) e (5.15) foi empregada a notação:

$$\ell = \begin{cases} \frac{n-1}{2} & \text{se } n \in \mathcal{O} \\ \frac{n}{2} & \text{se } n \in \mathcal{E}. \end{cases}$$

Seja α um número real tal que $0 < \alpha < 1$, define-se então

★ o α -quantil amostral de $\mathbf{a} = \hat{\mathbf{a}}_\alpha$:

$$\begin{cases} a_{N:\alpha N} & \text{se } \alpha N \text{ é inteiro} \\ a_{N:[\alpha N]+1} & \text{se } \alpha N \text{ não é inteiro.} \end{cases}$$

★ O coeficiente de variação amostral de $\mathbf{a} = \tilde{C}(\mathbf{a}) = \frac{S(\mathbf{a})}{\bar{\mathbf{a}}}$.

En (96) pode ver-se um resumo bastante completo das propriedades mais relevantes das quantidades mencionadas acima, quando \mathbf{a} é um vetor tal que (a_1, \dots, a_N) são ocorrências de variáveis aleatórias independentes identicamente distribuídas.

Dado o vetor $\mathbf{y} = (y_1, \dots, y_v)^T$ de observações independentes e identicamente distribuídas segundo uma distribuição $\mathcal{R}(\xi)$, serão vistas algumas formas de se estimar o parâmetro ξ .

Uma técnica de estimação, muito usada, é a que consiste em se empregar a média amostral:

$$\hat{\xi}_{\text{MO}} = \frac{1}{v} \sqrt{\frac{2}{\pi}} \sum_{1 \leq i \leq v} y_i. \quad (5.16)$$

Outra, a de máxima verossimilhança, é a dada pela seguinte equação:

$$\hat{\xi}_{\text{ML}} = \sqrt{\frac{1}{2v} \sum_{1 \leq i \leq v} y_i^2}. \quad (5.17)$$

Ambos estimadores têm boas propriedades assintóticas, porém, podem apresentar problemas se forem aplicados a amostras que não satisfaçam as hipóteses de independência e igualdade de distribuição. A seguir são formulados dois modelos para o vetor de observações \mathbf{y} , que serão de utilidade para entender o problema de *contaminação* e, conseqüentemente, propor soluções para o mesmo (24).

Definição 5.1 *As variáveis aleatórias Y_1, \dots, Y_v satisfazem um modelo Rayleigh puro com parâmetro ξ se Y_1, \dots, Y_v são variáveis aleatórias independentes identicamente distribuídas, cada uma com densidade f_ξ . Denota-se $\mathcal{R}^n(\xi)$ a função de distribuição acumulada de $\mathbf{Y} = (Y_1, \dots, Y_v)^T$.*

Definição 5.2 *As variáveis aleatórias Y_1, \dots, Y_v satisfazem um modelo Rayleigh contaminado com parâmetro ξ e*

- *proporção de contaminação inferior α (com $\alpha > 0$),*
- *escala de contaminação inferior ξ_0 (com $0 < \xi_0 < \xi$);*
- *proporção de contaminação superior β (com $\beta > 0$ e $\alpha + \beta < \frac{1}{2}$), and*
- *escala de contaminação superior ξ_2 (com $\xi < \xi_2 < +\infty$)*

se

- *Y_1, \dots, Y_v são independentes,*
- *$Y_1, \dots, Y_{\lfloor \alpha v \rfloor}$ são identicamente distribuídas, cada uma com densidade f_{ξ_0} ,*
- *$Y_{\lfloor \alpha v \rfloor + 1}, \dots, Y_{v - \lfloor \beta v \rfloor}$ são identicamente distribuídas, cada uma com densidade f_ξ .*
- *$Y_{v - \lfloor \beta v \rfloor + 1}, \dots, Y_v$ são identicamente distribuídas, cada uma com densidade f_{ξ_2} .*

Denota-se $\mathcal{RC}^n(\xi_0, \xi, \xi_2; \alpha, \beta)$ a função de distribuição acumulada do vetor $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ neste caso.

5.4 Estimadores Robustos

A robustez estatística consiste em se propor estimadores que, quando aplicados a amostras de dados *contaminados* apresentam melhor desempenho que os estimadores clássicos (obtidos pelos métodos dos momentos ou de máxima verossimilhança, por exemplo); por outro lado, quando aplicados a amostras *puras* de dados, não apresentam um desempenho muito ruim com respeito aos estimadores clássicos. Uma definição de robustez, dada por (79) e citada por (17)), é a seguinte:

Robustez [...] um procedimento estatístico é chamado **robusto** se não é muito sensível a desvios das suposições sobre as quais se baseia. (Traduzido por O. H. Bustos)

Algumas referências gerais para assunto relacionados à robustez de estimadores são (8,17,18). Para ver algumas aplicações específicas de robustez ao processamento de imagens, podem consultar-se os artigos (14,84,88).

Seja o vetor y_W de observações independentes identicamente distribuídas segundo uma distribuição Rayleigh, de tamanho v). Pode-se definir, então, os seguintes estimadores:

$$\hat{\xi}_{\text{TML}} = \sqrt{\frac{1}{2(v-2a)} \sum_{i=a+1}^{v-a} y_{v:i}^2} \quad (5.18)$$

$$\hat{\xi}_{\text{TMO}} = \sqrt{\frac{2}{\pi} \frac{1}{v-2a} \sum_{i=a+1}^{v-a} y_{v:i}} \quad (5.19)$$

$$\hat{\xi}_{\text{MAD}} = \frac{1}{K_1} Q_2(\mathbf{z}_W), \quad (5.20)$$

$$\hat{\xi}_{\text{IQR}} = \frac{Q_3(\mathbf{y}_W) - Q_1(\mathbf{y}_W)}{K_2}, \quad (5.21)$$

$$\hat{\xi}_{\text{Med}} = \frac{1}{K_3} Q_2(\mathbf{y}_W), \quad (5.22)$$

onde $a = \lfloor v\alpha_0 \rfloor$, $0 \leq \alpha_0 < 1/2$, e $(y_{v:1}, \dots, y_{v:v})$ denota o vetor $\mathbf{y}_W = (y_1, \dots, y_v)$ ordenado em forma crescente; $\mathbf{z} = (z_1, \dots, z_v)$, com $z_i = |y_i - Q_2(\mathbf{y})|$ para todo $i \in W$. As observações aparadas são as $\lfloor v\alpha_0 \rfloor$ menores e as $\lfloor v\alpha_0 \rfloor$ maiores.

As equações acima definem os métodos de estimação por: *máxima verossimilhança (ML) aparada* com proporção de observações aparadas igual a $2\alpha_0$ (equação (5.18)), *momentos (MO) aparados* com proporção de observações aparadas igual a $2\alpha_0$ (equação (5.19)), mediana do desvio absoluto (MAD) (*Median Absolute Deviation*) (5.20), distância inter-quartil (5.21), e mediana (5.22).

As constantes K_1 , K_2 e K_3 são calculadas de tal forma que os respectivos estimadores são assintoticamente consistentes, da seguinte maneira: para K_2 e para K_3 é suficiente conhecer os valores $y_{1/4}$, $y_{1/2}$ e $y_{3/4}$, e aplicar o método de substituição. Desta maneira, usando as equações (5.7) e (5.8), chega-se aos valores $K_2 = \sqrt{2 \ln 4} - \sqrt{2 \ln 4/3}$ e $K_3 = \sqrt{2 \ln 2}$. A constante K_1 pode obter-se usando ferramentas numéricas, e o seu valor aproximado é $K \approx 0.4485$.

Para se obter este último valor é necessário conhecer a distribuição da variável aleatória $Z = |X - \text{Med}(X)|$, onde $X \sim \mathcal{R}(1)$, e definir $K_1^{-1} = \text{Med}(Z)$. Para tanto, seja $m = \sqrt{2 \ln 2}$; então a função de distribuição acumulada de Z é dada por:

$$F_Z(y) = \begin{cases} 0 & \text{se } y < 0, \\ \exp(-\frac{1}{2}(m-y)^2) - \exp(-\frac{1}{2}(m+y)^2) & \text{se } 0 \leq y < m, \\ 1 - \exp(-\frac{1}{2}(m+y)^2) & \text{se } y \geq m. \end{cases}$$

De posse desta função de distribuição acumulada, é possível calcular as densidades para diferentes tamanhos de amostra. Na Figura 5.7 tem-se o desenho desta densidade, e das densidades do estimador MAD para amostras de tamanhos cinco e nove. Desta figura conclui-se que quanto maior for a amostra, mais *normalmente* distribuído estará o estimador MAD.

5.5 Avaliação de Estimadores

A seguir descreve-se uma experiência Monte Carlo para avaliar o desempenho dos estimadores para os modelos puro e contaminado (24). Fixar os valores iniciais $N = 30000$, $I = 10$, $\alpha_0 = .05$. Fixar os valores dos parâmetros $\xi = 0.5$, $\xi_0 = 0.1$ e $\xi_2 = 1$. Para cada $i = 1, 2, \dots, 10$ definir $n(i) = 100i$ e

$$M(i) = \lfloor N/n(i) \rfloor.$$

- E1) Atribuir $d = 1$.
- E2) Atribuir $i = 1$.
- E3) Atribuir $m = 1$.
- E4) Atribuir $a(m) = (m - 1)n(i)$.
- E5) Gerar $z_{a(m)+1}, \dots, z_{a(m)+n(i)}$, pegando amostras do vetor aleatório $\mathbf{Y} = (Y_1, \dots, Y_{n(i)})^T$ com função de distribuição acumulada $\mathcal{R}^{n(i)}(\xi)$ se $d = 1$, e com $\mathcal{RC}^{n(i)}(\xi_0, \xi, \xi_2; \alpha, \beta)$ se $d = 2$.
- E6) Atribuir $\mathbf{z} = (z_{a(m)+1}, \dots, z_{a(m)+n(i)})$ e calcular
- $\xi_{\text{ML},n(i)}(m) := \xi_{\text{ML}}(\mathbf{z})$,
 - $\xi_{\text{MO},n(i)}(m) := \xi_{\text{MO}}(\mathbf{z})$,
 - $\xi_{\text{TML},n(i)}(m) := \xi_{\text{TML}}(\mathbf{z})$,
 - $\xi_{\text{TMO},n(i)}(m) := \xi_{\text{TMO}}(\mathbf{z})$ and
 - $\xi_{\text{MAD},n(i)}(m) := \xi_{\text{MAD}}(\mathbf{z})$.
- E7) Se $m = M(i)$ continuar no passo E8. Caso contrário, atribuir $m = m + 1$ e voltar ao passo E4.
- E8) Se $i = I$ continuar no passo E9. Caso contrário, atribuir $i = i + 1$ e continuar no passo E3.
- E9) Se $d = 2$ continuar no passo E10. Caso contrário, atribuir $d = d + 1$ e continuar no passo E2.
- E10) Para cada $d = 1, 2$ e para cada $i = 1, \dots, I$ definir os seguintes vetores:
- $\xi_{\text{ML},n(i)} = (\xi_{\text{ML},n(i)}(1), \dots, \xi_{\text{ML},n(i)}(M(i)))^T$,
 - $\xi_{\text{MO},n(i)} = (\xi_{\text{MO},n(i)}(1), \dots, \xi_{\text{MO},n(i)}(M(i)))^T$,
 - $\xi_{\text{TML},n(i)} = (\xi_{\text{TML},n(i)}(1), \dots, \xi_{\text{TML},n(i)}(M(i)))^T$,
 - $\xi_{\text{T},n(i)} = (\xi_{\text{TMO},n(i)}(1), \dots, \xi_{\text{TMO},n(i)}(M(i)))^T$ and
 - $\xi_{\text{MAD},n(i)} = (\xi_{\text{MAD},n(i)}(1), \dots, \xi_{\text{MAD},n(i)}(M(i)))^T$.
- Analisar a distribuição empírica de cada um dos vetores definidos acima.
- E11) Mostrar os resultados em uma tabela, por exemplo seguindo o formato sugerido em (96), frisando:

- Distribuição gerada.
- Tamanho total da amostra observada, da distribuição considerada, para cada $i = 1, 2, \dots, I$, por exemplo N .
- Tamanho da amostra sobre o qual o procedimento de estimação se baseia: n_1, \dots, n_I .
- Número de procedimentos observados, para cada $i = 1, 2, \dots, I$, por exemplo $M(1), \dots, M(I)$. Note-se que:

$$N \approx n(1)M(1) \approx n(2)M(2) \approx \dots \approx n(I)M(I). \quad (5.23)$$

A relação apresentada na equação (5.23) é importante pois, sob certas hipóteses (ver, por exemplo, a referência (95)), valem as seguintes expansões assintóticas:

$$\mathbb{E}(\Theta_n) = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots \quad (5.24)$$

$$\text{Var}(\Theta_n) = \frac{b_1}{n} + \frac{b_2}{n^2} + \frac{b_3}{n^3} + \dots \quad (5.25)$$

Assim sendo, é possível prosseguir o estudo analisando as regressões

$$Z_i = \theta + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + \varepsilon_i, \quad i = 1, \dots, I, \quad (5.26)$$

onde $Z_i = \bar{\Theta}_{n(i)}$ é o estimador de $\mathbb{E}(\Theta)_{n(i)}$, “média amostral” sobre amostras de $\Theta_{n(i)}$ de tamanho $M(i)$, $x_{i1} = \frac{1}{n(i)}$, $x_{i2} = \frac{1}{n(i)^2}$, $x_{i3} = \frac{1}{n(i)^3}$, $i = 1, \dots, I$. Assim, $\Theta_{n(i)}$ representa qualquer dos $\xi_{\text{ML},n(i)}, \dots, \xi_{\text{MAD},n(i)}$ (em total 100).

Dado que $\text{Var}(Z_i) = \frac{1}{M(i)} \text{Var}(\Theta_{n(i)})$, pela equação (5.25) tem-se que, quando $n(i)$ é “grande” vale que:

$$\text{Var}(Z_i) \approx \frac{b_1}{n(i)M(i)} \approx \frac{b_1}{N} \text{ para todo } i = 1, \dots, I,$$

usando a equação (5.24). Portanto, $\text{Var}(Z_1) \approx \text{Var}(Z_2) \approx \dots \approx \text{Var}(Z_I)$. Além disso, levando em consideração a estrutura da simulação, pode considerar-se que Z_1, \dots, Z_I são independentes (note-se que, para cada $i = 1, \dots, I$ a geração continua, em vez de retornar ao início).

Pode, então, aplicar-se um modelo de regressão ao modelo (5.26) considerando os valores $\bar{\theta}_{n(1)}, \dots, \bar{\theta}_{n(I)}$ como os valores observados de Z_1, \dots, Z_I .

Logo, por exemplo usando estimadores de mínimos quadrados, podem se estimar os coeficientes de (5.26) com:

$$\hat{\theta}_{\text{reg}}, \quad \hat{a}_{1,\text{reg}}, \quad \hat{a}_{2,\text{reg}}, \quad \hat{a}_{3,\text{reg}}. \quad (5.27)$$

Para se afirmar alguma coisa a respeito da precisão dos estimadores definidos em (5.27), poder-se-ia prosseguir a simulação obtendo-se, por exemplo R , valores como:

$$\begin{aligned} &\hat{\theta}_{\text{reg}}(1), \dots, \hat{\theta}_{\text{reg}}(R) \\ &\hat{a}_{1,\text{reg}}(1), \dots, \hat{a}_{1,\text{reg}}(R) \\ &\vdots \end{aligned}$$

Logo, poder-se-ia analisar as distribuições empíricas de:

$$\begin{aligned} &(\hat{\theta}_{\text{reg}}(1), \dots, \hat{\theta}_{\text{reg}}(R)) \\ &(\hat{a}_{1,\text{reg}}(1), \dots, \hat{a}_{1,\text{reg}}(R)), \text{ etc.}, \end{aligned}$$

e, desta análise, poderiam ser obtidas as precisões dos estimadores propostos para o modelo (5.26). Embora o interesse deste estudo esteja centralizado em θ , um estudo de a_1 também é interessante, pois pode fornecer informação a respeito do vício assintótico do procedimento.

Após serem feitas as R replicações, têm-se $RM(i)$ ocorrências da variável aleatória $\Theta_{n(i)}$: $(\theta_{n(i)}(1), \dots, \theta_{n(i)}(M(i)), \dots, \theta_{n(i)}(RM(i)))$. Esta amostra permitiria mais estudos sobre a distribuição empírica.

As Tabelas 5.1, 5.2 e 5.3 apresentam alguns resultados das experiências acima definidas. Os valores mostrados na Tabela 5.1 foram arredondados à quarta casa decimal, pois foram empregados para a construção da Tabela 5.4; os valores das Tabelas 5.2 e 5.3 foram arredondados à terceira casa decimal.

Tal como foi mencionado na definição do algoritmo, os valores de ξ_0 , ξ , ξ_2 , α e β são mantidos constantes durante a experiência toda; pode, assim, escrever-se $\mathcal{R}^{n(i)}$ e $\mathcal{RC}^{n(i)}$ como formas abreviadas de $\mathcal{R}^{n(i)}(\xi)$ e $\mathcal{RC}^{n(i)}(\xi_0, \xi, \xi_2; \alpha, \beta)$, respectivamente.

TABELA 5.1 - VALORES AMOSTRAIS DE MÉDIA, DESVIO PADRÃO MÉDIO E DESVIO PADRÃO DOS ESTIMADORES PARA $M(i) = \lfloor 30000/n(i) \rfloor$ VETORES DE TAMANHO $n(i) = 100i$

—	ξ_{ML}		ξ_{MO}		ξ_{TML}		ξ_{TMO}		ξ_{MAD}	
	$\mathcal{R}^{n(i)}$	$\mathcal{RC}^{n(i)}$								
1	.4995	.5226	.5021	.5056	.4742	.4777	.4921	.4870	.4983	.5424
	.0026	.0026	.0028	.0026	.0025	.0024	.0028	.0026	.0032	.0037
	.0454	.0458	.0482	.0451	.0436	.0420	.0480	.0448	.0553	.0646
2	.4990	.5286	.5021	.5093	.4732	.4803	.4921	.4894	.4941	.5476
	.0028	.0029	.0029	.0028	.0026	.0031	.0028	.0025	.0035	.0033
	.0337	.0356	.0354	.0347	.0318	.0308	.0349	.0341	.0428	.0408
3	.5000	.5204	.5005	.5022	.4723	.4734	.4895	.4828	.4981	.5377
	.0029	.0031	.0029	.0029	.0026	.0027	.0028	.0029	.0031	.0034
	.0290	.0307	.0292	.0293	.0262	.0269	.0282	.0289	.0314	.0432
4	.5005	.5186	.5018	.5004	.4724	.4724	.4906	.4811	.4993	.5410
	.0025	.0030	.0027	.0029	.0024	.0027	.0027	.0028	.0032	.0038
	.0218	.0263	.0238	.0248	.0211	.0231	.0237	.0244	.0273	.0328
5	.4998	.5195	.5000	.5001	.4711	.4724	.4886	.4806	.4980	.5510
	.0026	.0029	.0027	.0030	.0024	.0027	.0027	.0030	.0034	.0034
	.0205	.0227	.0213	.0235	.0185	.0211	.0206	.0236	.0266	.0267
6	.4974	.5254	.4979	.5054	.4690	.4770	.4867	.4856	.4944	.5492
	.0028	.0025	.0027	.0024	.0025	.0022	.0026	.0027	.0038	.0041
	.0199	.0193	.0194	.0185	.0173	.0171	.0186	.0191	.0265	.0287
7	.4994	.5169	.5007	.4971	.4716	.4691	.4899	.4772	.4958	.5423
	.0025	.0032	.0027	.0029	.0023	.0026	.0026	.0029	.0030	.0035
	.0165	.0207	.0172	.0190	.0152	.0170	.0167	.0186	.0197	.0229
8	.4999	.5247	.5005	.5053	.4712	.4761	.4890	.4852	.4979	.5436
	.0029	.0035	.0029	.0033	.0026	.0029	.0028	.0031	.0032	.0040
	.0174	.0215	.0175	.0201	.0155	.0174	.0168	.0188	.0192	.0197
9	.5037	.5273	.5035	.5057	.4746	.4759	.4920	.4845	.5039	.5469
	.0029	.0025	.0030	.0024	.0026	.0025	.0028	.0025	.0041	.0047
	.0168	.0142	.0171	.0137	.0149	.0143	.0163	.0145	.0233	.0271
10	.4989	.5255	.4992	.5075	.4705	.4790	.4882	.4881	.4981	.5459
	.0021	.0027	.0022	.0026	.0019	.0023	.0022	.0026	.0025	.0031
	.0113	.0150	.0121	.0144	.0104	.0127	.0119	.0145	.0136	.0168

FONTE: (24)

TABELA 5.2 - VALORES AMOSTRAIS DE MÍNIMO, MEDIANA E MÁXIMO
 DOS ESTIMADORES PARA $M(i) = \lfloor 30000/n(i) \rfloor$ VETORES DE
 TAMANHO $n(i) = 100i$

—	ξ_{ML}		ξ_{MO}		ξ_{TML}		ξ_{TMO}		ξ_{MAD}	
	$\mathcal{R}^{n(i)}$	$\mathcal{RC}^{n(i)}$								
1	.392	.389	.383	.355	.373	.331	.373	.332	.375	.397
	.499	.525	.502	.508	.475	.481	.492	.488	.498	.539
	.636	.651	.647	.622	.607	.595	.642	.593	.687	.784
2	.431	.445	.435	.430	.413	.409	.427	.411	.408	.455
	.495	.529	.497	.509	.469	.479	.488	.487	.490	.549
	.640	.629	.654	.607	.608	.569	.639	.590	.690	.637
3	.454	.453	.455	.439	.429	.420	.443	.421	.433	.447
	.495	.520	.495	.501	.469	.472	.486	.481	.496	.539
	.588	.633	.595	.607	.561	.575	.585	.585	.583	.660
4	.453	.461	.452	.454	.432	.429	.446	.436	.442	.472
	.498	.515	.495	.496	.468	.471	.484	.476	.498	.536
	.549	.608	.552	.585	.516	.547	.541	.563	.576	.639
5	.464	.473	.462	.453	.441	.424	.451	.430	.450	.486
	.498	.522	.500	.500	.470	.471	.489	.479	.500	.548
	.567	.584	.565	.574	.528	.545	.548	.558	.587	.628
6	.460	.487	.461	.464	.437	.443	.452	.445	.441	.499
	.497	.522	.499	.504	.470	.475	.489	.486	.491	.542
	.544	.567	.538	.555	.510	.528	.524	.540	.559	.601
7	.465	.482	.466	.465	.441	.442	.457	.445	.456	.495
	.500	.516	.499	.497	.469	.465	.486	.474	.495	.542
	.541	.570	.539	.550	.511	.518	.528	.533	.565	.594
8	.466	.488	.463	.468	.440	.444	.452	.450	.447	.497
	.501	.523	.504	.505	.473	.473	.490	.483	.497	.545
	.536	.569	.533	.538	.501	.512	.518	.517	.545	.589
9	.473	.504	.470	.486	.448	.455	.461	.466	.478	.489
	.500	.528	.500	.503	.472	.473	.489	.483	.500	.544
	.537	.552	.542	.530	.506	.499	.530	.511	.567	.592
10	.478	.504	.479	.479	.453	.454	.470	.456	.472	.518
	.498	.521	.497	.507	.470	.479	.486	.489	.499	.544
	.528	.561	.537	.535	.501	.505	.526	.520	.542	.591

FONTE: (24)

TABELA 5.3 - QUARTIS INFERIOR E SUPERIOR DOS ESTIMADORES:

 $M(i) = \lfloor 30000/n(i) \rfloor$ VETORES DE TAMANHO $n(i) = 100i$

—	ξ_{ML}		ξ_{MO}		ξ_{TML}		ξ_{TMO}		ξ_{MAD}	
	$\mathcal{R}^{n(i)}$	$\mathcal{RC}^{n(i)}$								
1	.469	.486	.470	.472	.445	.448	.460	.458	.460	.498
	.527	.553	.531	.536	.499	.505	.521	.519	.532	.579
2	.477	.505	.479	.489	.452	.461	.470	.468	.464	.521
	.519	.551	.525	.531	.493	.499	.513	.511	.519	.577
3	.477	.498	.480	.483	.454	.456	.471	.467	.476	.513
	.519	.539	.519	.521	.490	.490	.511	.500	.517	.560
4	.483	.502	.484	.486	.456	.459	.474	.466	.482	.521
	.518	.534	.521	.514	.489	.485	.510	.494	.516	.563
5	.472	.501	.482	.486	.458	.461	.472	.466	.481	.529
	.501	.535	.514	.515	.482	.484	.501	.494	.513	.561
6	.483	.510	.484	.492	.455	.465	.470	.473	.478	.528
	.510	.535	.510	.516	.481	.489	.499	.497	.509	.575
7	.486	.504	.487	.485	.460	.456	.477	.463	.485	.525
	.511	.527	.513	.506	.482	.478	.500	.486	.506	.558
8	.486	.508	.487	.490	.460	.465	.476	.471	.485	.532
	.512	.540	.515	.524	.484	.494	.504	.501	.531	.554
9	.493	.517	.494	.494	.463	.464	.482	.472	.486	.530
	.520	.540	.514	.521	.488	.489	.502	.500	.516	.568
10	.492	.515	.492	.499	.462	.471	.481	.477	.490	.536
	.504	.538	.504	.520	.475	.489	.578	.493	.502	.553

FONTE: (24)

TABELA 5.4 - O ESTIMADOR ESTÁ DENTRO DO INTERVALO EM TORNO DO VALOR ESTIMADO?

—	ξ_{ML}		ξ_{MO}		ξ_{TML}		ξ_{TMO}		ξ_{MAD}	
	$\mathcal{R}^{n(i)}$	$\mathcal{RC}^{n(i)}$								
1	S(2)	S(4)	S(4)	S(3)	S(1)	S(1)	S(3)	S(2)	S(5)	S(5)
2	S(2)	S(4)	S(4)	S(3)	S(1)	S(1)	S(3)	S(2)	S(5)	S(5)
3	S(3)	S(4)	S(4)	S(3)	S(1)	S(1)	S(3)	S(2)	S(5)	S(5)
4	S(2)	S(4)	S(4)	S(3)	S(1)	S(1)	S(3)	S(2)	S(5)	S(5)
5	S(2)	S(2)	S(4)	S(3)	S(1)	S(1)	S(3)	S(4)	S(5)	S(5)
6	S(4)	S(4)	S(3)	S(2)	S(1)	S(1)	S(2)	S(3)	S(5)	S(5)
7	S(2)	S(4)	S(4)	S(3)	S(1)	S(1)	S(3)	S(2)	S(5)	S(5)
8	S(3)	S(4)	S(4)	S(3)	S(1)	S(1)	S(2)	S(2)	S(5)	N
9	S(3)	S(2)	S(4)	S(1)	S(1)	S(3)	S(2)	S(4)	S(5)	S(5)
10	S(1)	S(4)	S(3)	S(2)	N	S(1)	S(2)	S(3)	S(4)	N

FONTE: Adaptada de (24)

A conclusão mais importante surge da Tabela 5.4. Para todos os casos considerados, o estimador que *atinge* o valor verdadeiro do parâmetro a ser estimado, isto é aquele que $\xi \in [\bar{\xi} \pm 2s(\hat{\xi})]$, com a menor dispersão em torno da média amostral é o ξ_{TML} , com exceção dos seguintes casos:

- $\mathcal{R}^{n(10)}$, quando o *melhor* estimador é o ξ_{ML} ;
- $\mathcal{RC}^{n(9)}$, quando o *melhor* estimador é o ξ_{MO} .

Para amostras puras, a teoria afirma que o estimador ξ_{ML} é assintoticamente consistente (não viciado) e eficiente (de uniformemente mínima variância). Desta experiência conclui-se um fato interessante: que o estimador ξ_{TML} é melhor do que o estimador ξ_{ML} , para amostras de até 900 observações. Este fato motiva uma futura análise mais profunda das propriedades destes, e de outros estimadores, para amostras finitas.

Da Tabela 5.5 pode-se concluir que um bom modelo de regressão polinomial para todos os estimadores é o identicamente constante. Desta maneira, constata-se que não há evidência de uma grande dependência do comportamento dos estimadores em relação ao tamanho da amostra, para as condições da experiência.

TABELA 5.5 - COEFICIENTES DA REGRESSÃO SOBRE OS ESTIMADORES

—	ξ_{ML}		ξ_{MO}		ξ_{TML}		ξ_{TMO}		ξ_{MAD}	
	$\mathcal{R}^{n(i)}$	$\mathcal{RC}^{n(i)}$								
$\hat{\theta}_{reg}$.4998	.5230	.5008	.5039	.4720	.4754	.4899	.4842	.4978	.5448

FONTE: (24)

5.6 Estimador de Máxima Verossimilhança Aproximado para Amostras Aparadas

Uma alternativa interessante para o estimador de máxima verossimilhança aparado (equação (5.18)) é dado em (5). Este estimador provém de uma generalização do trabalho de Dyer e Whisenand (45), que consideram amostras *pequenas* com deleção de observações.

Neste último trabalho, pode se ver os seguintes resultados, que serão de utilidade para a obtenção dos estimadores: seja $\mathbf{X} = (X_1, \dots, X_n)$ um vetor de variáveis aleatórias independentes e identicamente distribuídas segundo uma distribuição Rayleigh padrão (consideramos apenas este caso por simplicidade, para maior generalidade basta lembrar a propriedade das distribuições Rayleigh formarem uma família de escala). Seja $\mathbf{X}_\bullet = (X_{1:n}, \dots, X_{n:n})$ o vetor das variáveis aleatórias ordenadas em ordem crescente. Em (11,76) pode se ver que a densidade conjunta de \mathbf{X}_\bullet é dada por:

$$f_{\mathbf{X}_\bullet}(\mathbf{x}) = \begin{cases} n! \prod_{i=1}^n x_i \exp\left(-\frac{x_i^2}{2}\right) & \text{se } x_1 < \dots < x_n, \\ 0 & \text{caso contrário;} \end{cases}$$

e que a densidade da ℓ -ésima componente do vetor \mathbf{X}_\bullet é dada por:

$$f_{X_{\ell:n}}(x) = \frac{n!}{(\ell-1)!(n-\ell)!} \left(1 - \exp\left(-\frac{x^2}{2}\right)\right)^{\ell-1} \left(\exp\left(-\frac{x^2}{2}\right)\right)^{n-\ell} x \exp\left(-\frac{x^2}{2}\right) \mathbb{I}_{\mathbb{R}_+}(x).$$

Assim, é possível escrever as seguintes relações:

$$\alpha_{\ell:n} = \mathbb{E}(X_{\ell:n}) = \sqrt{\frac{\pi}{2}} n \binom{n-1}{\ell-1} \sum_{r=0}^{\ell-1} (-1)^{\ell-1-r} \frac{\binom{\ell-1}{r}}{(n-r)^{\frac{3}{2}}},$$

$$\begin{aligned}\mathbb{E}(X_{\ell:n}^2) &= 2n \binom{n-1}{\ell-1} \sum_{r=0}^{\ell-1} (-1)^{\ell-1-r} \frac{\binom{\ell-1}{r}}{(n-r)^2}, \\ \beta_{\ell,m:n} &= \text{Cov}(X_{\ell}, X_m) = \\ &= k_{1,\ell,m,n} \sum_{r=0}^{m-\ell-1} \sum_{s=0}^{\ell-1} (-1)^{r+s} k_{2,\ell,m,n} H(m-\ell-r+s, n-m+r+1),\end{aligned}$$

onde

$$\begin{aligned}k_{1,\ell,m,n} &= \frac{n!}{(\ell-1)! (m-\ell-1)! (n-m)!} \\ k_{2,\ell,m,n} &= \binom{m-\ell-1}{r} \binom{\ell-1}{s} \\ H(a,b) &= \frac{\frac{\pi}{2} - \arctan \sqrt{\frac{a}{b}} + \sqrt{ab} \frac{a-b}{(a+b)^2}}{(ab)^{\frac{3}{2}}}.\end{aligned}$$

Para analisar amostras do vetor $\xi(X_{r_1+1:n}, \dots, X_{n-r_2:n})$, resultante de aparar as primeiras r_1 e as últimas r_2 observações de $\xi(X_1, \dots, X_n)$, serão de utilidade os seguintes vetor e matriz:

$$\begin{aligned}\boldsymbol{\alpha} &= (\alpha_{r_1+1:n}, \dots, \alpha_{n-r_2:n}); \\ \boldsymbol{\beta} &= \begin{pmatrix} \beta_{r_1+1, r_1+1:n} & \cdots & \beta_{r_1+1, n-r_2:n} \\ \vdots & \vdots & \vdots \\ \beta_{r_1+1, n-r_2:n} & \cdots & \beta_{n-r_2, n-r_2:n} \end{pmatrix},\end{aligned}$$

e é evidente que $\boldsymbol{\beta}$ é uma matriz quadrada definida simétrica de ordem $n - r_1 - r_2$. Denote-se, ainda, $\underline{\mathbf{Y}}^T = (Y_{r_1+1}, \dots, Y_{n-r_2})^T = \xi(X_{r_1+1}, \dots, X_{n-r_2})^T$.

O estimador BLUE (*Best Linear Unbiased Estimator*: melhor estimador linear não viciado) de ξ baseado no vetor $\underline{\mathbf{Y}}$ é dado por

$$\xi_{\text{BLUE}} = \begin{pmatrix} \boldsymbol{\alpha}' \boldsymbol{\beta}^{-1} \\ \boldsymbol{\alpha}' \boldsymbol{\beta}^{-1} \boldsymbol{\alpha} \end{pmatrix} \mathbf{Y} = \sum_{i=r_1+1}^{n-r_2} a_i Y_{i:n}, \quad (5.28)$$

e a sua variância é dada por

$$\text{Var}(\xi_{\text{BLUE}}) = \frac{1}{(\boldsymbol{\alpha}' \boldsymbol{\beta}^{-1} \boldsymbol{\alpha})^2} \boldsymbol{\alpha}' \boldsymbol{\beta}^{-1} (\xi^2 \boldsymbol{\beta}) \boldsymbol{\beta}^{-1} \boldsymbol{\alpha} = \frac{\xi^2}{\boldsymbol{\alpha}' \boldsymbol{\beta}^{-1} \boldsymbol{\alpha}}.$$

Em (45) encontram-se tabulados os valores do vetor $\boldsymbol{\alpha}$ para amostras pequenas. Dada a aplicação que será dada a este estimador, somente será

considerado o caso $r_1 = r_2 = r$, e valores de $n = 5$ e $n = 9$, e os valores reproduzidos na Tabela 5.6. Para outros valores de n se fazem necessárias algumas aproximações, que serão vistas mais adiante.

TABELA 5.6 - COEFICIENTES DO ESTIMADOR “BLUE” PARA PEQUENAS AMOSTRAS

n	r	$a_{i:n}$
5	1	$a_{2:5} = 0.1588$
—	—	$a_{3:5} = 0.1537$
—	—	$a_{4:5} = 0.4355$
9	1	$a_{2:9} = 0.0593$
—	—	$a_{3:9} = 0.0535$
—	—	$a_{4:9} = 0.0646$
—	—	$a_{5:9} = 0.0757$
—	—	$a_{6:9} = 0.0876$
—	—	$a_{7:9} = 0.1011$
—	—	$a_{8:9} = 0.2555$
9	2	$a_{3:9} = 0.1106$
—	—	$a_{4:9} = 0.0737$
—	—	$a_{5:9} = 0.0863$
—	—	$a_{6:9} = 0.0996$
—	—	$a_{7:9} = 0.3712$
9	3	$a_{4:9} = 0.1854$
—	—	$a_{5:9} = 0.1001$
—	—	$a_{6:9} = 0.5020$

FONTE: Adaptada de (45), p. 32

Embora seja possível, usando ferramentas computacionais, determinar os valores das componentes do vetor α para outros casos de n e de r , em (5) propõe-se uma técnica alternativa bastante atraente. Considere-se o vetor

$\underline{Y} = (Y_{r+1:n}, \dots, Y_{n-r})^T$; a sua densidade conjunta é

$$\begin{cases} \frac{n!}{(r!)^2} \xi^{n-2r} \left(\left(1 - \exp\left(-\frac{y^2}{2}\right) \right) \exp\left(-\frac{y^2}{2}\right) \right)^r \prod_{i=r+1}^{n-r} y_{i:n}^2 \exp\left(-\frac{y_{i:n}^2}{2}\right) & \text{se} \\ & y_{r+1:n} < \dots < y_{n-r:n} \\ 0 & \text{caso contrário,} \end{cases} \quad (5.29)$$

A equação (5.29), em geral, não admite solução para ξ mas, fazendo-se algumas aproximações (para detalhes, ver (5)), pode chegar-se à seguinte aproximação:

$$\begin{aligned} \hat{\xi}_{\text{BLUE}} &\approx \frac{\sqrt{B^2 + 8AC} - B}{4A}, \text{ onde} & (5.30) \\ A &= n - 2r, \\ B &= \frac{r(n-r)(n+1)}{(r+1)^2} \left(-2 \ln \left(\frac{n-r}{n+1} \right) \right)^{3/2}, \\ C &= \sum_{i=r+1}^{n-r} Y_{i:n}^2 + rY_{n-r:n}^2 - \frac{r(n-r)}{r+1} \left(1 + \frac{2(r+1)}{n+1} \ln \left(\frac{n-r}{n+1} \right) \right) Y_{r+1:n}^2. \end{aligned}$$

5.7 Exemplo

Foi feita uma experiência de simulação para ilustrar o funcionamento dos estimadores propostos no decorrer deste Capítulo. Para tanto, foram geradas independentemente quinze amostras da variável aleatória $Y \sim \mathcal{R}(100)$ e aplicados os estimadores. Os resultados são apresentados arredondados à quarta casa decimal, porém, a experiência foi feita, e os estimadores calculados, sem serem afetados arredondamentos.

O primeiro vetor apresentado em (5.31), \mathbf{y} , é o vetor das observações. A partir dele e de sua forma ordenada (o segundo vetor, \mathbf{y}_\bullet) pode-se calcular os valores $\widehat{m}_1 = 125.5013$, $\widehat{m}_2 = 2.0136 \cdot 10^4$, $Q_1(\mathbf{y}) = 72.2633$, $Q_2(\mathbf{y}) = 118.7659$, $Q_3(\mathbf{y}) = 162.7093$ e o terceiro vetor mostrado em (5.31), \mathbf{z} . A mediana do vetor \mathbf{z} , definido na pág. 129, é $Q_2(\mathbf{z}) = 44.9490$.

Destes valores, e das definições apresentadas com $\alpha_0 = 0.225$, podem-se calcular os estimadores apresentados na última coluna de (5.31).

$$\begin{aligned}
\mathbf{y} = & \begin{pmatrix} 118.7659 \\ 162.7093 \\ 219.0815 \\ 32.7565 \\ 228.3457 \\ 117.6209 \\ 138.3288 \\ 160.2150 \\ 42.4557 \\ 112.7258 \\ 123.8475 \\ 34.8808 \\ 244.7061 \\ 73.8169 \\ 72.2633 \end{pmatrix}, \quad \mathbf{y}_\bullet = \begin{pmatrix} 32.7565 \\ 34.8808 \\ 42.4557 \\ 72.2633 \\ 73.8169 \\ 112.7258 \\ 117.6209 \\ 118.7659 \\ 123.8475 \\ 138.3288 \\ 160.2150 \\ 162.7093 \\ 219.0815 \\ 228.3457 \\ 244.7061 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} 86.0094 \\ 83.8851 \\ 76.3102 \\ 46.5026 \\ 44.9490 \\ 6.0401 \\ 1.1450 \\ 0 \\ 5.0816 \\ 19.5629 \\ 41.4491 \\ 43.9434 \\ 100.3156 \\ 109.5798 \\ 125.9402 \end{pmatrix}; \\
& \xi_{MV} = 100.3404 \\
& \xi_{MO} = 100.1356 \\
& \xi_{TMV} = 87.5386 \\
& \xi_{TMO} = 95.7722 \\
& \xi_{MAD} = 100.2208 \\
& \xi_{Med} = 100.8705 \\
& \xi_{IQR} = 99.7660 \\
& \xi_{BLUE} = 98.2868.
\end{aligned} \tag{5.31}$$

Para esta amostra, provinda do modelo puro e obtida pelo uso do pacote MatLab, o melhor estimador nos sentidos de erro quadrático e de erro absoluto, é o ξ_{MO} .

5.8 Filtros Robustos para Deteção Linear e Uma Visada

Com os resultados distribucionais a respeito da distribuição Rayleigh apresentados no decorrer deste trabalho, é possível projetar filtros (estimadores) que apresentam características de robustez.

Para se amenizar o aspecto ruidoso das imagens SAR (notadamente as de uma visada), é possível aplicar um filtro da média local. Tal como foi visto, o filtro da média pode ser interpretado como o estimador do método dos momentos. Também poderia ser aplicado o estimador de máxima verossimilhança.

Esses estimadores apresentam um bom desempenho somente quando a amostra sobre a qual são aplicados é de observações *puras*. Isto é, quando cada observação provém da mesma distribuição e quando todas são coletivamente independentes. Esta hipótese pode não ser verificada nas imagens reais por vários motivos: o mais evidente é quando se realiza a filtragem em regiões de borda, ou seja, perto da separação entre duas ou mais regiões com valores de retoespalhamento diferentes. Outro motivo possível para a não validade desta hipótese é o aparecimento de correlação entre observações, fato decorrente das propriedades do sistema de imageamento e/ou do tipo de alvo imageado.

Muitos filtros espaciais têm sido desenvolvidos para a redução do *speckle* e para o aumento da relação sinal-ruído, objetivando uma melhoria na imagem final. Estes filtros podem ser divididos em dois grupos: os filtros convolucionais e os filtros pontuais, adaptativos ou não. No primeiro grupo encontram-se:

- Filtro da Média: é o filtro mais simples para a redução do ruído. Consiste em se trocar o nível de cinza de um *pixel* (valor observado) pela média aritmética dos níveis de cinza de uma janela de tamanho $N \times N$, centrada neste *pixel*. É um filtro geral, por não adotar nenhum modelo específico para o ruído.
- Filtro de Frost (57): é um filtro convolucional linear, derivado da minimização do erro médio quadrático sobre o modelo multiplicativo do ruído. Neste filtro incorpora-se a dependência estatística do sinal original, uma vez que se supõe uma função de correlação espacial Exponencial entre os *pixels*. É um filtro adaptativo que preserva a estrutura de bordas.

e no segundo grupo encontram-se:

- Filtro Sigma de Lee (92): teve seu desenvolvimento baseado no fato de que, em muitas imagens degradadas, o ruído tem distribuição gaussiana e é aditivo ao sinal. O valor do *pixel* filtrado resulta da observação dos níveis de cinza numa janela de tamanho $N \times N$. Dessas observações, consideram-se apenas aquelas cujo valor encontra-se dentro do intervalo (estimado) $2\hat{\sigma}$;

calcula-se, então, a média aritmética destas. Trata-se, portanto, de um filtro da média *aparada*.

- Filtro de Lee (91): adota um modelo multiplicativo para o ruído e obedece o critério de *local linear minimum mean square error*. *Local*, porque utiliza estatísticas locais do *pixel* a ser filtrado, admitindo a não estacionaridade da média e da variância do sinal; *linear*, porque realiza uma linearização por expansão em série de Taylor da multiplicação do sinal e o ruído em torno da média, utilizando apenas os termos lineares. O resultado da linearização transforma o modelo multiplicativo do ruído em aditivo, ou seja, o ruído e o sinal tornam-se independentes; e, finalmente, *minimum mean square error*, porque minimiza o erro médio quadrático através do filtro de Wiener. É um filtro adaptativo e geral.
- Filtro de Kuan/Nathan (87,114): adota um modelo multiplicativo. O procedimento é semelhante àquele de Lee, onde a estimação ponto a ponto é feita utilizando-se o filtro de Wiener. A diferença entre eles, entretanto, consiste no fato de que no filtro de Kuan/Nathan não se realiza nenhuma aproximação. É também um filtro adaptativo e geral.
- Filtro MAP para imagens de radar (125): adota um modelo multiplicativo para o ruído. O *pixel* filtrado poderá ser a média local se o coeficiente de variação dos valores observados for menor que o coeficiente de variação do ruído ou ainda poderá ser a solução da equação MAP caso esta solução esteja entre a média e a observação. É um filtro adaptativo e geral, dependendo apenas de sua modelagem.

Além destes filtros, existem na literatura várias outras propostas de filtros que preservam bordas. Ver, por exemplo, o filtro de Nagao e Matsuyama (113) e as referências aí indicadas.

Sempre é considerada a seguinte situação: serão filtrados apenas os *pixels* com um número de vizinhos suficiente; aqueles que não satisfaçam esta propriedade, ficarão com o valor original. Isto é, somente serão filtrados os *pixels*

$S \in S: W_s \subset S$. As imagens original e filtradas serão escritas $y = [y_s]_{s \in S}$ e $\hat{x} = [\hat{x}_s]_{s \in S}$, respectivamente.

Denota-se \mathcal{F} um filtro qualquer; ele é uma função da forma $\mathcal{F}: \Xi^W \rightarrow \Xi$, onde Ξ é o contradomínio de cada *pixel* (por exemplo, nas aplicações aqui apresentadas, tem-se que $\Xi = \{0, \dots, 255\}$) e supõe-se que todos os *pixels* têm o mesmo contradomínio. Acima $W \subset \mathbf{Z}^2$ é denominado o suporte do filtro, ou *janela*, que é da forma $\{-t, \dots, 0, \dots, t\} \times \{-t, \dots, 0, \dots, t\}$, onde se escreve $s = (s_1, s_2)$ e diz-se que o filtro tem tamanho $2t + 1$. Portanto, o filtro usa $v = (2t + 1)^2$ observações para cada *pixel*. Define-se ainda a translação da janela do filtro pelo ponto $r = (r_1, r_2)$ como $W_r = \{s \in \mathbf{Z}^2: s - r \in W\}$.

Finalmente, a operação de filtragem pode ser escrita como:

$$\hat{x}_s = \begin{cases} \sqrt{\pi/2} \mathcal{F}(y_{W_s}) & \text{se } W_s \subset S \\ y_s & \text{caso contrário.} \end{cases}$$

Caso haja necessidade de se filtrar todos os *pixels* da imagem, basta que se redefina o suporte do filtro para cada caso. Preferiu-se não fazer isto por simplicidade computacional, e para usar as bordas (observações não processadas) como uma forma rápida de comparação da imagem filtrada com a imagem original.

Poderiam ser consideradas outras formas de janela, por exemplo as W derivadas da idéia de *norma-L*: $W = \{s \in S: \|s\|_L \leq d\}$, onde $s = (s_1, s_2)$, $d \in \mathbb{N}$ e:

$$\|s\|_L = \begin{cases} |s_1| + |s_2| & \text{se } L = 1, \\ \sqrt{s_1^2 + s_2^2} & \text{se } L = 2, \text{ e} \\ \text{máx}\{|s_1|, |s_2|\} & \text{se } L = \infty. \end{cases}$$

Uma pequena modificação deve ser introduzida nas definições acima para que os filtros preservem o contradomínio Ξ (um conjunto discreto) da imagem. Considerar-se-á como *valor filtrado* ao inteiro mais próximo do resultado da operação de filtragem, i. e., $[\frac{1}{2} + \sqrt{\frac{\pi}{2}} \mathcal{F}(y_{W_s})]$.

Ainda, se todas as observações dentro da janela forem iguais, anulando o valor dos estimadores MAD e IQR, considerar-se-á como *valor filtrado*

a própria observação.

Como deseja-se manter o nível médio de cinza na imagem, e este se relaciona com o parâmetro ξ através da equação (5.9), na implementação dos filtros cada resultado é multiplicado pelo fator $\sqrt{\pi/2}$. Fica claro que $\hat{x} = \mathcal{F}(y)$ é um estimador da imagem não observada $x = \sqrt{\pi/2} [\xi_s]_{s \in S}$ baseado na imagem observada y , e justifica-se assim a notação empregada.

A seguir pode ser visto um exemplo da aplicação desta técnica a uma imagem SAR aerotransportado obtida pelo sensor SAR-580 (Banda L), sobre a região de Freiburg, Alemanha, com uma visada e detecção linear.

A Figura 5.6 é um pedaço de 128×128 *pixels* da imagem original, apresentando o ruído *speckle*, típico das imagens SAR. Nele observam-se duas regiões homogêneas: uma clara, de floresta e uma escura, de solo exposto.

Resta, antes de aplicar os filtros propostos, verificar se a hipótese de distribuição Rayleigh é satisfeita pelos dados a serem tratados. Na Tabela 5.7 mostram-se algumas quantidades estimadas sobre amostras homogêneas da imagem apresentada na Figura 5.6. Daí conclui-se que, embora as médias para as regiões clara e escura sejam bastante diferentes, o número equivalente de visadas (NEV), definido na pág. 161, é aproximadamente 1 para ambas. Na Tabela 5.8 mostram-se os p -valores percentuais dos testes de Kolmogorov-Smirnov e χ^2 para o ajuste dos dados às Distribuições 2.3, 2.5, 2.6 e 2.8. Daí conclui-se que não existe evidência para se rejeitar a hipótese de Distribuição Rayleigh para os dados.

TABELA 5.7 - QUANTIDADES ESTIMADAS

Amostra	N	\bar{x}	s	$100\hat{C}$	NEV
escura	3600	36.50	18.40	50.41	1.08
clara	2444	97.45	51.39	52.74	0.98

TABELA 5.8 - p -VALORES PERCENTUAIS DOS TESTES KOLMOGOROV-SMIRNOV E χ^2

Amostra	\mathcal{R}		$\mathcal{KA1}$		\mathcal{N}		\mathcal{LN}		β	
	p_{KS}	p_{χ^2}	p_{KS}	p_{χ^2}	p_{KS}	p_{χ^2}	p_{KS}	p_{χ^2}	p_{KS}	p_{χ^2}
escura	2.28	26.53	2.28	26.53	0.003	0.000	0.000	0.000	0.324	31.830
clara	80.93	48.12	0.00	0.00	0.001	0.000	0.000	0.000	15.937	1.130

As Figuras 5.8, 5.9, 5.10, 5.11, 5.12, 5.13, 5.14 são, respectivamente, as imagens filtradas pelos estimadores MV, MO, TMV, TMO, MAD, IQR e Med. Todas as imagens foram geradas se empregando um filtro de tamanho $t = 5$, e os algoritmos aparados usaram $\alpha_0 = 0.225$.

Para poder medir o efeito redutor de ruído, foi empregado \widehat{C}^{-1} , a recíproca do coeficiente de variação sobre regiões homogêneas. Para se verificar o efeito *normalizador* dos filtros foram estimadas a assimetria e kurtose sobre as mesmas áreas. Os resultados, arredondados à terceira casa decimal, são mostrados nas Tabelas 5.9 para um total de 3500 *pixels* e 5.10 para um total de 2679 *pixels*.

TABELA 5.9 - VALORES ESTIMADOS SOBRE A REGIÃO ESCURA

	\widehat{C}^{-1}	Assimetria	Kurtose
Teórico	1.913	0.631	3.245
Original	1.965	0.665	0.276
MV	4.060	1.593	8.366
MO	4.057	1.419	7.061
TMV	3.881	1.688	8.206
TMO	4.018	1.201	5.296
MAD	3.088	0.914	2.061
IQR	3.082	1.103	2.364
Med	3.881	0.960	3.429

FONTE: Adaptada de (56), p. 172.

TABELA 5.10 - VALORES ESTIMADOS SOBRE A
REGIÃO CLARA

	\widehat{C}^{-1}	Assimetria	Kurtose
Teórico	1.913	0.631	3.245
Original	1.879	0.573	-0.155
MV	4.494	0.683	4.100
MO	4.424	0.660	3.472
TMV	4.140	0.903	3.820
TMO	4.199	0.595	2.135
MAD	3.155	0.627	0.652
IQR	3.185	0.611	0.579
Med	3.971	0.491	1.100

FONTE: Adaptada de (56), p. 172.

O melhor algoritmo em relação ao critério dado pelo valor de \widehat{C}^{-1} é o filtro MV: ele fornece um aumento de aproximadamente 107% e 139% nas regiões escura e clara, respectivamente. O pior algoritmo, em relação ao mesmo critério, é o MAD pois produz aumentos de, aproximadamente, 57% e 68% nas mesmas áreas. Assim, a diferença entre o melhor e o pior não é muito grande, sendo que o borramento introduzido pelas técnicas robustas é consideravelmente menor que o introduzido pelos filtros MV e MO.

A Figura 5.15 mostra os histogramas das imagens original e filtradas pelos métodos apresentados. É evidente que, após a filtragem, aparece uma separação entre as duas classes (os *vales* nos histogramas) que não era perceptível na imagem original. Assim, verifica-se uma outra aplicação destas técnicas: além de melhorar a qualidade visual das imagens, aumentam a separabilidade entre classes.

5.8.1 Extensões

Pretende-se dar continuidade a este assunto pelo uso de estimadores consistentes, em vez de estimadores *assintoticamente* consistentes. Tam-

bém pretende-se implementar o filtro derivado do estimador “BLUE”. Com a metodologia apresentada em (125) será possível avaliar o desempenho destes filtros em relação a vários aspectos: redução de ruído, preservação de bordas, etc.

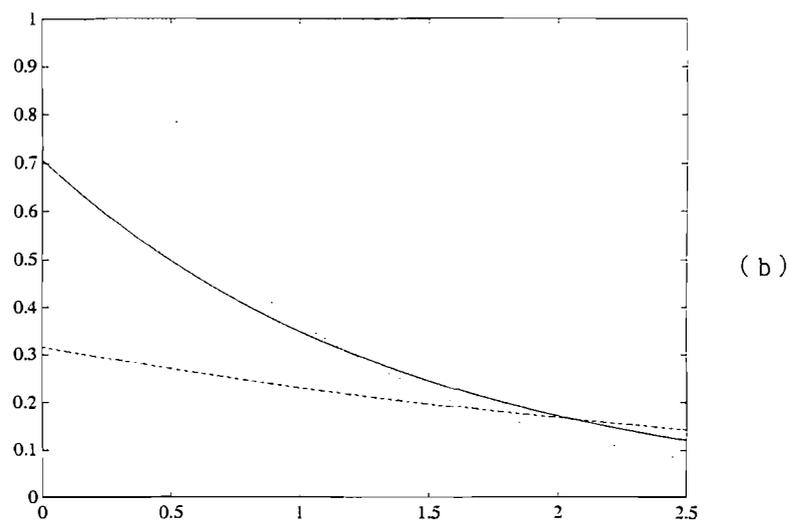
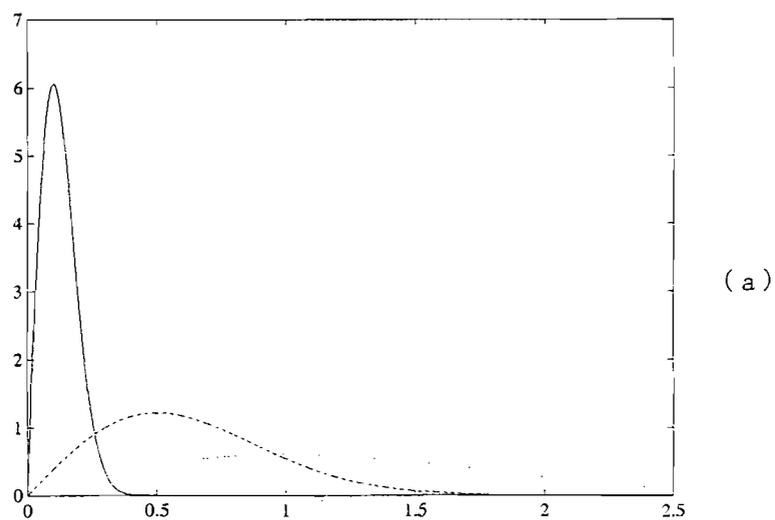


Fig. 5.2 - (a) Função de densidade de probabilidade Rayleigh para vários valores do parâmetro. (b) Função de densidade de probabilidade Exponencial para vários valores do parâmetro.

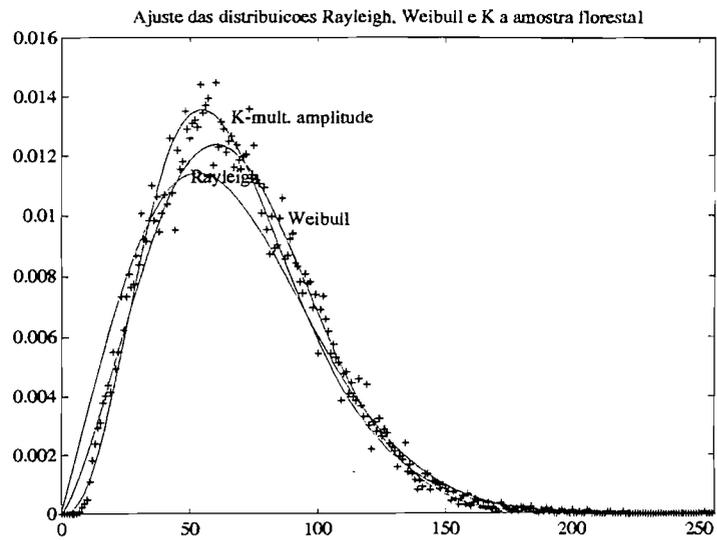


Fig. 5.3 - Ajuste das distribuições Rayleigh, Weibull e $\mathcal{KA6}$ a uma amostra florestal.

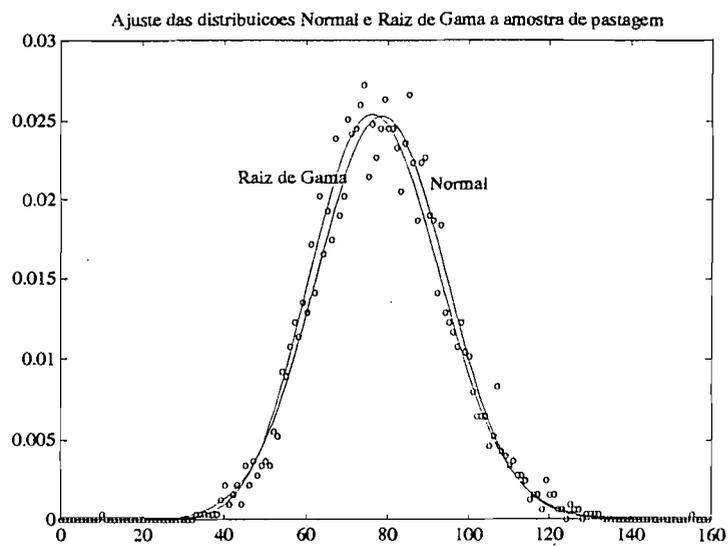


Fig. 5.4 - Ajuste das distribuições Normal e Raiz de Gamma a uma amostra de pastagem.

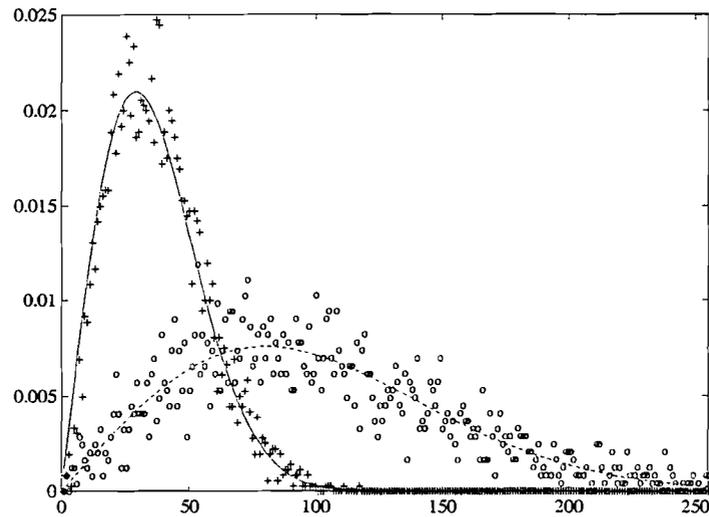


Fig. 5.5 - Histogramas (“+” e “o”) e densidades Rayleigh com parâmetros estimados (linhas contínua e tracejada) para as regiões escura e clara, respectivamente: $\xi = 28.90$ e $\xi = 77.90$ respectivamente.
 FONTE: (56), p. 167.

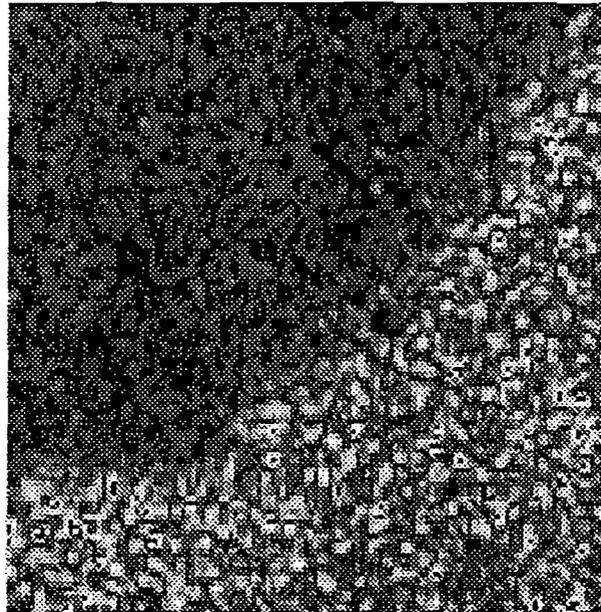


Fig. 5.6 - Imagem SAR-580, banda L.
 FONTE: (56), p. 167.

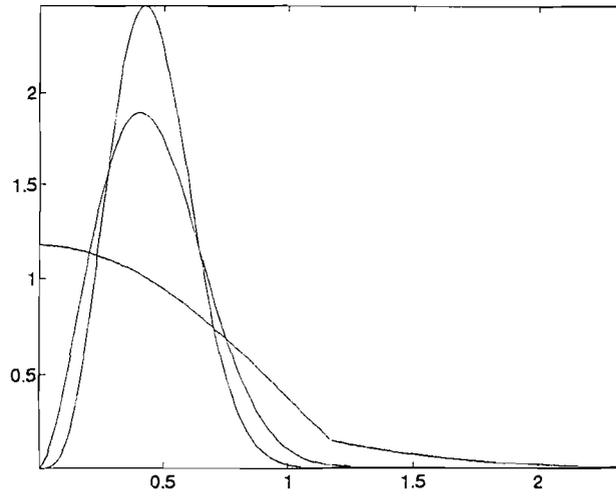


Fig. 5.7 - Densidades do estimador MAD para amostras de tamanho um, cinco e nove.

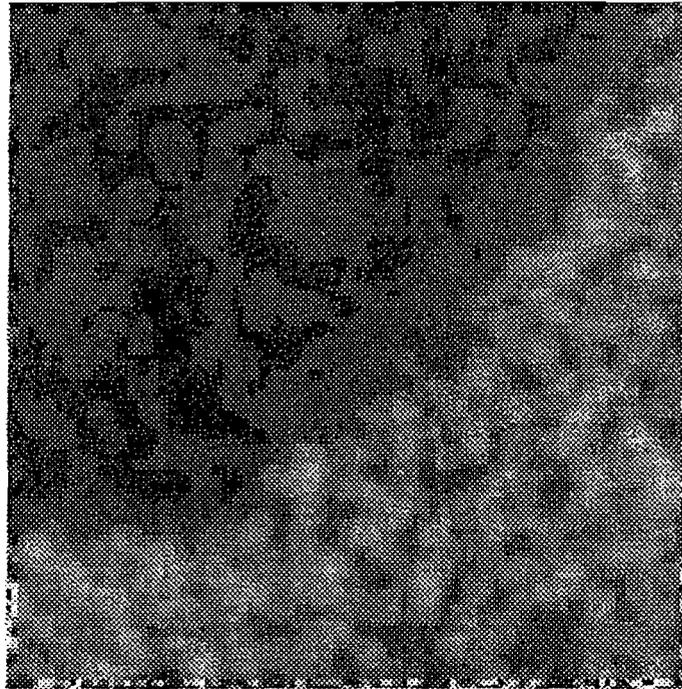


Fig. 5.8 - Imagem filtrada pelo estimador MV.

FONTE: (56), p. 170.

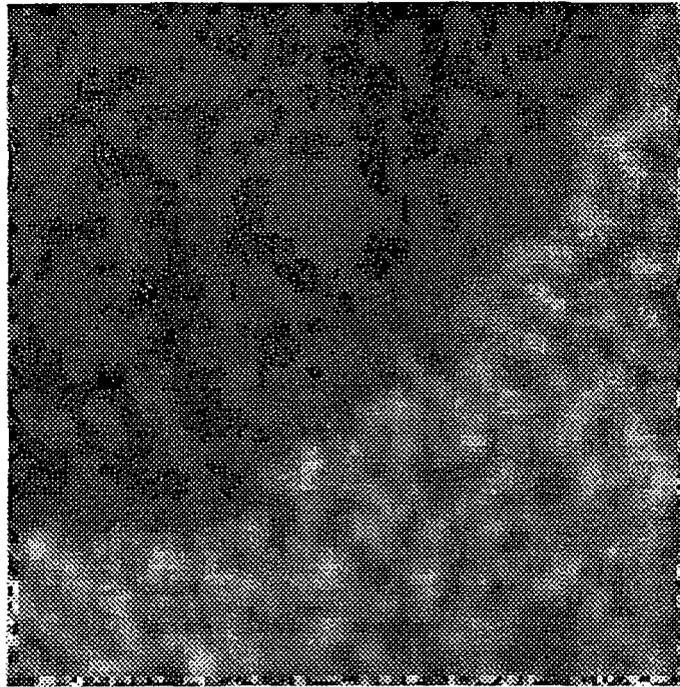


Fig. 5.9 - Imagem filtrada pelo estimador MO.
FONTE: (56), p. 170.

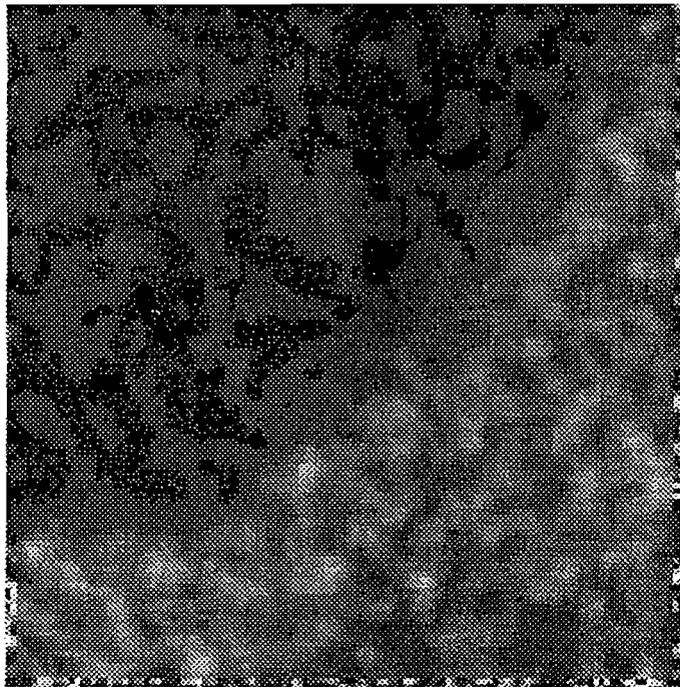


Fig. 5.10 - Imagem filtrada pelo estimador TMV.
FONTE: (56), p. 167.

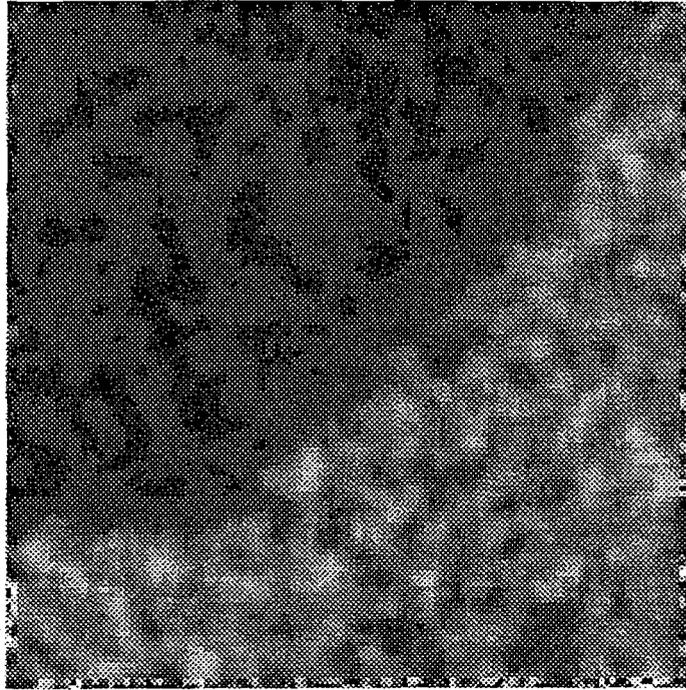


Fig. 5.11 - Imagem filtrada pelo estimador TMO.

FONTE: (56), p. 171.

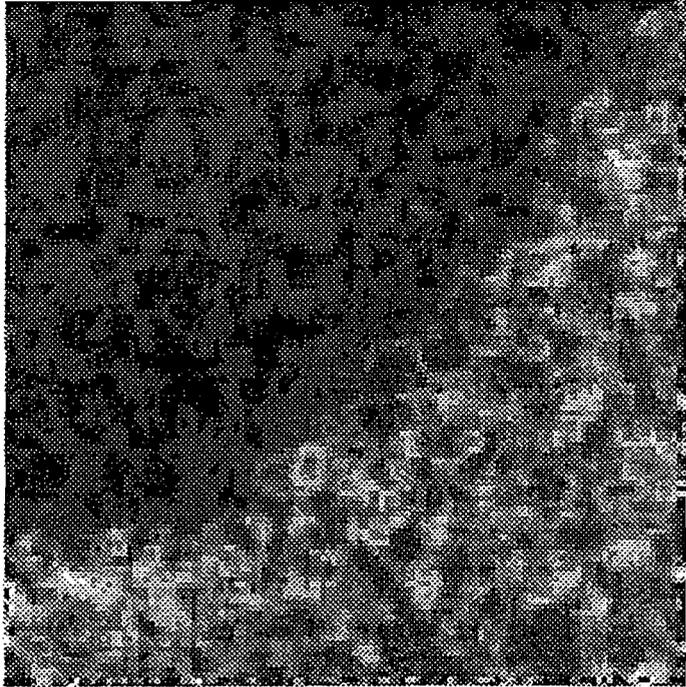


Fig. 5.12 - Imagem filtrada pelo estimador MAD.

FONTE: (56), p. 171.

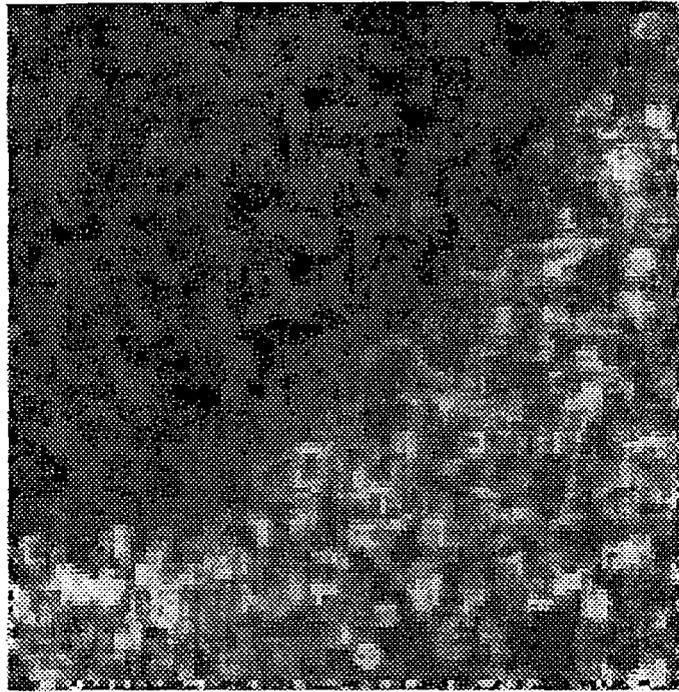


Fig. 5.13 - Imagem filtrada pelo estimador IQR.

FONTE: (56), p. 171.

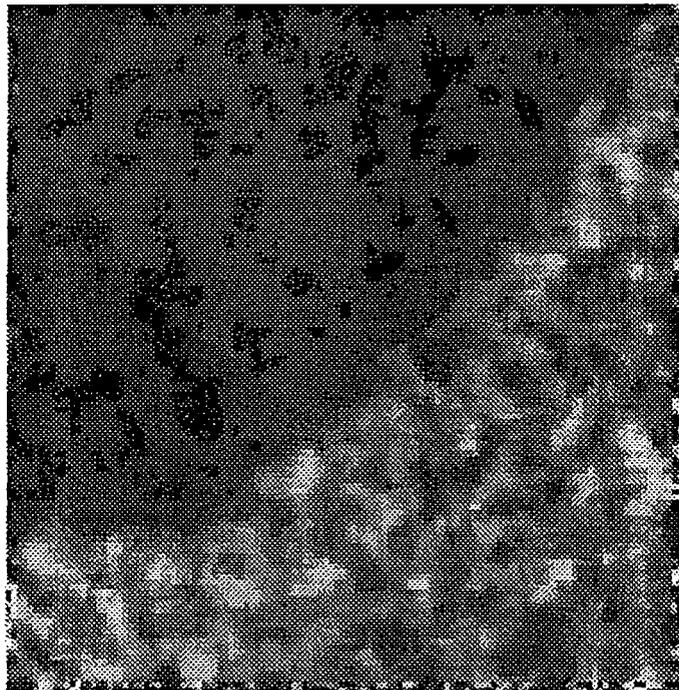


Fig. 5.14 - Imagem filtrada pelo estimador Med.

FONTE: (56), p. 172.

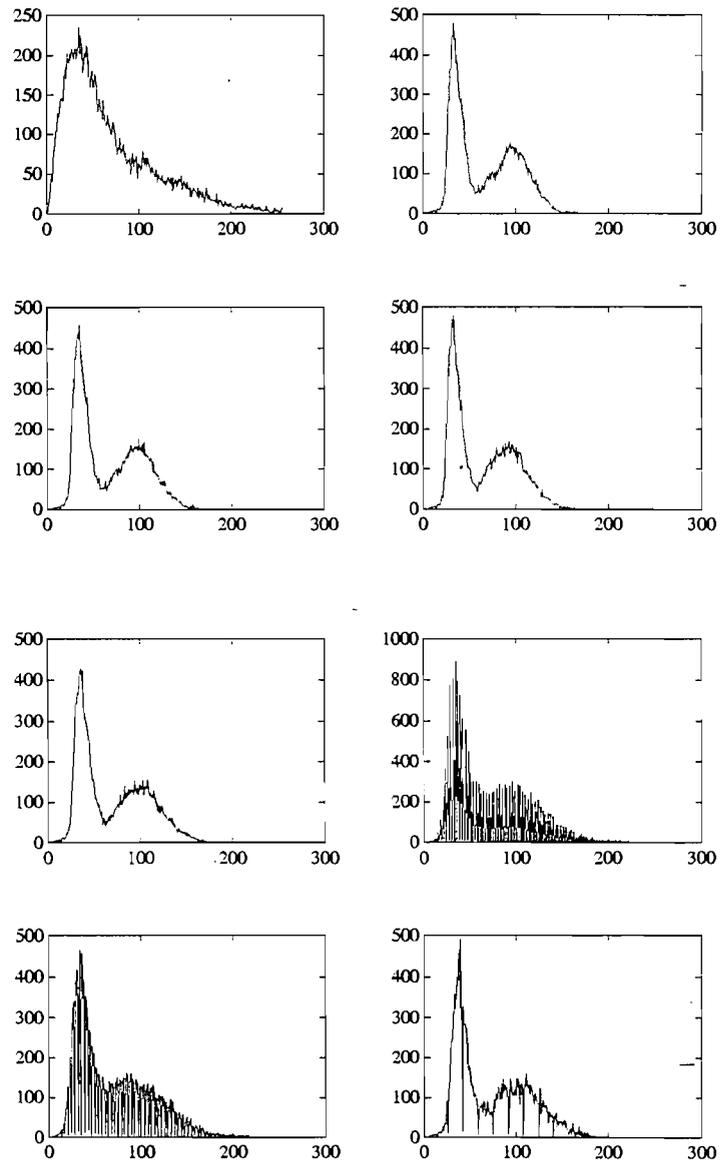


Fig. 5.15 - Histogramas, de esquerda para direita e de cima para abaixo, das imagens original e filtradas pelos algoritmos MV, MO, TMV, TMO, MAD, IQR e Med.

FONTE: (56), p .174.

CAPÍTULO 6

APLICAÇÃO DE ANÁLISE DE ESTATÍSTICA DE DADOS A IMAGENS DA MISSÃO SAREX SOBRE A RESERVA FLORESTAL DE TAPAJÓS

6.1 Introdução

Em abril de 1992, durante uma missão conjunta Canadá-Brasil (Missão SAREX) foi sobrevoada, entre outras áreas, uma reserva florestal na região de Tapajós pelo sistema SAR-580 pertencente ao Centro Canadense de Sensoriamento Remoto (CCRS). Imagens polarimétricas HH e VV, banda C, foram adquiridas com o objetivo de avaliar o uso e utilidade dessas imagens na tarefa de classificação dos tipos de florestas que se possam definir na região amazônica e também o mapeamento de áreas urbanas e desflorestadas (41,42).

Os pontos que o grupo de processamento de imagens de radar de abertura sintética da Divisão de Processamento de Imagens do INPE pretende estudar, usando as imagens da missão SAREX, são:

- 1) Estatísticas da Imagem: o fato das florestas se constituírem de alvos extensos tornam as mesmas áreas propícias para o estudo do comportamento estatístico do *speckle*, da estrutura de covariância das bandas, e de quais modelos estatísticos melhor se adaptam às texturas das florestas. Esses estudos são importantes para a definição dos classificadores mais adequados para esse tipo de imagens, principalmente quando se quer levar a textura em consideração. Mais adiante discute-se como a hipótese de Gaussianidade dos dados não é a mais adequada para estas imagens, fazendo com que novos algoritmos devam ser implementados. Deseja-se saber se os comportamentos estatísticos observados correspondem aos modelos divulgados na literatura (98,142).

- 2) Classificação dos tipos de florestas: um primeiro passo consiste na identificação dos tipos de classes que serão utilizadas. Para isso é necessário identificar quais classes são separáveis usando-se os dados disponíveis. Dois métodos serão utilizados: o visual e o automático. No método visual são identificadas regiões com aparência distinta. Tal identificação pode, eventualmente, ser auxiliada por computador para o realce pictórico (44), a partir de fotos preto e branco e coloridas. O método automático de agrupamento ISODATA é utilizado para descobrir agrupamentos naturais nos dados, sendo o resultado comparado com o método visual. Visitas ao campo podem ser necessárias, nessa fase, para verificar se o que se distingue na imagem realmente corresponde a alvos distintos no campo. Uma vez fixadas as classes, é possível realizar a classificação supervisionada sobre as imagens utilizando-se os seguintes métodos:
- (a) Classificação pontual: as imagens são pré-filtradas por filtros de redução de speckle, com a finalidade de diminuir a quantidade de ruído e tornar gaussiana a distribuição dos pixels. Usa-se em seguida a classificação de máxima verossimilhança e filtros de pós-processamento para a diminuição do ruído de classificação.
 - (b) Segmentação e classificação de amostras: Dois métodos de segmentação serão utilizados (104): o primeiro método é baseado em detecção de bordas e o segundo é do tipo crescimento de regiões. Uma vez obtidas as regiões, estas são classificadas usando-se métodos de similaridade estatística (43). Estes métodos, a serem desenvolvidos, deverão levar em consideração as correlações presentes neste tipo de dados (139).
- 3) Comparação com outros sensores: os procedimentos apontados no item (b) serão repetidos comparando-se com dados de outros sensores de radar tais como imagens ERS-1 (já adquiridas), JERS-1 e Almaz II se disponíveis. Usar-se-ão também imagens ópticas para a comparação, corregistradas ou não com as imagens do SAR.

Com os resultados dessa análise pretende-se compreender melhor o processo de geração das imagens de radar e os mecanismos de interação de microondas com o alvo do tipo floresta tropical úmida nos diversos tipos de comprimento de onda e polarizações. A qualidade desse tipo de dados e os meios de melhor utilizá-los serão investigados.

O desempenho e os resultados obtidos com os algoritmos implementados (42,46,54,55,56) mostraram que é relevante o desenvolvimento de ferramentas específicas, projetadas para os problemas em questão.

Neste Capítulo apresenta-se o desenvolvimento de uma ferramenta computacional para o teste de aderência, a diversas distribuições e por dois critérios, dos dados SAREX colhidos sobre a região de Tapajós, com o intuito de discriminar diferentes ocupações do solo. Comentam-se os problemas encontrados para a implementação desta ferramenta, e os devidos ao formato dos dados.

6.2 Análise Preliminar

Foi feita uma segmentação visual (140) da imagem com polarização HH, procurando-se discriminar regiões de ocupação homogênea do solo: floresta e não floresta. As Tabelas 6.1 e 6.2 mostram os valores de: tamanho de amostra considerada (N), média amostral da amplitude (\bar{y}), desvio padrão amostral da amplitude (s), coeficiente de variação amostral vezes cem ($100\hat{C}$) e número equivalente amostral de visadas (NEV). Esta última quantidade é calculada pela relação:

$$NEV = \sqrt{\frac{4 - \pi}{\pi}} \left(\frac{\bar{y}}{s}\right)^2. \quad (6.1)$$

O valor de NEV empregado para a análise das amostras ($NEV = 6$) foi estimado pelo arredondamento, ao inteiro mais próximo, da média dos NEVs obtidos sobre as amostras de não floresta (ver Tabela 6.2).

A razão de serem utilizados somente amostras de não floresta para a estimação do NEV é que essas amostras são, supostamente, de regiões onde o retroespalhamento é constante. Para amostras deste tipo, os momentos da dis-

tribuição marginal dos dados são conhecidos (equação (2.20)), a partir dos quais é obtida a equação (6.1).

Das Tabelas 6.1 e 6.2 pode concluir-se que o coeficiente de variação poderia ser empregado na discriminação de ocupação do solo.

6.3 Testes de Qualidade de Ajuste

No decorrer deste Capítulo supor-se-á que $\mathbf{Y} = (Y_1, \dots, Y_N)$ é um vetor de variáveis aleatórias independentes e identicamente distribuídas, e $\mathbf{y} = (y_1, \dots, y_N)$ é o vetor amostra de tamanho N , ocorrência de \mathbf{Y} . Seja H_0 , ou hipótese nula, a hipótese de que $Y_i \sim \mathcal{D}(\boldsymbol{\theta})$, com $\boldsymbol{\theta} \in \mathbb{R}^\ell$ o vetor de parâmetros que caracterizam a distribuição \mathcal{D} . Suponha-se conhecido ou estimado o vetor $\boldsymbol{\theta}$, e conhecidas a densidade e a função de distribuição acumulada que caracterizam $\mathcal{D}(\boldsymbol{\theta})$: $f: \mathbb{R} \rightarrow \mathbb{R}_+$ e $F: \mathbb{R} \rightarrow [0, 1]$, respectivamente.

Seja $t: \mathbb{R}^N \rightarrow \mathbb{R}$ uma função, $p_t(y_1, \dots, y_N)$ (ou simplesmente p_t) será o p -valor da estatística $t(Y_1, \dots, Y_N)$ sob H_0 e tal que o teste rejeita a hipótese nula para valores *grandes* de t (existem testes de outras formas), isto é:

$$p_t(y_1, \dots, y_N) = \Pr(t(Y_1, \dots, Y_N) \geq t(y_1, \dots, y_N)).$$

Lembre-se que se $p_t(y_1, \dots, y_N) = \alpha$, então diz-se que o teste de H_0 definido por t rejeita H_0 ao nível α ou, equivalentemente, “a probabilidade de errar ao rejeitar H_0 é menor ou igual que $p_t(y_1, \dots, y_N)$ ”. Daí que rejeita-se H_0 se p_t é “pequeno”, e não rejeita-se H_0 se p_t é “grande” (valores típicos para se considerar p_t pequeno são $p_t < 0.01$ ou $p_t < 0.05$).

TABELA 6.1 - QUANTIDADES ESTIMADAS NAS AMOSTRAS DE FLORESTA.

Amostra	N	\bar{y}	s	$100\hat{C}$	NEV
flor10A	21166	73.60	34.40	46.74	1.25
flor1A	5023	65.11	27.72	42.57	1.51
flor1Ai	7442	70.12	31.25	44.56	1.38
flor1Aii	9024	81.90	29.37	35.86	2.13
flor1B	30056	65.40	32.92	50.33	1.08
flor1Bi	9600	70.17	35.05	49.95	1.10
flor1Bii	11319	70.22	33.45	47.63	1.20
flor1C	6912	71.89	33.17	46.14	1.28
flor1Ci	5888	81.09	35.20	43.40	1.45
flor1Cii	7396	83.29	40.15	48.20	1.18
flor1_1Ai	960	72.20	28.04	38.84	1.81
flor1_1Aii	837	84.16	25.80	30.65	2.91
flor1_1C	378	85.69	27.05	31.57	2.74
flor1_1Cii	280	90.27	28.08	31.11	2.82
flor1_2Ai	627	82.46	28.43	34.48	2.30
flor2A	15600	71.25	26.87	37.72	1.92
flor2Ai	12420	80.27	37.16	46.29	1.28
flor2Aii	17280	66.12	32.06	48.49	1.16
flor2B	11049	67.40	35.81	53.13	0.97
flor2Bi	8800	74.60	37.14	49.79	1.10
flor2Bii	7448	81.12	44.70	55.10	0.90
flor2C	10488	74.17	40.23	54.24	0.93
flor2Ci	1710	83.95	37.04	44.12	1.40
flor2Cii	8137	75.90	37.31	49.16	1.13

(continua)

TABELA 6.1 - Continuação.

Amostra	N	\bar{y}	s	$100\hat{C}$	NEV
flor2_1Ai	644	82.23	28.23	34.32	2.32
flor2_1Bi	486	84.07	32.43	38.57	1.84
flor2_2Ai	513	105.73	35.11	33.21	2.48
flor3A	2597	72.34	25.95	35.87	2.12
flor3Ai	12384	78.86	29.95	37.97	1.89
flor3B	7084	85.72	45.47	53.04	0.97
flor3Bi	10230	75.99	45.94	60.45	0.75
flor3Bii	3000	99.22	47.53	47.91	1.19
flor3C	12483	74.68	39.30	52.63	0.99
flor3Ci	10989	73.47	36.33	49.46	1.12
flor3Cii	15805	77.73	34.34	44.18	1.40
flor3_1A	1345	74.74	27.81	37.20	1.97
flor3_1Ai	957	89.64	30.21	33.70	2.41
flor3_1B	442	102.06	38.31	37.54	1.94
flor3_1Bi	414	102.86	40.06	38.95	1.80
flor3_1Bii	522	139.91	43.22	30.89	2.86
flor3_1Cii	700	82.52	25.77	31.23	2.80
flor3_2Ai	456	83.71	26.16	31.26	2.80
flor4A	2593	64.44	25.94	40.26	1.69
flor4Ai	19152	67.57	31.45	46.55	1.26
flor4B	7313	82.23	34.45	41.89	1.56
flor4Bi	7526	74.67	39.76	53.26	0.96
flor4Bii	6901	82.07	35.38	43.10	1.47
flor4C	8640	74.79	40.91	54.71	0.91

(continua)

TABELA 6.1 - Conclusão.

Amostra	N	\bar{y}	s	$100\hat{C}$	NEV
flor4Cii	17334	76.75	42.47	55.34	0.89
flor4_1B	616	86.24	30.25	35.08	2.22
flor4_1Bii	700	96.15	30.40	31.62	2.73
flor5A	3405	75.40	33.43	44.34	1.39
flor5Ai	19720	67.37	34.08	50.59	1.07
flor5B	10318	87.87	37.63	42.82	1.49
flor5Bi	11400	82.22	37.78	45.96	1.29
flor5Bii	13632	74.32	38.93	52.39	1.00
flor5_1Ai	408	79.68	27.05	33.95	2.37
flor5_1B	600	97.58	31.48	32.26	2.63
flor5_2Ai	494	75.00	28.04	37.39	1.95
flor6A	2755	62.27	28.75	46.17	1.28
flor6Ai	14560	75.11	40.99	54.58	0.92
flor6B	17745	71.10	34.81	48.95	1.14
flor6Bi	11514	69.61	36.66	52.67	0.99
flor7A	1281	71.26	28.12	39.47	1.75
flor7Ai	6075	81.70	41.30	50.55	1.07
flor7_1A	961	74.86	29.25	39.08	1.79
flor7_1Ai	792	96.49	39.79	41.24	1.61
flor7_TA	12446	68.43	30.78	44.97	1.35
flor8A	1190	77.49	32.57	42.03	1.55
flor9A	465	92.34	31.65	34.27	2.33

TABELA 6.2 - QUANTIDADES ESTIMADAS NAS AMOSTRAS DE NÃO FLORESTA.

Amostra	N	\bar{y}	s	$100\hat{C}$	NEV
past0B	341	79.27	17.57	22.16	5.56
past0aB	415	87.66	20.34	23.21	5.07
past11Ci	1915	86.70	19.15	22.09	5.60
past12Ci	4561	75.91	15.91	20.96	6.22
past13Ci	6931	74.09	16.28	21.98	5.66
past1B	1185	68.50	15.44	22.54	5.38
past1Bi	2245	68.93	15.61	22.65	5.33
past1Bii	3240	83.49	17.90	21.44	5.94
past1C	6328	72.93	16.66	22.85	5.23
past1Ci	1801	73.52	16.58	22.56	5.37
past1_1B	217	78.77	18.68	23.71	4.86
past2B	4051	77.24	19.46	25.19	4.31
past2Bi	1756	69.36	15.27	22.02	5.63
past2Bii	1821	84.77	17.85	21.05	6.17
past2C	1702	71.82	15.45	21.52	5.90
past2Ci	1291	80.99	17.48	21.58	5.86
past2_1B	1249	74.56	19.06	25.56	4.18
past2_2B	631	78.60	18.46	23.48	4.96

(continua)

TABELA 6.2 - Conclusão.

Amostra	N	\bar{y}	s	$100\hat{C}$	NEV
past3B	2117	77.86	18.22	23.40	4.99
past3Bi	2441	76.84	16.32	21.24	6.06
past3Bii	1750	71.34	15.10	21.17	6.10
past3C	3627	82.43	19.40	23.54	4.93
past3Ci	1240	72.83	15.49	21.27	6.04
past4B	3479	78.52	18.23	23.21	5.07
past4Bi	2458	63.99	12.86	20.10	6.76
past4Bii	4641	69.71	14.54	20.86	6.28
past4Ci	2458	63.28	12.90	20.39	6.57
past5B	2302	72.48	17.69	24.41	4.59
past5Bi	2689	78.26	18.30	23.39	4.99
past5Bii	3589	82.00	18.16	22.15	5.57
past5Ci	3268	78.51	15.77	20.09	6.77
past6B	2807	75.19	16.00	21.28	6.04
past6Ci	2577	77.40	17.55	22.67	5.32
past7B	5221	76.48	16.47	21.54	5.89
past7Ci	2729	79.32	17.09	21.54	5.89
past8Ci	3781	75.40	15.87	21.05	6.17
past9Ci	2022	69.69	16.08	23.08	5.13

Note-se que para calcular p_t é necessário conhecer a distribuição de $t(Y_1, \dots, Y_N)$, coisa que nem sempre é possível; todavia, pode-se substituir a distribuição exata de $t(Y_1, \dots, Y_N)$ pela sua distribuição assintótica (se ela existir) ou lançar-se mão de testes Monte Carlo (23). Do ponto de vista prático, este fato não compromete o desenvolvimento de uma ferramenta computacional, pois na maioria das vezes o valor de p_t pode obter-se de tabelas adequadas e, freqüentemente, a mesma sub-rotina que calcula $t(u_1, \dots, u_N)$ fornece o valor p_t . Assim acontece com as implementações em pacotes como IMSL e NAG.

6.3.1 Teste χ^2

Seja $2 \leq k < N$ o número de intervalos em que será particionada a amostra, e sejam esses intervalos denotados C_j :

$$\begin{aligned} C_1 &= (c_0 = -\infty, c_1] \\ &\vdots \\ C_{k-1} &= (c_{k-2}, c_{k-1}] \\ C_k &= (c_{k-1}, c_k = +\infty). \end{aligned}$$

Para cada $i = 1, \dots, k$ sejam as proporções observadas no intervalo C_i e a probabilidade no intervalo C_i dadas, respectivamente, por:

$$f_i = \frac{1}{N} \# \left\{ 1 \leq j \leq N : y_j \in C_i \right\}, \text{ e } I_i = \int_{C_i} f(x) dx.$$

Seja agora a estatística do teste dada por:

$$t = \sum_{1 \leq i \leq k} \frac{(f_i - I_i)^2}{I_i}.$$

Tem-se que $p_t \approx \text{IPr}(Y \geq t)$, com $Y \sim \chi_{k-1}^2$ (denota-se “aproximadamente igual” com “ \approx ”). Os valores de k e de N devem, pelo menos em princípio, respeitar uma certa relação; por exemplo, se $N = 1000$ então $k = 59$ (ver a referência (69) para mais detalhes).

Ainda mais, devido a considerações práticas, é desejável que sejam satisfeitos os seguintes critérios para todo $i = 1, \dots, k$: $Nf_i \geq 5$, $NI_i \geq 1$, e $f_i \approx f_j$ para todo $i \neq j$.

6.3.2 Teste de Kolmogorov-Smirnov

Seja $(y_{(1:N)}, \dots, y_{(N:N)})$ o vetor (y_1, \dots, y_N) com os elementos permutados em forma tal que: $y_{(1:N)} \leq \dots \leq y_{(N:N)}$. Por definição, sejam $y_{(0:N)} = 0$ e $y_{(N+1:N)} = 1$. Seja $\hat{F}: \mathbb{R} \rightarrow [0, 1]$ dada por:

$$\hat{F}(x) = \frac{i}{N} \text{ se } y_{(i:N)} \leq x < y_{(i+1:N)}, \quad i = 0, 1, \dots, N$$

e seja $t = \max\{|\hat{F}(x) - F(x)|: x \in [0, 1]\}$.

Pode ver-se, por exemplo em (63) ou na documentação da subrotina correspondente (G08CAF no NAG (112)) que, para N “grande” e sob H_0 , $p_t \approx 1 - L(\sqrt{N}t)$, onde:

$$L(z) = \begin{cases} 0 & \text{se } z \leq 0.27 \\ \frac{\sqrt{2\pi}}{z} \sum_{1 \leq k \leq 3} \exp\left\{-\frac{(2k-1)^2 \pi^2}{8z^2}\right\} & \text{se } 0.27 < z < 1 \\ 1 - 2 \sum_{1 \leq k \leq 4} (-1)^{k-1} \exp\{-2k^2 z^2\} & \text{se } 1 \leq z < 3.1 \\ 1 & \text{se } z \geq 3.1. \end{cases}$$

6.4 Discriminação da Ocupação do Solo

Os testes Kolmogorov-Smirnov e χ^2 foram aplicados a amostras de floresta e de não floresta numa imagem SAREX sobre a reserva florestal de Tapajós. Essas amostras foram escolhidas por meio de uma classificação visual da imagem, auxiliada por uma imagem LANDSAT da mesma área. As mesmas áreas foram previamente analisadas nas Tabelas 6.1 e 6.2.

As distribuições ajustadas foram: Raiz de Gamma (2.15), Gamma (2.14), \mathcal{K} de amplitude e 6 visadas (2.12), Normal (2.3), Log-Normal (2.5), Weibull (2.7) e Beta (2.6), já apresentadas no Capítulo 2. O número de visadas empregado foi estimado se arredondando ao inteiro mais próximo a média da quarta coluna da Tabela 6.2: a média foi 5.89 e, portanto, o número equivalente de visadas empregado foi $[5.89] = 6$.

Nas Tabelas 6.3 e 6.4 mostram-se as distribuições que melhor se ajustaram aos dados pelos quatro seguintes critérios:

$\text{máx}p_{KS}$: o melhor ajuste é atribuído àquela distribuição cujo p -valor é máximo para o teste Kolmogorov-Smirnov;

$\text{máx}p_{\chi^2}$: o melhor ajuste é atribuído àquela distribuição cujo p -valor é máximo para o teste χ^2 ;

$\text{máx}\{p_{KS}, p_{\chi^2}\}$: o melhor ajuste é atribuído àquela distribuição cujo p -valor é máximo para ambos os testes;

$\text{máx}(p_{KS} + p_{\chi^2})/2$: o melhor ajuste é atribuído àquela distribuição cujo p -valor é máximo para a média dos testes Komogorov-Smirnov e χ^2 .

Para construir as Tabelas 6.3 e 6.4 foi, previamente, construída uma tabela para cada tipo de amostra onde, para cada distribuição, tem-se o p -valor de cada teste. Esta tabela não é mostrada devido ao seu tamanho e a que não é simples, a partir dela, inferir os resultados resumidos nas Tabelas 6.3 e 6.4.

Em cada uma das quatro colunas, correspondentes aos quatro critérios mencionados, escreve-se a distribuição escolhida, seguindo a notação já empregada, os parâmetros estimados e, embaixo, o p -valor correspondente multiplicado por cem. Toda vez que o máximo de algum dos quatro critérios acima mencionados foi zero, não foi possível escolher nenhuma distribuição; nesse caso, foi escrito “—”.

Para se facilitar a interpretação dos resultados apresentados nas Tabelas 6.3 e 6.4 foram construídas as Tabelas 6.5 e 6.6, um *resumo* das anteriores. Nelas é mostrado o número de amostras atribuído pelos quatro critérios empregados às distribuições consideradas.

TABELA 6.3 - DISTRIBUIÇÕES ESCOLHIDAS POR QUATRO CRITÉRIOS
PARA AMOSTRAS DE FLORESTA

Amostra	$\text{máx}_{p_{KS}}$	$\text{máx}_{p_{\chi^2}}$	$\text{máx}\{p_{KS}, p_{\chi^2}\}$	$\text{máx}(p_{KS} + p_{\chi^2})/2$
flor1A0Ai.i	$\sqrt{\Gamma}(1.24, 0.000187)$ 1.40	$\mathcal{KA6}(1.58, 6599.98)$ 0.006	$\sqrt{\Gamma}(1.24, 0.000187)$ 1.40	$\sqrt{\Gamma}(1.24, 0.000187)$ 0.938
flor1Ai.i	$\beta(3.85, 11.24)$ 24.92	$\mathcal{KA6}(1.98, 50007.94)$ 0.797	$\beta(3.85, 11.24)$ 24.92	$\beta(3.85, 11.24)$ 12.49
flor1Aii.i	$\mathcal{W}(2.39, 0.012642)$ 16.09	$\sqrt{\Gamma}(1.35, 0.000230)$ 1.34	$\mathcal{W}(2.39, 0.012642)$ 16.09	$\mathcal{W}(2.39, 0.012642)$ 8.05
flor1Aiii.i	$\sqrt{\Gamma}(2.05, 0.000271)$ 28.19	$\sqrt{\Gamma}(2.05, 0.000271)$ 24.76	$\sqrt{\Gamma}(2.05, 0.000271)$ 28.19	$\sqrt{\Gamma}(2.05, 0.000271)$ 26.48
flor1Bi.i	$\beta(2.68, 7.77)$ 0.001	— 0.000	$\beta(2.68, 7.77)$ 0.001	$\beta(2.68, 7.77)$ 0.001
flor1Bii.i	$\mathcal{KA6}(1.35, 6151.49)$ 1.26	$\mathcal{KA6}(1.35, 6151.49)$ 0.000	$\mathcal{KA6}(1.35, 6151.49)$ 1.26	$\mathcal{KA6}(1.35, 6151.49)$ 0.629
flor1Biii.i	$\beta(2.92, 7.68)$ 15.15	$\mathcal{KA6}(1.51, 6049.73)$ 0.007	$\beta(2.92, 7.68)$ 15.15	$\beta(2.92, 7.68)$ 7.57
flor1Ci.i	$\beta(3.09, 7.87)$ 35.04	$\beta(3.09, 7.87)$ 8.39	$\beta(3.09, 7.87)$ 35.04	$\beta(3.09, 7.87)$ 21.72
flor1Cii.i	$\mathcal{KA6}(1.89, 7814.07)$ 17.69	$\sqrt{\Gamma}(5.31, 0.065463)$ 9.93	$\mathcal{KA6}(1.89, 7814.07)$ 17.69	$\mathcal{KA6}(1.89, 7814.07)$ 10.28
flor1Ciii.i	$\mathcal{KA6}(1.47, 8549.02)$ 23.28	$\mathcal{KA6}(1.47, 8549.02)$ 0.003	$\mathcal{KA6}(1.47, 8549.02)$ 23.28	$\mathcal{KA6}(1.47, 8549.02)$ 11.64
flor1_1Aii.i	$\mathcal{KA6}(2.51, 5999.77)$ 70.65	$\mathcal{KA6}(2.51, 5999.77)$ 63.44	$\mathcal{KA6}(2.51, 5999.77)$ 70.65	$\mathcal{KA6}(2.51, 5999.77)$ 67.05
flor1_1Aiii.i	$\beta(6.80, 13.80)$ 81.51	$\mathcal{KA6}(5.18, 7749.06)$ 37.55	$\beta(6.80, 13.80)$ 81.51	$\beta(6.80, 13.80)$ 49.19
flor1_1Ci.i	$\beta(6.33, 12.50)$ 90.69	$\beta(6.33, 12.50)$ 23.37	$\beta(6.33, 12.50)$ 90.69	$\beta(6.33, 12.50)$ 57.03
flor1_1Ciii.i	$\mathcal{LN}(4.47, 0.314618)$ 79.77	$\mathcal{KA6}(4.92, 8937.03)$ 84.86	$\mathcal{KA6}(4.92, 8937.03)$ 84.86	$\mathcal{KA6}(4.92, 8937.03)$ 80.61
flor1_2Aii.i	$\beta(5.37, 11.23)$ 89.70	$\beta(5.37, 11.23)$ 87.54	$\beta(5.37, 11.23)$ 89.70	$\beta(5.37, 11.23)$ 88.62
flor2Ai.i	$\sqrt{\Gamma}(1.86, 0.000321)$ 15.80	$\beta(4.79, 12.34)$ 2.68	$\sqrt{\Gamma}(1.86, 0.000321)$ 15.80	$\sqrt{\Gamma}(1.86, 0.000321)$ 8.24
flor2Aii.i	$\sqrt{\Gamma}(1.26, 0.000161)$ 35.90	$\sqrt{\Gamma}(1.26, 0.000161)$ 0.172	$\sqrt{\Gamma}(1.26, 0.000161)$ 35.90	$\sqrt{\Gamma}(1.26, 0.000161)$ 18.04
flor2Aiii.i	$\beta(2.89, 8.26)$ 1.24	— 0.000	$\beta(2.89, 8.26)$ 1.24	$\beta(2.89, 8.26)$ 0.621
flor2Bi.i	$\mathcal{KA6}(1.18, 5824.48)$ 1.04	— 0.000	$\mathcal{KA6}(1.18, 5824.48)$ 1.04	$\mathcal{KA6}(1.18, 5824.48)$ 0.518

(continua)

TABELA 6.3 - Continuação

Amostra	$\max p_{KS}$	$\max p_{\chi^2}$	$\max\{p_{KS}, p_{\chi^2}\}$	$\max(p_{KS} + p_{\chi^2})/2$
flor2Bii.i	$\beta(2.56, 6.19)$ 0.677	— 0.000	$\beta(2.56, 6.19)$ 0.677	$\beta(2.56, 6.19)$ 0.339
flor2Biii.i	$\Gamma(3.29, 0.04)$ 0.071	— 0.000	$\Gamma(3.29, 0.04)$ 0.071	$\Gamma(3.29, 0.04)$ 0.035
flor2Ci.i	$\Gamma(3.40, 0.05)$ 0.156	— 0.000	$\Gamma(3.40, 0.05)$ 0.156	$\Gamma(3.40, 0.05)$ 0.078
flor2Cii.i	$\mathcal{KA6}(1.81, 8420.61)$ 57.95	$\mathcal{KA6}(1.81, 8420.61)$ 14.48	$\mathcal{KA6}(1.81, 8420.61)$ 57.95	$\mathcal{KA6}(1.81, 8420.61)$ 35.55
flor2Ciii.i	$\mathcal{KA6}(1.40, 7153.36)$ 4.05	$\Gamma(4.14, 0.06)$ 0.000	$\mathcal{KA6}(1.40, 7153.36)$ 4.05	$\mathcal{KA6}(1.40, 7153.36)$ 2.03
flor2_1Aii.i	$\sqrt{\Gamma}(2.23, 0.000295)$ 98.81	$\Gamma(8.49, 0.10)$ 99.67	$\Gamma(8.49, 0.10)$ 99.67	$\sqrt{\Gamma}(2.23, 0.000295)$ 98.39
flor2_1Bii.i	$\mathcal{KA6}(2.56, 8119.76)$ 73.56	$\Gamma(6.72, 0.08)$ 12.91	$\mathcal{KA6}(2.56, 8119.76)$ 73.56	$\mathcal{KA6}(2.56, 8119.76)$ 42.43
flor2_2Aii.i	$\sqrt{\Gamma}(2.37, 0.000191)$ 97.99	$\Gamma(9.07, 0.09)$ 22.85	$\sqrt{\Gamma}(2.37, 0.000191)$ 97.99	$\mathcal{KA6}(3.96, 12412.67)$ 56.01
flor3Aii.i	$\beta(5.28, 13.34)$ 70.11	$\beta(5.28, 13.34)$ 54.66	$\beta(5.28, 13.34)$ 70.11	$\beta(5.28, 13.34)$ 62.39
flor3Aiii.i	$\sqrt{\Gamma}(1.84, 0.000258)$ 6.25	$\sqrt{\Gamma}(1.84, 0.000258)$ 0.036	$\sqrt{\Gamma}(1.84, 0.000258)$ 6.25	$\sqrt{\Gamma}(1.84, 0.000258)$ 3.14
flor3Bii.i	$\mathcal{W}(1.97, 0.010342)$ 5.51	— 0.000	$\mathcal{W}(1.97, 0.010342)$ 5.51	$\mathcal{W}(1.97, 0.010342)$ 2.75
flor3Biii.i	— 0.000	— 0.000	— 0.000	— 0.000
flor3Cii.i	$\mathcal{KA6}(1.49, 12104.23)$ 16.46	$\Gamma(4.36, 0.04)$ 0.002	$\mathcal{KA6}(1.49, 12104.23)$ 16.46	$\mathcal{KA6}(1.49, 12104.23)$ 8.23
flor3Ci.i	$\Gamma(3.61, 0.05)$ 0.000	— 0.000	$\Gamma(3.61, 0.05)$ 0.000	$\Gamma(3.61, 0.05)$ 0.000
flor3Ciii.i	$\Gamma(4.09, 0.06)$ 0.635	— 0.000	$\Gamma(4.09, 0.06)$ 0.635	$\Gamma(4.09, 0.06)$ 0.318
flor3Ciiii.i	$\Gamma(5.12, 0.07)$ 3.42	$\Gamma(5.12, 0.07)$ 0.000	$\Gamma(5.12, 0.07)$ 3.42	$\Gamma(5.12, 0.07)$ 1.71
flor3_1Aii.i	$\sqrt{\Gamma}(1.91, 0.0003)$ 88.31	$\mathcal{KA6}(2.83, 6359.07)$ 61.63	$\sqrt{\Gamma}(1.91, 0.0003)$ 88.31	$\sqrt{\Gamma}(1.91, 0.0003)$ 67.60
flor3_1Aiii.i	$\sqrt{\Gamma}(2.31, 0.000258)$ 79.69	$\mathcal{KA6}(3.78, 8947.46)$ 36.97	$\sqrt{\Gamma}(2.31, 0.000258)$ 79.69	$\sqrt{\Gamma}(2.31, 0.000258)$ 57.30
flor3_1Bii.i	$\mathcal{LN}(4.57, 0.38)$ 97.97	$\Gamma(7.10, 0.07)$ 19.86	$\mathcal{LN}(4.57, 0.38)$ 97.97	$\Gamma(7.10, 0.07)$ 55.80

(continua)

TABELA 6.3 - Continuação

Amostra	máxp_{KS}	máxp_{χ^2}	$\text{máx}\{p_{KS}, p_{\chi^2}\}$	$\text{máx}(p_{KS} + p_{\chi^2})/2$
flor3_1Bii.i	$\sqrt{\Gamma}(1.75, 0.000143)$ 87.31	$\mathcal{KA6}(2.49, 12185.61)$ 9.43	$\sqrt{\Gamma}(1.75, 0.000143)$ 87.31	$\sqrt{\Gamma}(1.75, 0.000143)$ 47.53
flor3_1Biii.i	$\mathcal{KA6}(5.04, 21443.42)$ 98.23	$\Gamma(10.48, 0.074918)$ 2.00	$\mathcal{KA6}(5.04, 21443.42)$ 98.23	$\Gamma(10.48, 0.074918)$ 50.08
flor3_1Ciii.i	$\mathcal{LN}(4.38, 0.30)$ 68.93	$\mathcal{LN}(4.38, 0.30)$ 99.83	$\mathcal{LN}(4.38, 0.30)$ 99.83	$\mathcal{LN}(4.38, 0.30)$ 84.38
flor3_2Aii.i	$\mathcal{LN}(4.39, 0.32)$ 97.68	$\Gamma(10.24, 0.12)$ 45.32	$\mathcal{LN}(4.39, 0.32)$ 97.68	$\mathcal{LN}(4.39, 0.32)$ 65.40
flor4Ai.i	$\sqrt{\Gamma}(1.64, 0.000340)$ 84.51	$\beta(4.36, 12.89)$ 13.67	$\sqrt{\Gamma}(1.64, 0.000340)$ 84.51	$\sqrt{\Gamma}(1.64, 0.000340)$ 45.84
flor4Aii.i	$\beta(3.13, 8.68)$ 0.761	— 0.000	$\beta(3.13, 8.68)$ 0.761	$\beta(3.13, 8.68)$ 0.381
flor4Bi.i	$\sqrt{\Gamma}(1.52, 0.000191)$ 12.50	$\mathcal{KA6}(2.06, 7948.21)$ 0.301	$\sqrt{\Gamma}(1.52, 0.000191)$ 12.50	$\sqrt{\Gamma}(1.52, 0.000191)$ 6.25
flor4Bii.i	$\mathcal{KA6}(1.17, 7156.01)$ 0.840	— 0.000	$\mathcal{KA6}(1.17, 7156.01)$ 0.840	$\mathcal{KA6}(1.17, 7156.01)$ 0.420
flor4Biii.i	$\sqrt{\Gamma}(1.44, 0.000180)$ 4.99	$\mathcal{KA6}(1.92, 7987.58)$ 0.262	$\sqrt{\Gamma}(1.44, 0.000180)$ 4.99	$\sqrt{\Gamma}(1.44, 0.000180)$ 2.50
flor4Ci.i	$\mathcal{KA6}(1.10, 7267.39)$ 0.138	— 0.000	$\mathcal{KA6}(1.10, 7267.39)$ 0.138	$\mathcal{KA6}(1.10, 7267.39)$ 0.069
flor4Ciii.i	$\Gamma(3.26, 0.04)$ 0.000	— 0.000	$\Gamma(3.26, 0.04)$ 0.000	$\Gamma(3.26, 0.04)$ 0.000
flor4_1Bi.i	$\Gamma(8.13, 0.09)$ 96.82	$\mathcal{KA6}(3.35, 8351.66)$ 14.56	$\Gamma(8.13, 0.09)$ 96.82	$\mathcal{KA6}(3.35, 8351.66)$ 55.63
flor4_1Biii.i	$\sqrt{\Gamma}(2.61, 0.000257)$ 99.76	$\mathcal{KA6}(4.65, 10169.50)$ 46.08	$\sqrt{\Gamma}(2.61, 0.000257)$ 99.76	$\sqrt{\Gamma}(2.61, 0.000257)$ 69.98
flor5Ai.i	$\sqrt{\Gamma}(1.37, 0.000201)$ 76.82	$\mathcal{KA6}(1.79, 6802.83)$ 12.06	$\sqrt{\Gamma}(1.37, 0.000201)$ 76.82	$\sqrt{\Gamma}(1.37, 0.000201)$ 44.40
flor5Aii.i	$\mathcal{KA6}(1.31, 5699.92)$ 0.600	— 0.000	$\mathcal{KA6}(1.31, 5699.92)$ 0.600	$\mathcal{KA6}(1.31, 5699.92)$ 0.300
flor5Bi.i	$\mathcal{KA6}(1.95, 9137.07)$ 2.02	$\mathcal{KA6}(1.95, 9137.07)$ 0.002	$\mathcal{KA6}(1.95, 9137.07)$ 2.02	$\mathcal{KA6}(1.95, 9137.07)$ 1.01
flor5Bii.i	$\mathcal{KA6}(1.64, 8187.26)$ 1.26	— 0.000	$\mathcal{KA6}(1.64, 8187.26)$ 1.26	$\mathcal{KA6}(1.64, 8187.26)$ 0.629
flor5Biii.i	$\mathcal{KA6}(1.21, 7038.70)$ 0.010	— 0.000	$\mathcal{KA6}(1.21, 7038.70)$ 0.010	$\mathcal{KA6}(1.21, 7038.70)$ 0.005
flor5_1Aii.i	$\beta(5.65, 12.44)$ 98.40	$\beta(5.65, 12.44)$ 97.67	$\beta(5.65, 12.44)$ 98.40	$\beta(5.65, 12.44)$ 98.03

(continua)

TABELA 6.3 - Conclusão

Amostra	$\text{máx}p_{KS}$	$\text{máx}p_{\chi^2}$	$\text{máx}\{p_{KS}, p_{\chi^2}\}$	$\text{máx}(p_{KS} + p_{\chi^2})/2$
flor5_1Bi.i	$\sqrt{\Gamma}(2.51, 0.000239)$ 64.49	$\Gamma(9.61, 0.10)$ 48.99	$\sqrt{\Gamma}(2.51, 0.000239)$ 64.49	$\mathcal{KA6}(4.35, 10513.82)$ 56.09
flor5_2Aii.i	$\sqrt{\Gamma}(1.89, 0.000295)$ 94.78	$\mathcal{KA6}(2.79, 6410.76)$ 83.04	$\sqrt{\Gamma}(1.89, 0.000295)$ 83.04	$\sqrt{\Gamma}(1.89, 0.000295)$ 87.19
flor6Ai.i	$\sqrt{\Gamma}(1.27, 0.000269)$ 43.79	$\beta(3.30, 10.22)$ 13.31	$\sqrt{\Gamma}(1.27, 0.000269)$ 43.79	$\sqrt{\Gamma}(1.27, 0.000269)$ 23.67
flor6Aii.i	$\mathcal{KA6}(1.11, 7321.10)$ 0.011	— 0.000	$\mathcal{KA6}(1.11, 7321.10)$ 0.011	$\mathcal{KA6}(1.11, 7321.10)$ 0.006
flor6Bi.i	$\beta(2.73, 7.06)$ 1.88	— 0.000	$\beta(2.73, 7.06)$ 1.88	$\beta(2.73, 7.06)$ 0.938
flor6Bii.i	$\mathcal{KA6}(1.20, 6189.09)$ 0.153	— 0.000	$\mathcal{KA6}(1.20, 6189.09)$ 0.153	$\mathcal{KA6}(1.20, 6189.09)$ 0.077
flor7Ai.i	$\mathcal{W}(2.74, 0.01)$ 94.58	$\beta(4.35, 11.21)$ 75.41	$\mathcal{W}(2.74, 0.01)$ 94.58	$\mathcal{W}(2.74, 0.01)$ 82.18
flor7Aii.i	$\mathcal{KA6}(1.32, 8379.99)$ 0.695	$\Gamma(3.91, 0.05)$ 0.003	$\mathcal{KA6}(1.32, 8379.99)$ 0.695	$\mathcal{KA6}(1.32, 8379.99)$ 0.348
flor7_1Ai.i	$\beta(4.33, 10.42)$ 94.52	$\beta(4.33, 10.42)$ 90.57	$\beta(4.33, 10.42)$ 94.52	$\beta(4.33, 10.42)$ 92.54
flor7_1Aii.i	$\sqrt{\Gamma}(1.57, 0.000144)$ 82.34	$\mathcal{KA6}(2.15, 10894.40)$ 13.51	$\sqrt{\Gamma}(1.57, 0.000144)$ 82.34	$\sqrt{\Gamma}(1.57, 0.000144)$ 47.41
flor7_TAi.i	$\sqrt{\Gamma}(1.33, 0.000236)$ 15.22	$\beta(3.35, 9.13)$ 0.000	$\sqrt{\Gamma}(1.33, 0.000236)$ 15.22	$\sqrt{\Gamma}(1.33, 0.000236)$ 7.61
flor8Ai.i	$\mathcal{KA6}(2.05, 7066.13)$ 91.54	$\mathcal{KA6}(2.05, 7066.13)$ 74.84	$\mathcal{KA6}(2.05, 7066.13)$ 91.54	$\mathcal{KA6}(2.05, 7066.13)$ 83.19
flor9Ai.i	$\mathcal{W}(3.20, 0.009699)$ 96.32	$\sqrt{\Gamma}(2.23, 0.000234)$ 45.42	$\mathcal{W}(3.20, 0.009699)$ 96.32	$\sqrt{\Gamma}(2.23, 0.000234)$ 69.01

TABELA 6.4 - DISTRIBUIÇÕES ESCOLHIDAS PARA AMOSTRAS DE NÃO FLORESTA POR QUATRO CRITÉRIOS.

Amostra	$\text{máx}_{p_{KS}}$	$\text{máx}_{p_{\chi^2}}$	$\text{máx}\{p_{KS}, p_{\chi^2}\}$	$\text{máx}(p_{KS} + p_{\chi^2})/2$
past0Bi.i	$\mathcal{N}(79.27, 17.57)$ 57.54	$\Gamma(20.36, 0.256815)$ 54.40	$\mathcal{N}(79.27, 17.57)$ 57.54	$\sqrt{\Gamma}(20.36, 0.256815)$ 49.59
past0aBi.i	$\mathcal{N}(87.66, 20.34)$ 82.23	$\Gamma(18.57, 0.211824)$ 66.57	$\mathcal{N}(87.66, 20.34)$ 82.23	$\sqrt{\Gamma}(4.76, 0.000587)$ 70.10
past11Cii.i	$\Gamma(20.50, 0.236426)$ 84.08	$\mathcal{LN}(4.45, 0.223041)$ 38.92	$\Gamma(20.50, 0.236426)$ 84.08	$\Gamma(20.50, 0.236426)$ 59.19
past12Cii.i	$\sqrt{\Gamma}(5.81, 0.000966)$ 53.70	$\mathcal{KA6}(101, 6014.97)$ 45.74	$\sqrt{\Gamma}(5.81, 0.000966)$ 53.70	$\sqrt{\Gamma}(5.81, 0.000966)$ 45.39
past13Cii.i	$\sqrt{\Gamma}(5.29, 0.000920)$ 5.11	$\mathcal{KA6}(45.09, 5754.23)$ 4.42	$\sqrt{\Gamma}(5.29, 0.000920)$ 5.11	$\sqrt{\Gamma}(5.29, 0.000920)$ 2.62
past1Bi.i	$\sqrt{\Gamma}(5.04, 0.001021)$ 45.76	$\mathcal{KA6}(31.46, 4930.02)$ 63.25	$\mathcal{KA6}(31.46, 4930.02)$ 63.25	$\sqrt{\Gamma}(5.04, 0.001021)$ 46.11
past1Bii.i	$\sqrt{\Gamma}(4.99, 0.000999)$ 47.60	$\mathcal{KA6}(29.69, 49994.51)$ 3.09	$\sqrt{\Gamma}(4.99, 0.000999)$ 47.60	$\sqrt{\Gamma}(4.99, 0.000999)$ 24.04
past1Biii.i	$\mathcal{LN}(4.41, 0.218066)$ 46.34	$\mathcal{KA6}(75.09, 7291.39)$ 19.95	$\mathcal{LN}(4.41, 0.218066)$ 46.34	$\mathcal{KA6}(75.09, 7291.39)$ 27.27
past1Ci.i	$\mathcal{LN}(4.28, 0.228764)$ 12.34	$\Gamma(19.16, 0.262711)$ 5.14	$\mathcal{LN}(4.28, 0.228764)$ 12.34	$\mathcal{LN}(4.28, 0.228764)$ 6.17
past1Cii.i	$\mathcal{KA6}(19.65, 0.267272)$ 45.69	$\Gamma(19.65, 0.267272)$ 30.61	$\mathcal{KA6}(19.65, 0.267272)$ 45.69	$\mathcal{KA6}(19.65, 0.267272)$ 34.06
past1_1Bi.i	$\mathcal{N}(78.77, 18.67)$ 95.95	$\beta(11.99, 26.82)$ 24.07	$\mathcal{N}(78.77, 18.67)$ 95.95	$\mathcal{N}(78.77, 18.67)$ 56.66
past2Bi.i	$\mathcal{LN}(4.33, 0.255481)$ 26.10	$\mathcal{KA6}(12.55, 6343.82)$ 34.31	$\mathcal{KA6}(12.55, 6343.82)$ 34.31	$\mathcal{KA6}(12.55, 6343.82)$ 23.72
past2Bii.i	$\mathcal{KA6}(43.59, 5043.84)$ 70.65	$\Gamma(20.62, 0.297314)$ 85.14	$\Gamma(20.62, 0.297314)$ 85.14	$\mathcal{KA6}(43.59, 5043.84)$ 73.32
past2Biii.i	$\beta(14.73, 29.58)$ 55.49	$\sqrt{\Gamma}(5.76, 0.000767)$ 44.90	$\beta(14.73, 29.58)$ 55.49	$\beta(14.73, 29.58)$ 49.75
past2Ci.i	$\sqrt{\Gamma}(5.52, 0.001022)$ 69.88	$\mathcal{KA6}(68.80, 5397.54)$ 64.75	$\sqrt{\Gamma}(5.52, 0.001022)$ 69.88	$\sqrt{\Gamma}(5.52, 0.001022)$ 63.30
past2Cii.i	$\mathcal{N}(80.90, 17.48)$ 65.66	$\sqrt{\Gamma}(5.48, 0.000799)$ 11.65	$\mathcal{N}(80.90, 17.48)$ 65.66	$\mathcal{N}(80.90, 17.48)$ 34.44
past2_1Bi.i	$\mathcal{LN}(4.29, 0.262733)$ 54.82	$\mathcal{KA6}(11.53, 5921.88)$ 26.91	$\mathcal{LN}(4.29, 0.262733)$ 54.82	$\mathcal{KA6}(11.53, 5921.88)$ 35.01

(continua)

TABELA 6.4 - Conclusão.

Amostra	máxp_{KS}	máxp_{χ^2}	$\text{máx}\{p_{KS}, p_{\chi^2}\}$	$\text{máx}(p_{KS} + p_{\chi^2})/2$
past2_2Bi.i	$\sqrt{\Gamma}(4.65, 0.000713)$ 97.41	$\beta(12.24, 27.46)$ 30.19	$\sqrt{\Gamma}(4.65, 0.000713)$ 97.41	$\sqrt{\Gamma}(4.65, 0.000713)$ 62.42
past3Bi.i	$\sqrt{\Gamma}(4.68, 0.000732)$ 49.25	$\mathcal{KA6}(21.38, 6393.40)$ 81.37	$\mathcal{KA6}(21.38, 6393.40)$ 81.37	$\beta(12.38, 28.17)$ 53.07
past3Bii.i	$\sqrt{\Gamma}(5.66, 0.000917)$ 46.47	$\beta(15.19, 35.22)$ 68.00	$\beta(15.19, 35.22)$ 68.00	$\beta(15.19, 35.22)$ 57.08
past3Biii.i	$\mathcal{N}(71.34, 15.10)$ 66.98	$\mathcal{KA6}(101, 5316.94)$ 63.95	$\mathcal{N}(71.34, 15.10)$ 66.98	$\sqrt{\Gamma}(5.69, 0.001071)$ 46.72
past3Ci.i	$\sqrt{\Gamma}(4.63, 0.000645)$ 47.78	$\sqrt{\Gamma}(4.63, 0.000645)$ 60.43	$\sqrt{\Gamma}(4.63, 0.000645)$ 60.43	$\sqrt{\Gamma}(4.63, 0.000645)$ 53.11
past3Cii.i	$\mathcal{KA6}(95.67, 5543.58)$ 89.47	$\Gamma(22.11, 0.303622)$ 47.20	$\mathcal{KA6}(95.67, 5543.58)$ 89.47	$\mathcal{KA6}(95.67, 5543.58)$ 61.45
past4Bi.i	$\mathcal{KA6}(23.00, 6497.40)$ 71.76	$\Gamma(18.56, 0.236330)$ 8.30	$\mathcal{KA6}(23.00, 6497.40)$ 71.76	$\mathcal{KA6}(23.00, 6497.40)$ 37.88
past4Bii.i	$\mathcal{N}(63.99, 12.86)$ 69.63	$\sqrt{\Gamma}(6.31, 0.001481)$ 33.89	$\mathcal{N}(63.99, 12.86)$ 69.63	$\mathcal{N}(63.99, 12.86)$ 36.90
past4Biii.i	$\mathcal{N}(69.71, 14.54)$ 23.71	$\beta(16.42, 43.65)$ 8.02	$\mathcal{N}(69.71, 14.54)$ 23.71	$\mathcal{N}(69.71, 14.54)$ 11.86
past4Cii.i	$\mathcal{LN}(4.14, 0.206068)$ 42.92	$\beta(17.83, 54.04)$ 27.15	$\mathcal{LN}(4.14, 0.206068)$ 42.92	$\beta(17.83, 54.04)$ 28.89
past5Bi.i	$\mathcal{N}(72.48, 17.69)$ 6.86	$\sqrt{\Gamma}(4.31, 0.000774)$ 9.80	$\sqrt{\Gamma}(4.31, 0.000774)$ 9.80	$\sqrt{\Gamma}(4.31, 0.000774)$ 5.50
past5Bii.i	$\mathcal{KA6}(21.46, 6458.96)$ 70.65	$\sqrt{\Gamma}(18.28, 0.233572)$ 96.64	$\sqrt{\Gamma}(18.28, 0.233572)$ 96.64	$\mathcal{KA6}(21.46, 6458.96)$ 83.27
past5Biii.i	$\mathcal{N}(82.00, 18.16)$ 32.68	$\mathcal{KA6}(39.87, 7053.99)$ 16.30	$\mathcal{N}(82.00, 18.16)$ 32.68	$\sqrt{\Gamma}(5.21, 0.000739)$ 16.77
past5Cii.i	$\mathcal{N}(78.51, 15.77)$ 37.14	$\sqrt{\Gamma}(6.31, 0.000984)$ 92.95	$\sqrt{\Gamma}(6.31, 0.000984)$ 92.95	$\mathcal{N}(78.51, 15.77)$ 61.96
past6Bi.i	$\mathcal{N}(75.19, 16.00)$ 21.85	$\mathcal{KA6}(94.16, 5908.89)$ 26.25	$\mathcal{KA6}(94.16, 5908.89)$ 26.25	$\mathcal{KA6}(94.16, 5908.89)$ 19.50
past6Cii.i	$\mathcal{LN}(4.34, 0.235128)$ 15.70	$\sqrt{\Gamma}(4.98, 0.000791)$ 0.0013	$\mathcal{LN}(4.34, 0.235128)$ 15.70	$\mathcal{LN}(4.34, 0.235128)$ 7.85
past7Bi.i	$\mathcal{LN}(4.33, 0.215823)$ 34.91	$\Gamma(21.55, 0.281778)$ 49.69	$\Gamma(21.55, 0.281778)$ 49.69	$\Gamma(21.55, 0.281778)$ 30.57
past7Cii.i	$\mathcal{LN}(4.36, 0.217458)$ 59.91	$\Gamma(21.55, 0.271644)$ 19.38	$\mathcal{LN}(4.36, 0.217458)$ 59.91	$\mathcal{LN}(4.36, 0.217458)$ 30.03
past8Cii.i	$\sqrt{\Gamma}(5.76, 0.000970)$ 46.20	$\mathcal{KA6}(101, 5937.70)$ 90.78	$\mathcal{KA6}(101, 5937.70)$ 90.78	$\sqrt{\Gamma}(5.76, 0.000970)$ 64.75
past9Cii.i	$\mathcal{LN}(4.23, 0.235095)$ 53.24	$\mathcal{KA6}(24.33, 5115.65)$ 78.03	$\mathcal{KA6}(24.33, 5115.65)$ 78.03	$\mathcal{KA6}(24.33, 5115.65)$ 64.37

Da Tabela 6.5 pode concluir-se que o teste χ^2 , além dos problemas apontados na Seção 6.5, fracassou em vinte amostras de floresta, isto é, enquanto o teste de Kolmogorov-Smirnov em apenas uma ocasião produziu um p -valor identicamente nulo para todas as distribuições consideradas e as setenta amostras, o teste χ^2 não conseguiu ajustar vinte dessas amostras. Da mesma tabela pode concluir-se que as três distribuições que melhor se ajustaram aos dados de floresta são a $\mathcal{KA6}$, a $\sqrt{\Gamma}$ e a β .

Da Tabela 6.6 pode concluir-se que, para amostras de não floresta, nenhum dos testes fracassou no ajuste das distribuições consideradas. Além disto, as quatro distribuições que melhor se ajustaram aos dados de não floresta foram a $\mathcal{KA6}$, a $\sqrt{\Gamma}$, a Γ e a \mathcal{N} .

TABELA 6.5 - DISTRIBUIÇÕES *PREFERIDAS* PELAS AMOSTRAS DE FLORESTA.

Critério	$\sqrt{\Gamma}$	β	\mathcal{W}	$\mathcal{KA6}$	\mathcal{LN}	Γ	Total
$\max p_{KS}$	20	14	4	20	4	7	69
$\max p_{\chi^2}$	6	11	0	21	1	11	50
$\max\{p_{KS}, p_{\chi^2}\}$	19	14	4	21	3	8	69
$\max(p_{KS} + p_{\chi^2})/2$	19	14	3	23	2	8	69

TABELA 6.6 - DISTRIBUIÇÕES *PREFERIDAS* PELAS AMOSTRAS DE NÃO FLORESTA.

Critério	\mathcal{N}	Γ	$\sqrt{\Gamma}$	\mathcal{LN}	$\mathcal{KA6}$	β	Total
$\max p_{KS}$	12	1	10	8	5	1	37
$\max p_{\chi^2}$	0	12	5	1	14	5	37
$\max\{p_{KS}, p_{\chi^2}\}$	8	4	9	5	9	2	37
$\max(p_{KS} + p_{\chi^2})/2$	5	4	10	4	10	4	37

6.5 Influência do Formato dos Dados nos Testes de Ajuste

Na implementação do programa empregado na Seção 6.4 foram encontrados vários problemas. Um deles foi o fato das rotinas apresentadas nas

referências (119,135) não serem totalmente adequadas às necessidades: a função $K_\alpha(t)$, requerida para as densidades e funções de distribuição acumuladas das Distribuições 2.10, 2.11, 2.12 e 2.13 somente admite $\alpha \in \mathbf{Z}$ e $t \in \mathbb{R}_+$. Como o parâmetro α pode ser estimado, é extremamente pouco provável se ter uma estimação que satisfaça o requerimento imposto pelas rotinas.

Tentou-se superar esse problema criando-se uma interface entre as linguagens de programação C++ e FORTRAN para, assim, poder aproveitar a rotina S18DCF da biblioteca de rotinas numéricas NAG (111), desenvolvida nesta última linguagem. A implementação da biblioteca NAG é para valores de α complexos, e t complexo de argumento positivo. Foi, portanto, especializada ao caso $\alpha \in \mathbb{R}$ e $t \in \mathbb{R}_+$ usando as relações apresentadas em (1).

Feita esta interface, e usada a rotina S18DCF, constatou-se que para os valores de α superiores a, aproximadamente, 3 apareciam ainda problemas numéricos. Foi, então criada uma interface para, desde o programa principal em C++ poder chamar rotinas do pacote MatLab Versão 4 (132). Cabe ressaltar que a Versão 3.5h do mesmo pacote, anteriormente usada, fornecia rotinas numéricas pouco adequadas às necessidades computacionais encontradas. Já a função `besselk`, da Versão 4, mostrou-se muito adequada aos requerimentos encontrados.

Um outro problema, que é abordado nesta Seção, é o da influência do formato dos dados nos testes de ajuste. Os testes de qualidade de ajuste apresentados nas Seções 6.3.1 e 6.3.2 requerem a especificação da distribuição considerada como hipótese nula. Foram consideradas as Distribuições 2.2, 2.3, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.11, 2.12, 2.13, 2.14 e 2.15, sendo que todas elas têm como suporte um conjunto não enumerável (\mathbb{R} ou \mathbb{R}_+). Por outro lado, todos os dados têm como contradomínio o conjunto $\{0, \dots, 255\}$. Assim, embora fosse observada a ocorrência da variável aleatória $Y \sim \mathcal{D}$, com \mathcal{D} uma das distribuições mencionadas, os testes χ^2 e Kolmogorov-Smirnov seriam aplicados a dados provindos do resultado de se observar ocorrências de $\max\{\min\{\lfloor Y + \frac{1}{2} \rfloor, 255\}, 0\}$.

Para se avaliar a influência que tem sobre os testes o fato de serem aplicados a $\max\{\min\{\lfloor y + \frac{1}{2} \rfloor, 255\}, 0\}$ em vez de serem aplicados a y , foi feita a experiência Monte Carlo descrita a seguir. Procurou-se discriminar o efeito que têm a discretização, o tamanho da amostra considerada e o(s) parâmetro(s) relevantes. A discretização pode afetar nos testes de, pelo menos, duas maneiras diferentes:

- 1) alterando o(s) valor(es) estimado(s) do(s) parâmetro(s), pois os estimadores são calculados sobre a amostra discretizada e truncada, e
- 2) modificando a qualidade dos ajustes efetuados.

Para se obter resultados de interesse à área de aplicação considerada neste trabalho foram, primeiro, estudados os tamanhos de amostra *típicos* que apareceram na Seção 6.4 e, depois, para as distribuições que se mostraram mais adequadas ao estudo, os valores dos seus parâmetros. Chegou-se à conclusão que amostras de tamanho 500 caracterizam bem a situação de *amostras pequenas*. Para a distribuição Rayleigh (Distribuição 2.8), observou-se que o parâmetro variando no intervalo [40, 150] fornece uma boa modelagem das situações encontradas.

Foram, assim, simulados 5000 vetores de tamanho 500 da distribuição Rayleigh e parâmetro em [40(10)150]. Cada uma dessas amostras foi submetida aos testes χ^2 e Kolmogorov-Smirnov, aos níveis 90%, 95% e 99% (amostras independentes para cada nível) e foram calculadas a média e o desvio padrão do número de vetores rejeitados. Os resultados são mostrados nas Tabelas 6.7, 6.8 e 6.9, e nas Figuras 6.1, 6.2 e 6.3.

Das propriedades dos testes aplicados (11) sabe-se que, assintoticamente e sem efetuar o arredondamento imposto pelo formato de imagens usado, deveria ser rejeitado o $(1 - \alpha)\%$ das amostras, pois os dados (antes de serem arredondados) têm a distribuição da hipótese nula. Entretanto, das Tabelas 6.7, 6.8 e 6.9 segue que, para a situação considerada (tamanho da amostra, discretização e truncamento antes da estimação e dos testes), a rejeição é muito maior do que o esperado.

TABELA 6.7 - MÉDIA E DESVIO PADRÃO DO NÚMERO DE AMOSTRAS RAYLEIGH DE TAMANHO 500 REJEITADAS AO 90% PELOS TESTES KOLMOGOROV-SMIRNOV E χ^2

ξ	Kolmogorov-Smirnov	χ^2
40	0.961801(0.000764)	0.988000(0.000240)
50	0.886400(0.002272)	0.930200(0.001396)
60	0.800601(0.003994)	0.863999(0.002731)
70	0.715000(0.014300)	0.781401(0.004378)
80	0.642200(0.012844)	0.717601(0.005652)
90	0.571400(0.008574)	0.637401(0.007253)
100	0.524800(0.010497)	0.593800(0.008125)
110	0.458200(0.009165)	0.531601(0.010633)
120	0.416000(0.011681)	0.491800(0.009837)
130	0.381600(0.007634)	0.446201(0.008925)
140	0.354600(0.007093)	0.415600(0.008314)
150	0.310600(0.013788)	0.397200(0.012057)

TABELA 6.8 - MÉDIA E DESVIO PADRÃO DO NÚMERO DE AMOSTRAS RAYLEIGH DE TAMANHO 500 REJEITADAS AO 95% PELOS TESTES KOLMOGOROV-SMIRNOV E χ^2

ξ	Kolmogorov-Smirnov	χ^2
40	0.878600(0.002428)	0.953001(0.000940)
50	0.756400(0.004872)	0.836201(0.003276)
60	0.654599(0.006909)	0.758401(0.015168)
70	0.530000(0.010601)	0.638801(0.012777)
80	0.458200(0.010837)	0.563201(0.008737)
90	0.394000(0.007882)	0.480400(0.009609)
100	0.336200(0.006725)	0.423000(0.008461)
110	0.288600(0.014229)	0.374800(0.012504)
120	0.256601(0.005133)	0.343800(0.006877)
130	0.226000(0.015480)	0.311000(0.013780)
140	0.191600(0.003832)	0.288000(0.014240)
150	0.180000(0.016400)	0.267200(0.005348)

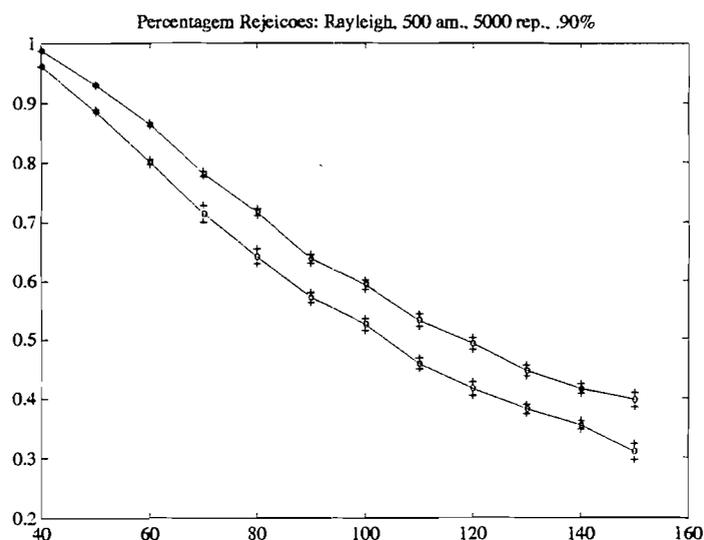


Fig. 6.1 - Média (“o”) e desvio padrão (“+”) do número de amostras Rayleigh rejeitadas ao 90% (linha superior χ^2 ; inferior Kolmogorov-Smirnov).

Esse excesso de rejeição diminui conforme o parâmetro considerado aumenta (ver Figuras 6.1, 6.2 e 6.3), porém, é consistente com o nível ao qual os testes são efetuados. Outra conclusão é que o teste χ^2 parece mais sensível às condições da experiência do que o teste Kolmogorov-Smirnov; o primeiro rejeita sistematicamente mais vetores do que o segundo.

Desta maneira, o fato dos testes considerados rejeitarem a hipótese de distribuição Rayleigh numa dada amostra pode ser questionado. Essa incerteza a respeito do p -valor dos testes foi quantizada em função do parâmetro, para um dado tamanho de amostra.

A experiência comentada será efetuada para algumas das outras distribuições mencionadas; por exemplo as apresentadas nas Tabelas 6.3 e 6.4, com valores dos parâmetros variando em um intervalo de tal maneira que sejam empregados todos os valores observados na prática. Com esses resultados será possível se ter uma idéia a respeito da *confiança* que pode se ter nos p -valores fornecidos pelos testes quando aplicados a amostras sob as condições consideradas.

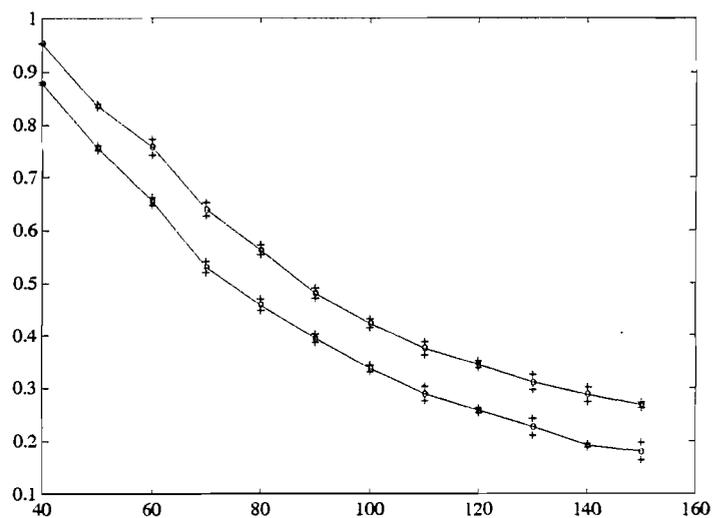


Fig. 6.2 - Média (“o”) e desvio padrão (“+”) do número de amostras Rayleigh rejeitadas ao 95% (linha superior χ^2 ; inferior Kolmogorov-Smirnov).

TABELA 6.9 - MÉDIA E DESVIO PADRÃO DO NÚMERO DE AMOSTRAS RAYLEIGH DE TAMANHO 500 REJEITADAS AO 99% PELOS TESTES KOLMOGOROV-SMIRNOV E χ^2

40	0.580400(0.008394)	0.768400(0.004637)
50	0.411800(0.008237)	0.590200(0.008197)
60	0.294200(0.005885)	0.458200(0.009165)
70	0.215000(0.015700)	0.335200(0.013296)
80	0.163800(0.003276)	0.268000(0.005361)
90	0.127800(0.002556)	0.211800(0.004236)
100	0.101600(0.002032)	0.175000(0.003500)
110	0.083000(0.001660)	0.153600(0.003081)
120	0.069600(0.018608)	0.134800(0.017304)
130	0.048800(0.000976)	0.107000(0.017860)
140	0.041200(0.000824)	0.093600(0.001872)
150	0.040000(0.000800)	0.086600(0.001732)

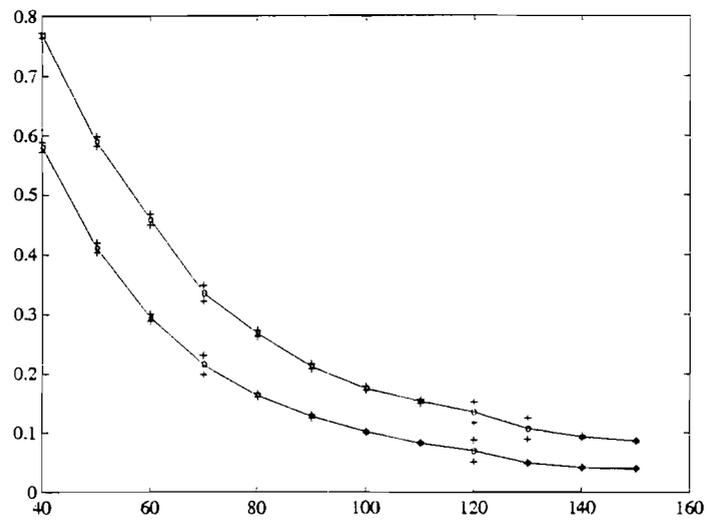


Fig. 6.3 - Média (“o”) e desvio padrão (“+”) do número de amostras Rayleigh rejeitadas ao 99% (linha superior χ^2 ; inferior Kolmogorov-Smirnov).

CAPÍTULO 7

CONCLUSÕES

O problema estudado neste trabalho foi o tocante ao desenvolvimento de ferramentas estatísticas específicas para a síntese, análise e processamento de imagens de radar de abertura sintética. A abordagem estatística foi escolhida devido à natureza predominantemente estocástica do ruído *speckle*, típico deste tipo de imageamento. Alguns destes tópicos foram abordados de tal maneira que as ferramentas desenvolvidas podem, também, ser aplicadas a imagens provindas de sensores ópticos. A seguir, são comentadas as principais contribuições desta tese, bem como os problemas que ficaram em aberto.

No Capítulo 2 foi apresentada uma forma fechada, e computacionalmente tratável, das densidades e funções de distribuição acumuladas das distribuições \mathcal{K} de amplitude e intensidade para múltiplas visadas. São calculados estimadores para os parâmetros destas distribuições. Resta avaliar o desempenho, relativo a critérios de consistência, eficiência, distribuição assintótica e custo computacional, das técnicas de estimação por máxima verossimilhança e pelo método de substituição para os modelos de interesse. Também seria de interesse avaliar o desempenho dos estimadores comentados na Seção 2.6.4 e, se possível, fazer uma avaliação global de todas as técnicas de segmentação.

No Capítulo 3 é estudado o modelo de Potts-Straus, que foi empregado como distribuição *a priori* para os problemas de síntese e segmentação (análise automática) de imagens. Comenta-se a influência que o fenômeno de transição de fase pode ter nestas aplicações. São fornecidos estimadores computacionalmente tratáveis para o parâmetro que caracteriza este modelo. Estudam-se as propriedades de convergência dos algoritmos de Swendsen-Wang e de Wolff, e essas propriedades são empregadas para o desenvolvimento de algoritmos para síntese de imagens. Esta ferramenta permite simular imagens, com um número arbitrário de

classes, e observações distribuídas, marginalmente, segundo a distribuição de uma média de variáveis aleatórias Rayleigh. Dessa maneira, é possível ver o efeito que o número de visadas tem na qualidade visual das imagens de amplitude. Além disso, as observações podem ser espacialmente independentes ou não. Uma ferramenta análoga é projetada para a simulação de imagens ópticas multibandas. Prova-se que, sob condições bastante gerais, a distribuição *a posteriori* de um campo Markoviano, dada a versão submetida a borramento e ruído, é Markoviana, estendendo o teorema de Geman e Geman. Resta, do ponto de vista teórico, estudar as propriedades distribucionais assintóticas dos estimadores de máxima pseudoverossimilhança, entre outros problemas em aberto a respeito desta técnica de estimação. Do ponto de vista computacional, seria de interesse estender a aplicabilidade da ferramenta de simulação para incluir, por exemplo, detecção quadrática e outros modelos de ruído.

No Capítulo 4 foram desenvolvidas várias versões do algoritmo ICM, para imagens SAR e uma visada e para imagens ópticas. Foi possível *automatizar* esta técnica de segmentação pelo uso dos estimadores de máxima pseudoverossimilhança. Desta maneira, e para o algoritmo ICM, foi resolvido um dos principais impedimentos para a divulgação do uso de técnicas Markovianas de segmentação: o fato de, em geral, não se dispor dos valores dos parâmetros dos quais dependem as soluções. Avalia-se o uso do algoritmo ICM como filtro. Assim sendo, a versão do algoritmo ICM para imagens ópticas foi incorporada ao sistema SPRING. Resta avaliar quantitativamente o desempenho do algoritmo ICM, para o qual poderiam ser empregadas as ferramentas de simulação propostas no Capítulo 3.

No Capítulo 5 são estudadas as principais propriedades de distribuição Rayleigh. Usando essas propriedades, são propostas técnicas robustas para a redução do ruído *speckle* em imagens de uma visada e detecção linear. É feita uma avaliação do desempenho das diferentes técnicas de filtragem. Poder-se-ia prosseguir o trabalho iniciado neste Capítulo pela avaliação das técnicas de estimação levando-se em conta as características específicas do problema de filtragem (dados truncados e arredondados, correlação espacial, etc.). Essa avaliação envolve problemas teóricos e computacionais.

No Capítulo 6 é desenvolvida uma ferramenta para a aplicação dos testes Kolmogorov-Smirnov e χ^2 a dados em formato de imagens (discretizados em 256 níveis de cinza, ou formato *byte*). Comenta-se o uso dos dados derivados da aplicação desses testes no problema da discriminação do uso do solo. Seria possível prosseguir o trabalho procurando-se quais os atributos derivados (*p*-valores dos testes, parâmetros estimados, etc.) que permitem uma melhor discriminação entre tipos de floresta, de regeneração, etc., e procurar novos atributos (por exemplo, pelo estimação da função de autocorrelação espacial). Inicia-se o estudo da influência que o formato *byte* tem nos testes de ajuste, aplicado à distribuição Rayleigh, e seria de interesse extendê-lo para outras distribuições.

REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Abramowitz, M.; Stegun, I. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. New York, Dover, 1964. 1046 p.
- (2) Ahrens, J.H.; Dieter, U. Extensions of Forsythe's method for random sampling from the normal distribution. *Mathematics of Computation*, 27(124):927-937, Oct. 1973.
- (3) Arsenault, H.H.; April, G. Properties of speckle integrated with a finite aperture and logarithmically transformed. *Journal of the Optical Society of America: Speckle in Optics*, 66(11):1160-1163, 1976.
- (4) Baillie, C.; Coddington, P. Comparison of cluster algorithms for two-dimensional Potts models. *Physical Review B*, 43(13-A):10617-10621, May 1991.
- (5) Balakrishnan, N. Approximate MLE of the scale parameter of the Rayleigh distribution with censoring. *IEEE Transactions on Reliability*, 38(3):355-357, Aug. 1989.
- (6) Balakrishnan, N.; Clifford Cohen, A. *Order statistics and inference: estimation methods*. San Diego, CA, Academic, 1991. 377 p. (Statistical Modelling and Decision Science).
- (7) Banon, G. *Processamento digital de imagens: notas de aula dadas no INPE*. São José dos Campos, 3. período, 1990. Manuscrito.
- (8) Barnett, V.; Lewis, T. *Outliers in statistical data*. Great Britain, John Wiley & Sons, 1978.
- (9) Besag, J. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society B*, 48(3):259-302, 1986.

- (10) Besag, J. Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.
- (11) Bickel, P.J.; Doksum, K.A. *Mathematical statistics: basic ideas and selected topics*. USA, Holden-Day, 1977. 493 p.
- (12) Binder, K., ed. *Monte Carlo methods in statistical physics*. 2. ed. Berlin, Springer-Verlag, 1986.
- (13) Binder, K. *Applications of the Monte Carlo method in statistical physics*. 2. ed. Berlin, Springer-Verlag, 1987.
- (14) Bovik, A.C.; Huang, T.S.; Munson, D.C. A generalization of median filtering using linear combination of order statistics. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-31(6):1342–1349, Dec. 1983.
- (15) Brown, B.W.; Lovato, J. *RANLIB.C: library of C routines for random number generation*. Houston, TX, s.ed., s.d.
- (16) Brown, W.M.; Palermo, C.J. *Random processes, communications and radar*. New York, NY, McGraw-Hill, 1969.
- (17) Bustos, O.H. *Estimação robusta no modelo de posição*. Rio de Janeiro, Instituto de Matemática Pura e Aplicada, 1981. 135 p. Notas de curso do Colóquio Brasileiro de Matemática, 13., Poços de Caldas, MG, 12–24 jul. 1981.
- (18) ——— Outliers e robustez. *Revista Brasileira de Estatística*, 49(191):5–25, jan.–jun. 1988.
- (19) ——— *A TURBO C implementation of a GFSR[0,1] pseudorandom number algorithm*. Rio de Janeiro, IMPA, 1990. (Informes de Matemática B-62).
- (20) ——— *Transición de fase: dificultad al modelar imágenes por distribuciones de Gibbs*. Palestra ministrada no INPE, São José dos Campos, nov. 1993.

- (21) Bustos, O.H.; Frery, A.C. Markovian fields as limits of Markov chains. In: Jornadas Internacionales de Uso de Computadoras en Actividades Científicas y Tecnológicas, 2., Mendoza, Argentina, 1991. Mendoza, EDIUM, 1991, v. 1., p. 98–107.
- (22) ——— A contribution to the study of Markovian degraded images: an extension of a theorem by Geman and Geman. *Matemática Aplicada e Computacional*, 11(1):17–29; (3):281–285, 1992a.
- (23) ——— *Simulação estocástica: teoria e algoritmos (versão completa)*. Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)/Instituto de Matemática Pura e Aplicada, 1992b. 148 p. (Monografias de Matemática, 49).
- (24) ——— *Reporting Monte Carlo experiences in statistics: suggestions and an example*. 1993. Submetido para publicação à *Revista de la Sociedad Chilena de Estadística*.
- (25) Bustos, O.H.; Ojeda, S. *Campos aleatorios Markovianos en procesamiento de imágenes: notas de aula dadas no INPE*. São José dos Campos, 1. período, 1993. Datilografado.
- (26) Câmara, G.; Souza, R.C.M.; Freitas, U.M.; Casanova, M.A. SPRING: processamento de imagens e dados georeferenciados. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens, 5., Águas de Lindóia, nov. 1992. São José dos Campos, INPE/SBC, 1992, p. 233–242.
- (27) Carnevalli, P.; Coletti, L.; Patarnello, S. Image processing by simulated annealing. *IBM Journal of Research and Development*, 29(6):569–579, Nov. 1985.
- (28) Caves, R.G. *Automatic matching of features in synthetic aperture radar data to digital map data*. (PhD Thesis) – University of Sheffield, Sheffield, UK, June 1993. 228 p.

- (29) Chellappa, R.; Jain, A., ed. *Markov random fields: theory and application*. San Diego, CA, Academic, 1993. 581 p.
- (30) Chou, P.B. *The theory and practice of Bayesian image labelling*. (PhD Thesis) – Department of Computer Science, University of Rochester, Rochester, NY, Aug. 1999. 142 p. (Technical Report 258).
- (31) Chou, P.B.; Cooper, P.R.; Swain, M.J.; Brown, C.M.; Wixson, L.E. Probabilistic network inference for cooperative high and low level vision. In: Chellappa, R.; Jain, A., ed. *Markov random fields: theory and application*. San Diego, CA, Academic, 1993. p. 211–243.
- (32) Chung, K.L. *A course in probability theory*. 2. ed. New York, Academic, 1974.
- (33) Cooper, D.B.; Subrahmonia, J.; Hung, Y-P.; Cernuschi-Frias, B. The use of Markov random fields in estimating and recognizing objects in 3D space. In: Chellappa, R.; Jain, A., ed. *Markov random fields: theory and application*. San Diego, CA, Academic, 1993. p. 335–367.
- (34) Corti, M.A. *Evaluación de los métodos de generación de números aleatorios*. (Trabajo Final de Analista de Sistemas) – Facultad de Ingeniería, Universidad de Mendoza, Mendoza, Argentina, dic. 1992. 129 p.
- (35) Cross, G.R.; Jain, A.K. Markov random field texture model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5:149–163, 1983.
- (36) Derin, H.; Kelly, P.A.; Vézina, G.; Labitt, S.G. Modelling and segmentation of speckled images using complex data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(1):76–87, Jan. 1990.
- (37) Devroye, L. *Non-uniform random variate generation*. New York, Springer-Verlag, 1986. 843 p.

- (38) D’Onorio De Meo, M.; Heermann, D.; Binder, K. Monte Carlo study of the Ising model phase transition in terms of the percolation transition of “physical clusters”. *Journal of Statistical Physics*, 60(5-6):585–618, 1990.
- (39) Dubes, R.C.; Jain, A.K. Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.
- (40) Duda, R.; Hart, E. *Pattern classification and scene analysis*. USA, John Wiley & Sons, 1973.
- (41) Dutra, L.V.; Frery, A.C.; Hernandez, P.F.; Sant’Anna, S.J.S.; Yanasse, C.C.F. Análisis estadístico de datos SAREX sobre la región de Tapajós – Brasil. In: Simposio Latinoamericano de Percepción Remota, 6., Cartagena, Colombia, 3–8 oct. 1993a. No prelo.
- (42) Dutra, L.V.; Frery, A.C.; Hernandez, P.F.; Sant’Anna, S.J.S.; Yanasse, C.C.F. Avaliação do uso de Imagens SAR Polarimétricas para Levantamento de Recursos Terrestres na Amazônia: Reserva Florestal de Tapajós. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens, 6., Recife, PE, 19–22 out. 1993b. *Comunicações*. Recife, PE, SBC/UFPe, 1993b, p. 91–94.
- (43) Dutra, L.V.; Frery, A.C.; Krug, T.; Mascarenhas, N.D.A.; Sant’Anna, S.J.S.; Yanasse, C.C.F. *Alguns aspectos de modelagem estatística de dados de sensoriamento remoto. Tutorial I*. In: Simpósio Brasileiro de Sensoriamento Remoto, 7., Curitiba, 10–14 maio 1993. São José dos Campos, INPE, 1993. 25 p. Notas de curso ministrado no Simpósio Brasileiro de Sensoriamento Remoto, 7.
- (44) Dutra, L.V.; Mascarenhas, N.D.A. Realce de texturas em imagens de radar de abertura sintética (SAR) por filtros de redução de “speckle”. In: Simpósio Brasileiro de Sensoriamento Remoto, 5., Manaus, AM, 24–29 jun. 1990. *Anais*. São José dos Campos, INPE, 1990, v 1., p. 173–180.

- (45) Dyer, D.D.; Whisenand, C.W. Best linear unbiased estimator of the parameter of the Rayleigh distribution—part I: small sample theory for censored order statistics. *IEEE Transactions on Reliability*, R-22(1):27–34, Apr. 1973.
- (46) Erthal, G.J.; Frery, A.C. Segmentação de imagens multiespectrais pelo algoritmo ICM: integração ao ambiente SPRING. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens, 6., Recife, PE, 19–22 out. 1993. *Comunicações*. Recife, PE, SBC/UFPe, 1993, p. 33–36.
- (47) Farag, A.A.; Delp, E.J. Image segmentation based on composite random field models. *Optical Engineering*, 31(12):2594–2607, 1992.
- (48) Fau, R.; Béné, G.B.; Boucher, J.-M.; He, D.-C. Segmentation Markovienne pyramidale d'images. In: Canadian Symposium on Remote Sensing, 16. Congrès de L'Association Québécoise de Télédétection, 8, Canada, 1993, L'AQT/CRSS, p. 701–708.
- (49) Fernandes, D. *Formação de imagens de radar de abertura sintética e modelos da relação "speckle"-textura*. (Tese de Doutorado) – Instituto Tecnológico de Aeronáutica, São José dos Campos, SP, nov. 1993.
- (50) Frery, A.C. *Estatística na análise e na síntese de imagens*. (Dissertação de Mestrado) – Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1990. 99 p. (Informes de Matemática, D-34/90).
- (51) ——— Monte Carlo determination of the convergence time of two cluster flip algorithms in the Ising model. *Revista Brasileira de Física*, 21(4):482–491, Dec. 1991a.
- (52) ——— *Markovian fields: theory, simulation and parameter estimation*. Rio de Janeiro, Instituto de Matemática Pura e Aplicada – CNPq, fev. 1991b. (Informes de Matemática, B-064).

- (53) Frery, A.C.; Marengoni, M. The Swendsen-Wang dynamics in the simulation of the Ising model. In: *Jornadas Internacionales de Uso de Computadoras en Actividades Científicas y Tecnológicas*, 2., Mendoza, Argentina, 1991. Mendoza, EDIUM, 1991, v. 1., p. 98–107.
- (54) Frery, A.C.; Mascarenhas, N.D.A. Segmentação binária de imagens SAR pelo algoritmo ICM. In: *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, 5., Águas de Lindóia, nov. 1992. São José dos Campos, INPE/SBC, 1992, p. 73–79.
- (55) Frery, A.C.; Sant’Anna, S.J.S. Redução de ruído em imagens SAR pelo uso de filtros robustos. In: *Simpósio Brasileiro de Sensoriamento Remoto*, 7., Curitiba, 10–14 maio 1993. *Anais*. São José dos Campos, INPE, 1993a, v. 3., p. 433–443.
- (56) Frery, A.C.; Sant’Anna, S.J.S. Non-adaptive robust filters for speckle noise reduction. In: *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, 6., Recife, PE, 19–22 out. 1993b. *Anais*. Recife, PE, SBC/UFPe, 1993b, p. 165–174.
- (57) Frost, V.S.; Stiles, J.A.; Shanmugan, K.S.; Holtzman, J.C. A model for radar images and its applications to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis Machine Intelligence*, PAMI-4(2):157–166, 1982.
- (58) Geman, D. Random fields and inverse problems in imaging. In: Hennequin, P.L., ed. *École d’été de probabilités de Saint-Flour XVIII-1988*. Berlin, Springer-Verlag, 1988. (Lecture Notes in Mathematics, 1427).
- (59) Geman, S.; Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov. 1984.

- (60) Geman, D.; Geman, S.; Graffigne, C.; Dong, P. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12(7):609–628, 1990.
- (61) Georgii, H. *Gibbs measures and phase transitions*. Berlin, Walter De Gruyter, 1988. (De Gruyter Studies in Mathematics, 9).
- (62) Goldfinger, A.D. Estimation of spectra from speckled images. *IEEE Transactions on Aerospace and Electronic Systems*, AES-18(5):675–681, 1982.
- (63) González, T.; Sahni, S.; Frante, W.R. An efficient algorithm for the Kolmogorov-Smirnov and Lillefor tests. *ACM Transactions on Mathematical Software*, 3:60–64, 1977.
- (64) Goodman, J.W. Statistical properties of laser speckle patterns. In: Dainty, J.C., ed. *Laser speckle and related phenomena*. Berlin, Springer-Verlag, 1982. cap. 2.
- (65) Graham, R.; Knuth, D.; Patashnik, O. *Concrete mathematics: a foundation for computer science*. USA, Addison-Wesley, 1988. 625 p.
- (66) Green, P.J; Han, X.-L. Metropolis methods, Gaussian proposals and antithetic variables. In: Barone, P.; Frigessi, A.; Piccioni, M., ed. *Stochastic models, statistical methods, and algorithms in image analysis: proceedings of the special year on image analysis*. USA, Springer-Verlag, 1992. cap. 10, p. 142–164.
- (67) Haralick, R.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621, Nov. 1973.
- (68) Hayes, B. The wheel of fortune. *American Scientist*, 81:114–118, Mar.–Apr. 1993.
- (69) IMSL *Stat. Library IMSL*. Houston, TX, 1987.

- (70) Instituto Nacional de Pesquisas Espaciais. *Manual de normas para publicações técnico-científicas*. São José dos Campos, 1990. (INPE-5116-MAN/001).
- (71) Huang, K. *Statistical mechanics*. New York, John Wiley, 1963.
- (72) Jackeman, E. On the statistics of K-distributed noise. *Journal of Physics A: Mathematical and General*, 13(1):31-48, 1980.
- (73) Jackeman, E.; Pusey, P.N. The statistics of light scattered by a random phase screen. *Journal of Physics A: Mathematical and General*, 6:L89-L92, 1973.
- (74) Jackeman, E.; Pusey, P.N. A model for non-Rayleigh sea echo. *IEEE Transactions on Antennas and Propagation*, AP-24(6):806-814, 1976.
- (75) Jain, A.K. *Fundamentals of digital image processing*. USA, Prentice-Hall International Editions, 1989. 569 p.
- (76) James, B.R. *Probabilidade: um curso em nível intermediário*. Rio de Janeiro, Instituto de Matemática Pura e Aplicada – CNPq, 1981. 304 p. (Projeto Euclides).
- (77) Jensen, J.L.; Møller, J. *Pseudolikelihood for exponential family models of spatial processes*. Denmark, Department of Theoretical Statistics, Institute of Mathematics, University of Århus, s.d.
- (78) Johnson, M.E. *Multivariate statistical simulation*. New York, John Wiley & Sons, 1987. 230 p.
- (79) Kendall, M.G.; Buckland, W.R. *A dictionary of statistical terms*. Edinburg, Oliver and Boyd, 1971.
- (80) Kelly, P.A.; Derin, H.; Hartt, K.D. Adaptive segmentation of speckled images using a hierarchical random field model. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(10):1628–1641, Oct. 1988.

- (81) Klein, R.; Press, S.J. Contextual Bayesian classification of remotely sensed data. *Communications in Statistics, Part A: Theory and Methods*, 18:3177–3202, 1989.
- (82) ——— Bayesian classification of remotely sensed data when training data is part of the scene. *Revista Brasileira de Probabilidade e Estatística*, 4(1):43–67, 1990.
- (83) ——— Adaptive Bayesian classification of spatial data. *Journal of the American Statistical Association*, 87(419):844–851, 1992.
- (84) Kie-Bum, E.; Kashyap, R.L. Robust image model based noise suppression. The Scandinavian Conference on Image Analysis, 5., Stockholm, June 2–5, 1987. Stockholm, IEEE, 1987.
- (85) Kitagawa, G. On the use of AIC for the detection of outliers. *Technometrics*, 21(2):193–199, May 1979.
- (86) Koroliuk, V. *Manual de la teoría de probabilidades y estadística matemática*. Moscú, Editorial Mir, 1986. 579 p. 1. reimp.
- (87) Kuan, D.T.; Sawchuk, A.A.; Strand, T.C.; Chavel, P. Adaptive restoration of images with speckle. [Washington], SPIE, 1982. p. 28–38. (SPIE Proceeding v. 359).
- (88) Kundu, A.; Mitra, S.K.; Vaidyanathan, P.P. Application of two-dimensional generalized mean filtering for removal of impulse noises from images. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-32(3):600–609, June 1984.
- (89) Lamport, L. \LaTeX : a document preparation system. Reading, MA, Addison-Wesley, 1986. 242 p.
- (90) L’Ecuyer, P.; Côté, S. Implementing a random number package with splitting facilities. *ACM Transactions on Mathematical Software*, 17(1):98–111, Mar. 1991.

- (91) Lee, J.S. Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17(1):24–32, Sep. 1981a.
- (92) Lee, J.S. A simple speckle smoothing algorithm for synthetic aperture radar images. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(1):85–89, Jan.–Fev. 1981b.
- (93) Leguizamón, S. Análisis textural de imágenes mediante medición de entropía. In: *Jornadas Internacionales de Uso de Computadoras en Actividades Científicas y Tecnológicas*, 1., Mendoza, Argentina, 1989. Mendoza, EDIUM, 1989, p. 431–440.
- (94) Leguizamón, S. Descripción fractal de texturas naturales. *Jornadas Internacionales de Uso de Computadoras en Actividades Científicas y Tecnológicas*, 2., Mendoza, Argentina, 1991. Mendoza, EDIUM, 1991, v. 2, p. 112–120.
- (95) Lehmann, D.H. *Theory of point estimation*. New York, John Wiley & Sons, 1983.
- (96) Lewis, P.A.; Orav, E.J. *Simulation methodology for statisticians, operation analysts and engineers*. Pacific Grove, CA, Cole Advanced Books & Software, Wadsworth & Books, 1989. v. 1.
- (97) Lim, J.S. *Two-dimensional signal and image processing*. USA, Prentice-Hall International Editions, 1990. 694 p.
- (98) Lopes, A.; Laur, H.; Nezry, E. Statistical distribution and texture in multilook and complex SAR images. In: *International Geoscience and Remote Sensing Symposium*, 10., Washington, DC, May 20–24, 1990. *Remote Sensing Science for the Nineties*. New York, IEEE, 1990, v. 3, p. 2427–2430.

- (99) Lowenthal, S.; Arsenault, H. Image formation for coherent diffuse objects: statistical properties. *Journal of the Optical Society of America*, 60(11):1478–1483, 1970.
- (100) Madsen, S.N. *Speckle theory: modelling, analysis, and applications related to synthetic aperture radar data*. (PhD Thesis) – Technical University of Denmark, 1986.
- (101) Martinelli, F.; Olivieri, E.; Scoppola, E. On the Swendsen-Wang dynamics I: exponential convergence to equilibrium. *Journal of Statistical Physics*, 62(1/2):117–133, 1991a.
- (102) Martinelli, F.; Olivieri, E.; Scoppola, E. On the Swendsen-Wang dynamics II: critical droplets and homogeneous nucleation at low temperature for the 2 dimensional Ising model. *Journal of Statistical Physics*, 62(1/2):135–159, 1991b.
- (103) Mascarenhas, N.D.A. *Estimadores Markovianos alternativos para segmentação de imagens*. São José dos Campos, 1993. Comunicação pessoal.
- (104) Mascarenhas, N.D.A.; Velasco, F.R.D. *Processamento digital de imagens*. Buenos Aires, Editorial Kapelusz, 1989. Escola Brasileira-Argentina de Informática, 4.
- (105) Miller, R.J. *Correlation properties of coherently imaged non-uniform surfaces*. Malvern, England, Royal Signals and Radar Stablishment, 1986.
- (106) Miranda, E. Swendsen-Wang cluster relaxation time in the 3D Ising model. *Physica A*, 175(1/3):235–238, 1991a.
- (107) Miranda, E. Geometrical properties of Swendsen-Wang clusters. *Physica A*, 175(1/3):229–234, 1991b.

- (108) Modestino, J.; Zhang, J. A Markov random field model-based approach to image interpretation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA., 1989. IEEE, 1989, p. 458–465.
- (109) Møller, J. *Lectures on Markov random fields*. Denmark, Department of Theoretical Statistics, Århus University, 1988.
- (110) Müller-Krumbhaar, H.; Binder, K. Dynamic properties of the Monte Carlo method in statistical mechanics. *Journal of Statistical Physics*, 8(1):1–24, 1972.
- (111) NAG. *The NAG FORTRAN Library Manual, Mark-13*. Oxford, s.e., 1988.
- (112) NAG. *The NAG FORTRAN Library Manual, Mark-14*. Oxford, s.e., 1990.
- (113) Nagao, M.; Matsuyama, T. Edge preserving smoothing. *Computer Graphics and Image Processing*, 9(4):394–407, Apr. 1979.
- (114) Nathan, K.S.; Curlander, J.C. Speckle noise reduction of 1-look SAR imagery. In: 1987 International Geoscience and Remote Sensing Symposium (IGARSS'87). Ann Arbor, MI, 18–21 May 1987. *Proceedings*. p. 1457–1462.
- (115) Oliver, C.J. A model for non-Rayleigh scattering statistics. *Optica Acta*, 31:701–722, 1984.
- (116) Oliver, C.J. Correlated K-distributed clutter models. *Optica Acta*, 32:1515–1547, 1985.
- (117) Parisi, G. *Statistical field theory*. Redwood City, CA, Addison Wesley, 1988. 352 p. (Frontiers in Physics).
- (118) Pickard, D.K. Inference for discrete Markov fields: the simplest nontrivial case. *Journal of the American Statistical Association*, 82(1):90–96, 1987.

- (119) Press, W.H; Flannery, B.P.; Teulosky, S.A.; Vetterling, W.T. *Numerical recipes in C: the art of scientific computing*. USA, Cambridge University Press, 1990. 735 p. 2. rep.
- (120) Quegan, S. Interpolation and sampling in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, GRS-28(4):641-646, July 1990.
- (121) Ray, T.; Tamayo, P. Properties of metastable Ising models evolving under the Swendsen-Wang dynamics. *Journal of Statistical Physics*, 60(5-6):851-861, 1990.
- (122) Reingold, E.; Nievergelt, J.; Deo, N. *Combinatorial algorithms: theory and practice*. USA, Prentice-Hall, 1977.
- (123) Rignot, E; Chellappa, R. Segmentation of synthetic-aperture-radar complex data. *Journal of the Optical Society of America A: Optics and Image Science*, 8(9):1499-1509, Sep. 1991.
- (124) Ripley, B.D. The use of spatial models as image priors. In: Possolo, A., ed. *Spatial statistics and imagery*. s.l., s.ed., 1989. (IMS Lecture Notes).
- (125) Sant'Anna, S.J.S. Avaliação de filtros redutores de *speckle*. (Tese de Mestrado em Sensoriamento Remoto) - INPE, São José dos Campos. A ser apresentada.
- (126) Sagan, C. *Cosmos*. Barcelona, Espanha, Editorial Planeta, 1980. 366 p.
- (127) Sher, D. Minimizing the cost of errors with a Markov random field. *Pattern Recognition Letters*, 12(2): 85-89, 1991.
- (128) Scholefield, P.H.R. Statistical aspects of ideal radar targets. *Proceedings IEEE*, 55(4):587-590, Apr. 1967.
- (129) Sperry, R.H; Parker, K.J. Segmentation of speckled images based on level-crossing statistics. *Journal of the Optical Society of America A: Optics and Image Science*, 8(3):490-498, Mar. 1991.

- (130) Swendsen, R. Acceleration methods for Monte Carlo computer simulations. *Computer Physics Communications*, 65(1/3):281–288, 1991.
- (131) Swendsen, R.; Wang, J. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, 1987.
- (132) The MathWorks *External interface guide*. Natick, MA, 1993.
- (133) Torreão, J.R.A. *Métodos estocásticos em computação visual*. Porto Alegre, Instituto de Informática da UFRGS, Taisy Silva Weber, 1992. 148 p. Notas de curso da Escola de Computação, 8.
- (134) Tur, M.; Chin, K.C.; Goodman, J.W. When is speckle noise multiplicative? *Applied Optics*, 21(7):1157–1159, 1982.
- (135) Vetterling, W.T.; Teulosky, S.A.; Press, W.H.; Flannery, B.P.; *Numerical recipes: example book (C)*. USA, Cambridge University Press, 1990. 239 p. 2. rep.
- (136) Vinhas, L.; Frery, A.C. Classificação por máxima verossimilhança e por redes neurais: um estudo Monte Carlo. In: Congresso Nacional de Matemática Aplicada e Computacional, 16., Uberlândia, MG, 6–9 set. 1993. *Resumo das comunicações*. Uberlândia, SBMAC, 1993, p. 131.
- (137) Wolff, U. Collective Monte Carlo updating for spin systems. *Physical Review Letters*, 62(4):361–364, 1989a.
- (138) Wolff, U. Comparison between cluster Monte Carlo algorithms in the Ising model. *Physics Letters B*, 228(3):379–382, 1989b.
- (139) Yanasse, C.C.F. *Statistical analysis of synthetic aperture radar images and its applications to system analysis and change detection*. (PhD Thesis) – University of Sheffield, Sheffield, UK, Dec. 1991. 318 p.
- (140) Yanasse, C.C.F.; Frery, A.C.; Sant’Anna, S.J.S.; Dutra, L.V. Statistical analysis of SAREX data over Tapajós – Brazil. Final Results Workshop For Sarex-92. Paris, 1993.

- (141) Yuille, A.; Vincent, L.; Geiger, D. Statistical morphology. In: *Image Algebra and Morphological Image Processing I*, Billingham, WS, SPIE, 1991. p. 271–282. (SPIE Proceeding, v. 1568).
- (142) Zito, R.R. The shape of SAR histograms. *Computer Vision, Graphics and Image Processing*, 43:(3)281–293, Sept. 1988.

APÊNDICE A

INFORMAÇÃO COMPUTACIONAL

A.1 Geradores de Números Pseudoaleatórios

No Capítulo 3 foram vistos vários algoritmos que envolvem o uso de ocorrências de variáveis aleatórias com diferentes distribuições. Portanto, para implementá-los, é necessário contar com algoritmos que forneçam tais ocorrências.

O problema de geração de números pseudoaleatórios por computador é quase tão antigo como os computadores digitais. Parte do interesse que o assunto tem despertado na comunidade da Ciência da Computação deve-se, provavelmente, ao fato da geração destes números ir de encontro ao paradigma de máquina de calcular: deseja-se ter comportamento aleatório de uma engenhoca que *por definição* é determinística. Outra motivação, não menos intensa do que a comentada, é que muitas áreas requerem o uso de números que apresentam comportamento similar ao de ocorrências de variáveis aleatórias. Para uma discussão sobre estas e outras motivações, o leitor pode consultar o texto (23) e as referências aí indicadas.

Para as implementações apresentadas neste trabalho, inicialmente usou-se o gerador descrito em (19). Este gerador mostrou-se extremamente rápido, porém, em alguns casos gerou amostras uma estrutura de correlação indesejada. Além disso, em (34) mostra-se que as amostras por ele geradas são rejeitadas pelos testes Kolmogorov-Smirnov e χ^2 , aos níveis 90%, 95% e 99% mais vezes do que o esperado. Foi, então, obtida a biblioteca de subrotinas RANLIB (cuja documentação pode ver-se em (15)) através do nodo de distribuição STATLIB (disponível no endereço eletrônico statlib@lib.stat.cmu.edu). Esta biblioteca oferece um leque bastante largo de rotinas (nas linguagens de programação FORTRAN e C), em forma gratuita e com a única condição de se citar a fonte dos algoritmos empregados.

Os dois algoritmos mais empregados, nesta tese, foram o de

geração de ocorrências de variáveis aleatórias uniformes (`genunf` ver (90)) e o de ocorrências de variáveis aleatórias normais k -variadas (`genmn` (2)).

A.2 Aplicativos

Este trabalho foi redigido com o uso da linguagem descritora de documentos \LaTeX (89), procurando aproximar o formato exigido (70).

Alguns dos aplicativos que possibilitaram diferentes partes deste trabalho foram:

MatLab: nas versões 3.5h e 4. Foi empregado para cálculos numéricos e para apresentação de gráficos (funções, histogramas, etc.). Tal como já foi discutido, a Versão 4 teve um papel central no desenvolvimento do programa que efetua os testes Kolmogorov-Smirnov e χ^2 , pois permitiu a implementação da função de distribuição acumulada da distribuição 2.12.

SAO e Khoros: foram empregados para a visualização de imagens. O último forneceu um ambiente amigável para a ilustração da extensão do Teorema de Geman e Geman.

APÊNDICE B

LISTAGENS DE ALGUNS DOS PROGRAMAS IMPLEMENTADOS

B.1 Programa para Simulação do Modelo de Potts-Strauss e Contaminação Rayleigh com Número Arbitrário de Visadas e Detecção Linear

```

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <limits.h>
#include "pseudo.h"

/*
 * Programa "regs+multispeckle.c"
 *
 * Proposito: gerar um arquivo de regioes (ocorrencia do modelos de
 * Potts-Strauss para um numero arbitrario de classes) e um arquivo de
 * speckle (a imagem observada, dada a ocorrencia de regioes, para o modelo
 * multiplicativo de ruido speckle).
 *
 * Gerador de uniformes utilizado: "triplo"
 *
 * INPE: 22/9/92
 *
 * Alejandro C. Frery
 */

#define K 128
#define KM1 129
#define KM2 130
#define K_1 127
#define FORA -100
#define saturacao 3.1

/* variaveis globais */

unsigned      grafomax = 0, n_class;

double       teste;

struct PIXEL {
unsigned char  valor;
int           compnum;
}             imagem[KM2][KM2];

/* funcoes */

void          comp(unsigned i, unsigned j);
void          use_comp(void);
void          troca_grafo(void);
unsigned char unif_n(void);

/* programa principal */

```

```

main()
{
unsigned      i, j, n_looks, k, N;

char          campo_regioes[20], campo_speckle[20];

FILE          *f_campo_regioes, *f_campo_speckle;

unsigned char  c, v;

double        temp, raighley, *sigmas;

/* corpo */

printf("\nIngresse as tres sementes para o gerador [0<=semente<=%d]:\n",
SHRT_MAX);
scanf("%d %d %d", &triplo_seed1, &triplo_seed2, &triplo_seed3);

printf("\nIngresse o numero desejado de classes: ");
scanf("%u", &n_class);
sigmas = (double *) malloc((unsigned) n_class * sizeof(double));

printf("\nO valor de beta critico para este modelo e %2.12lf", log(1. +
sqrt(n_class)));

printf("\nIngresse o nome do arquivo destino do campo de regioes: ");
scanf("%s", campo_regioes);
if ((f_campo_regioes = fopen(campo_regioes, "wb")) == NULL) {
printf("\n\n>>ERRO ABRINDO ARQUIVO<<");
exit(0);
}
printf("\nIngresse o nome do arquivo destino do campo com speckle: ");
scanf("%s", campo_speckle);
if ((f_campo_speckle = fopen(campo_speckle, "wb")) == NULL) {
printf("\n\n>>ERRO ABRINDO ARQUIVO<<");
exit(0);
}
printf("\nIngresse beta ");
scanf("%lf", &teste);
teste = 1. - exp(-teste);
printf("\nProbabilidade de formar arco = %lf\n", teste);

printf("\nIngresse os %d valores dos back-scatters medios das classes em
ordem crescente [ 0 < sigma <= 1. ]: ", n_class);
for (i = 0; i < n_class; i++) {
printf("\nBack-scatter da classe %u: ", i);
scanf("%lf", sigmas + i);
}

printf("\nIngresse o numero de visadas: ");
scanf("%u", &n_looks);

/* inicializador da matriz */

for (i = 0; i < KM2; i++)
for (j = 0; j < KM2; j++) {
if (i != 0 && i != KM1 && j != 0 && j != KM1) {
imagem[i][j].valor = unif_n();
imagem[i][j].compnum = 0;
} else {
imagem[i][j].valor = UCHAR_MAX;
imagem[i][j].compnum = FORA;
}
}

/* realiza o numero de iteracoes desejadas */

```

```

printf("\nEntre com o numero de iteracoes [ < %d ] ", INT_MAX);
scanf("%d", &N);

for (k = 1; k <= N; k++) {
printf(" %d", k);

grafomax = 0;
use_comp();
troca_grafo();
}

printf("\nGravando os campos...");
double      conversao1 = 255. / (double) (n_class - 1);
double      conversao2 = 255. / saturacao;

for (i = 1; i < KM1; i++)
for (j = 1; j < KM1; j++) {
v = imagem[i][j].valor;
c = (unsigned char) floor(conversao1 * (double) v);
fwrite((char *) &c, sizeof(unsigned char), 1,
f_campo_regioes);

temp = 0.;

for (k = 0; k < n_looks; k++) {
while ((raighley = (double) triplo()) == 0.);
temp += sqrt(-2. * log(raighley));
}

temp *= (*(sigmas + (unsigned) v) / (double) n_looks);

if (temp < saturacao)
c = (unsigned char) floor(temp * conversao2);
else
c = (unsigned char) 255;

fwrite((char *) &c, sizeof(unsigned char), 1,
f_campo_speckle);
}

fclose(f_campo_regioes);
fclose(f_campo_speckle);
printf("\nO estado final das sementes e: %d %d %d\n", triplo_seed1,
triplo_seed2, triplo_seed3);
printf("FIM\n");

return (0);
} /* main */

/*****
/*****Funcoes de operacao sobre grafos e memoria*****/

void
comp(unsigned i, unsigned j)
{
unsigned char  val;

val = imagem[i][j].valor;
imagem[i][j].compnum = grafomax;

if (imagem[i - 1][j].valor == val && imagem[i - 1][j].compnum == 0 &&
triplo() < teste)
comp(i - 1, j);
if (imagem[i + 1][j].valor == val && imagem[i + 1][j].compnum == 0 &&

```

```

triplo() < teste)
comp(i + 1, j);
if (imagem[i][j - 1].valor == val && imagem[i][j - 1].compnum == 0 &&
    triplo() < teste)
comp(i, j - 1);
if (imagem[i][j + 1].valor == val && imagem[i][j + 1].compnum == 0 &&
    triplo() < teste)
comp(i, j + 1);

return;
} /* comp */

/*-----*/

void
troca_grafo(void)
{

/*
 * esta subrotina faz o teste para verificar se o valor do pixel no
 * grafo deve ser trocado
 */

unsigned      i, j;
unsigned char *ltroca;

ltroca = (unsigned char *) malloc(grafomax * sizeof(unsigned char));
if (ltroca == NULL) {
printf("\nErro em ltroca");
exit(0);
}
for (i = 0; i < grafomax; i++)
ltroca[i] = unif_n();

/* testa o valor da lista e executa a troca */

for (i = 1; i < KM1; i++)
for (j = 1; j < KM1; j++) {
/* troca o valor do pixel e libera o pixel do grafo */

imagem[i][j].valor = ltroca[(unsigned)
    (imagem[i][j].compnum - 1)];
imagem[i][j].compnum = 0;
}
free((char *) ltroca);
return;
} /* troca_grafo */

/*-----*/

void
use_comp(void)
{
unsigned      i, j;

for (i = 1; i < KM1; i++)
for (j = 1; j < KM1; j++)
if (imagem[i][j].compnum == 0) {
grafomax++;
comp(i, j);
}
return;
} /* use_comp */

/*-----*/

unsigned char
unif_n(void)
{

```

```
return ((unsigned char) floor((double) n_class * triplo()));
}
```

B.2 Programa para Segmentação em um Número Arbitrário de Classes, pelo Algoritmo ICM, de Imagens com Contaminação Rayleigh, uma Visada e Detecção Linear

```
/*
 * Programa: ICMest Proposito: obter o estimador ICM de uma imagem SAR
 * monoespectral a partir da segmentacao por maxima verossimilhanca estimando
 * o parametro da distribuicao a priori e os parametros decorrentes de cada
 * segmentacao em forma iterativa. Os estimadores utilizados sao MV, MD e
 * MAD. Para cada iteracao sao supostos como corretos os "mais afastados".
 * Autor: Alejandro C. Frery. INPE 31/3/92 - 8/5/92.
 *
 * Modificacao # 1: deixar de usar o re-escalamento a valores reais. A
 * estimacao, etc., sera feita sobre valores "unsigned char" toda vez que
 * possivel.
 *
 * 15-10-92
 *
 * Modificacao # 2: tirar o "mergulho" da imagem segmentada, que e heranca dos
 * algoritmos de simulacao; trabalhar em forma vetorial para qualquer tamanho
 * de imagem.
 *
 * 15-10-92
 *
 * Modificacao # 3: deixar somente a estimacao por maxima verossimilhanca
 *
 * 20-10-92
 *
 * Modificacao # 4: incorporar o estimador de maxima pseudoverossimilhanca para
 * um numero arbitrario de classes. O algoritmo foi desenvolvido e testado no
 * programa "pseudo_potts.c"
 *
 * 21-10-92
 *
 * Modificacao # 6: incorporar todas as mudancas para ter o algoritmo ICM para
 * um numero arbitrario de classes
 *
 * 22-10-92
 *
 * Modificacao # 7: eliminar o maior numero possivel de variaveis globais
 *
 * 26-10/92
 *
 * Versao operacional: 29-10-92
 *
 * Modificacao # 7a: o algoritmo para quando um numero inferior a um certo
 * numero especificado de posicoes muda de uma iteracao para a seguinte
 */
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <malloc.h>
#include <limits.h>
#include <values.h>
#include <time.h>

#define sigmo( a ) ( a / ( 1 + a ) )

/* Declaracao de funcoes */
```

```

int      compara_uchar(void *x, void *y);
int      compara_double(void *x, void *y);
int      compara_uchar-vesre(void *x, void *y);
unsigned char  segmenta(unsigned char val, double *y);
double   rtbis(double (*func) (double x, double *param), double *param);
double   plikehood_2(double search, double *param);
double   plikehood_3(double search, double *param);
double   plikehood_4(double search, double *param);
void     estimacao_geral(unsigned char *x, unsigned n_filas,
    unsigned n_colunas);
double   estimacao_global(unsigned char *x, unsigned n_class,
    unsigned n_filas, unsigned n_colunas);
double   estimates_2(unsigned char *x, unsigned n_filas,
    unsigned n_colunas);
double   estimates_3(unsigned char *x, unsigned n_filas,
    unsigned n_colunas);
double   estimates_4(unsigned char *x, unsigned n_class,
    unsigned n_filas, unsigned n_colunas);
unsigned char  atualiza_posicao(unsigned char *x, unsigned n_class,
    unsigned n_colunas, unsigned i, unsigned j,
    unsigned char *dados_observados,
    double *backscatters);

/* Declaracao de variaveis globais */
double     pp[10];
double     beta_est = 0.;
time_t     tempo_i, tempo_f;

main()
{
    unsigned     i, j, k, troca, contador_iteracoes = 0, *contador,
        opcao, ncm1, n_filas, n_colunas, n_class, temporario,
        min_pontos;

    double       *xi, *som_quad, beta_max, *probs_novos_valores,
        percentagem;

    char         imag_radar[20], maxver[20], ICM[20];

    unsigned char  uc, ucn, *x, *y, anterior, novo;

    FILE         *imagem_entrada, *f_maxver, *f_ICM;

    printf("\nIngresse, na ordem, o numero de filas e de colunas da imagem a
    ser processada: ");
    scanf("%u %u", &n_filas, &n_colunas);
    x = (unsigned char *) malloc(sizeof(unsigned char) * n_filas *
    n_colunas);
    y = (unsigned char *) malloc(sizeof(unsigned char) * n_filas *
    n_colunas);
    if (x == NULL || y == NULL) {
        printf("\n>>ERRO ALOCANDO MEMORIA PARA \"x\" ou \"y\"\nPROGRAMA
        ABORTADO<<\n");
        exit(0);
    }
    printf("\nIngresse a percentagem de pontos minima a ser trocada para o
    algoritmo prosseguir: ");
    scanf("%lf", &percentagem);
    printf("\nO algoritmo parara quando forem trocados %u ou menos
    posicoes", min_pontos = (unsigned) floor((percentagem * n_filas * n_colunas /
    100.)));

    printf("\nIngresse o numero desejado de classes: ");
    scanf("%u", &n_class);
    pp[9] = n_class;
    ncm1 = n_class - 1;
    printf("\nO valor de beta critico para este modelo de Potts-Strauss e

```

```

    %lf", log(1. + sqrt((double) n_class)));
    contador = (unsigned *) malloc(sizeof(unsigned) * n_class);
    xi = (double *) malloc(sizeof(double) * n_class);
    som_quad = (double *) malloc(sizeof(double) * n_class);
    probs_novos_valores = (double *) malloc(sizeof(double) * (n_class + 1));
    if (contador == NULL || xi == NULL || som_quad == NULL ||
        probs_novos_valores == NULL) {
        printf("\n>>ERRO ALOCANDO MEMORIA: para \"contador\", \"xi\",
            \"som_quad\" ou \"probs_novos_valores\"\nPROGRAMA ABORTADO<<\n");
        exit(0);
    }
    printf("\nIngresse os %d valores dos back-scatters medios das classes em
        ordem crescente: ", n_class);
    for (i = 0; i < n_class; i++) {
        printf("\nBack-scatter da classe %u: ", i);
        scanf("%lf", xi + i);
    }
    /*
     * calculo dos valores de segmentacao max-ver usando xi: em
     * probs_novos_valores estarao guardados os limiares para a
     * segmentacao
     */
    *(probs_novos_valores) = 0.;
    for (j = 0; j < ncm1; j++) {
        *(probs_novos_valores + j + 1) = 2. * sqrt(log(xi[j + 1] /
            xi[j]) / (pow(xi[j], -2.) - pow(xi[j + 1], -2.)));
        printf("\ny(%u) = %lf", j + 1, *(probs_novos_valores + j + 1));
    }

    *(probs_novos_valores + n_class) = MAXDOUBLE;

    printf("\nOpcoes para a estimacao de backscatters:
        \n\t[0] = SEM atualizacao de beta e NEM de backscatters,
        \n\t[1] = COM atualizacao de beta e SEM de backscatters,
        \n\t[2] = SEM atualizacao de beta e COM de backscatters,
        \n\t[3] = COM atualizacao de beta e de backscatters...");
    printf("\nIngresse a sua escolha: ");
    scanf("%d", &opcao);
    if (opcao == 0 || opcao == 2) {
        printf("\nIngresse o valor desejado de beta para todas as
            iteracoes: ");
        scanf("%lf", &beta_est);
    } else {
        printf("\nIngresse o valor maximo de beta que sera utilizado
            (ou um valor negativo qualquer para nao ter limitante superior): ");
        scanf("%lf", &beta_max);
        beta_max = (beta_max > 0. ? beta_max : MAXDOUBLE);

        printf("\nIngresse o minimo onde o zero da pseudo sera procurado
            ");
        scanf("%lf", pp);
        printf("\nIngresse o maximo onde o zero da pseudo sera procurado
            ");
        scanf("%lf", pp + 1);
        printf("\nIngresse a tolerancia com que o zero da pseudo sera
            procurado ");
        scanf("%lf", pp + 2);
    }

    loop2:
    printf("\nIngresse o nome do arquivo da imagem SAR: ");
    scanf("%s", imag_radar);
    imagem_entrada = fopen(imag_radar, "rb");
    while (imagem_entrada == NULL) {
        printf("\n>>ERRO<< Nao foi possivel abrir o arquivo");
        goto loop2;
    }
}

```

```

printf("\nIngresse o nome do arquivo destino da estimacao por maxima
verossimilhanca: ");
scanf("%s", maxver);
f_maxver = fopen(maxver, "wb");

printf("\nIngresse o nome do arquivo destino da estimacao ICM: ");
scanf("%s", ICM);
f_ICM = fopen(ICM, "wb");

/* Condicoes iniciais: a segmentacao max-ver */
for (j = 0; j < n_colunas; j++) {
for (i = 0; i < n_filas; i++) {

if (fread((char *) &uc, sizeof(unsigned char), 1,
imagem_entrada) == 0) {
printf("\n\n>>ERRO FATAL DE LEITURA DA
IMAGEM<<");
exit(0);
}
y[j + i * n_colunas] = uc;

/*
* segmentacao por max-ver usando os back - scatters
* medios. Re-escalamento dos valores segmentados
* para guardar em arquivo.
*/
x[j + i * n_colunas] = ucn = segmenta(uc,
probs_novos_valores);
uc = ucn * 255 / ncm1;
/*
* escrita no arquivo de maxima verossimilhanca dos
* valores de segmentacao por maxima verossimilhanca
*/
fwrite((char *) &uc, sizeof(unsigned char), 1,
f_maxver);
}
}
printf("\n>>Fim segmentacao max-ver, inicio do ICM<<\n");

fclose(f_maxver);
fclose(imagem_entrada);

/*
* loop de segmentacao por ICM. Parara quando nao houver mais mudanca
* de valores. Os valores de beta e dos parametros sao estimados
* antes de cada passo de atualizacao.
*/
time(&tempo_i);
do {
/*
* estimacao atualizada do beta baseada na segmentacao atual,
* e calculos associados
*/
if (opcao == 1 || opcao == 3) {
beta_est = estimacao_global(x, n_class, n_filas,
n_colunas);
beta_est = (beta_est < beta_max ? beta_est : beta_max);
}
printf("\nValor empregado de beta = %lf", beta_est);

/*
* bloco de atualizacao das estimacoes dos backscatters: sera
* ativado somente se for indicado pelo usuario
*/
if (opcao == 2 || opcao == 3) {

```

```

/*
 * estimacao atualizada dos backscatters baseada na
 * segmentacao atual, e calculos associados
 */
for (k = 0; k < n_class; k++) { /* inicializacao dos
 * acumuladores e dos
 * contadores */
som_quad[k] = 0.;
contador[k] = 0;
}

for (j = 0; j < n_colunas; j++) {
for (i = 0; i < n_filas; i++) {
uc = y[j + i * n_colunas];
/* uc contera a observacao */
ucn = x[j + i * n_colunas];
/* ucn contera a classe a qual foi atribuida a observacao */
som_quad[(unsigned) ucn] +=
(double) (uc * uc);
contador[(unsigned) ucn] += 1;
}
}

for (k = 0; k < n_class; k++) {
if ((temporario = contador[k]) > 0) {
/* atualiza o valor estimado do parametro desde que haja observacoes
atribuidas a essa classe; caso contrario mantem o valor estimado atual
*/
xi[k] = sqrt(som_quad[k] / (2. *
(double) temporario));
}
printf("\nValor atual do backscatter na classe
%d = %lf (%d observacoes)", k, xi[k], contador[k]);
}

}
troca = 0;

/* (1) */
for (j = 1; j < n_colunas - 1; j += 2) {
for (i = 1; i < n_filas - 1; i += 2) {

anterior = x[j + i * n_colunas];
novo = atualiza_posicao(x, n_class, n_colunas,
i, j, y, xi);
if (novo != anterior) {
troca++;
x[j + i * n_colunas] = novo;
}
}
}

/* (2) */
for (j = 2; j < n_colunas; j += 2) {
for (i = 2; i < n_filas; i += 2) {

anterior = x[j + i * n_colunas];
novo = atualiza_posicao(x, n_class, n_colunas,
i, j, y, xi);
if (novo != anterior) {
troca++;
x[j + i * n_colunas] = novo;
}
}
}
}

```

```

/* (3) */
for (j = 1; j < n_colunas - 1; j += 2) {
for (i = 2; i < n_filas; i += 2) {

anterior = x[j + i * n_colunas];
novo = atualiza_posicao(x, n_class, n_colunas,
    i, j, y, xi);
if (novo != anterior) {
troca++;
x[j + i * n_colunas] = novo;
}
}
}

/* (4) */
for (j = 2; j < n_colunas; j += 2) {
for (i = 1; i < n_filas - 1; i += 2) {

anterior = x[j + i * n_colunas];
novo = atualiza_posicao(x, n_class, n_colunas,
    i, j, y, xi);
if (novo != anterior) {
troca++;
x[j + i * n_colunas] = novo;
}
}
}

/* as quatro bordas e as esquinas estao faltando!!! */

printf("\nIteracao numero %u completa com %u trocas,
    entrando na proxima...", ++contador_iteracoes, troca);

} while (troca >= min_pontos);
time(&tempo_f);
printf("\nO tempo de CPU empregado para a segmentacao ICM foi de %d
    segundos", tempo_f - tempo_i);

for (j = 0; j < n_colunas; j++) {
for (i = 0; i < n_filas; i++) {
uc = x[j + i * n_colunas] * 255 / ncmi;

/* escrita no arquivo de ICM dos valores */
fwrite((char *) &uc, sizeof(unsigned char), 1, f_ICM);

}
}

/* fechamento do arquivo de saida e liberacao dos vetores
    empregados */
fclose(f_ICM);
free((char *) x);
free((char *) y);
free((char *) contador);
free((char *) xi);
free((char *) som_quad);
free((char *) probs_novos_valores);

printf("\n>>FIM<<\n");

return (0);
} /* main */

/*
* -----
*/

```

```

unsigned char
segmenta(unsigned char val, double *y)
{
    unsigned char    i = 0;
    double           s = *(y);

    while (s < (double) val) {
        i++;
        s += *(y + i);
    }

    return (i - 1);
}

/*
 * -----
 */

double
estimates_2(unsigned char *x, unsigned n_filas, unsigned n_colunas)
{
    int             val, viz;
    unsigned        i, j;
    int             cm1m4 = 0, c1m4 = 0, cm1m2 = 0, c1m2 = 0, cm12 = 0,
                  c12 = 0, cm14 = 0, c14 = 0;
    double          max = -MAXDOUBLE;
    *(pp + 3) = *(pp + 4) = *(pp + 5) = *(pp + 6) = *(pp + 7) = 0.;

    for (i = 1; i < n_filas - 1; i++) {
        for (j = 1; j < n_colunas - 1; j++) {
            val = (int) x[j + i * n_colunas];
            viz = (int) x[j + (i - 1) * n_colunas] +
                  (int) x[j + (i + 1) * n_colunas] +
                  (int) x[(j - 1) + i * n_colunas] +
                  (int) x[(j + 1) + i * n_colunas];

            if (val == 0) {
                switch (viz) {
                    case 0:
                        cm1m4++;
                        break;
                    case 1:
                        cm1m2++;
                        break;
                    case 3:
                        cm12++;
                        break;
                    case 4:
                        cm14++;
                        break;
                }
            } else {
                switch (viz) {
                    case 4:
                        c14++;
                        break;
                    case 3:
                        c12++;
                        break;
                    case 1:
                        c1m2++;
                        break;
                    case 0:
                        c1m4++;
                        break;
                }
            }
        }
    }
}

```

```

}
*(pp + 3) = (double) (2 * (c14 - c1m4) + c12 - c1m2);
*(pp + 4) = (double) (2 * (cm1m4 + c1m4));
*(pp + 5) = (double) (cm1m2 + c1m2);
*(pp + 6) = (double) (cm12 + c12);
*(pp + 7) = (double) (2 * (cm14 + c14));
}
}

/*
 * padronizacao das constantes envolvidas na equacao de
 * pseudo-verossimilhanca dividindo pelo valor maximo
 */
for (i = 3; i < 8; i++) {
max = (*(pp + i) > max ? *(pp + i) : max);
}
for (i = 3; i < 8; i++) {
*(pp + i) /= max;
}
/* fim padronizacao */

return (rtbis(plikehood_2, pp));
}
/* estimates */
/*
 * -----
 */

double
plikehood_2(double search, double *pp)
{
double t2, t4, tm2, tm4;
t2 = exp(2 * (double) search);
t4 = t2 * t2;
tm2 = 1 / t2;
tm4 = 1 / t4;
return (*(pp + 3) + *(pp + 4) * sigmo(tm4) + *(pp + 5) * sigmo(tm2) -
*(pp + 6) * sigmo(t2) - *(pp + 7) * sigmo(t4));
} /* plikehood */

/*
 * -----
 */

double
rtbis(double (*func) (double x, double *param), double *param)
/*
 * Metodo da biseccao para achar os zeros de uma funcao de dupla precisao com
 * parametros. Numerical Recipes in C cap 9. Versao del 22 de agosto de 1990.
 * INPE. O vetor (double)param contem os seguintes valores: param[0] = o
 * minimo do intervalo de procura param[1] = o maximo do intervalo de procura
 * param[2] = a tolerancia param[3] ... param[n] = parametros da funcao
 */
{
double dx, xmid;
int j;
double f = (*func) (*param, param), rtb;
double fmid = (*func) (*(param + 1), param);
if (f * fmid >= 0.)
printf("\n0 intervalo deve incluir a raiz\n");
rtb = f < 0. ? (dx = *(param + 1) - *param, *param) : (dx = *param -
*(param + 1), *(param + 1));
for (j = 1; j <= MAXINT; j++) {
fmid = (*func) (xmid = rtb + (dx *= .5), param);
if (fmid <= 0.)
rtb = xmid;
}
}

```

```

if (fabs(dx) < *(param + 2) || fmid == 0.)
return (rtb);
}
printf("\nUltrapassou o maximo de iteracoes permitido\n");
return *(param + 1));
}

/*
-----
*/

int
compara_uchar(void *x, void *y)
{
if (*(unsigned char *) x < *(unsigned char *) y)
return (-1);
if (*(unsigned char *) x > *(unsigned char *) y)
return (1);
if (*(unsigned char *) x == *(unsigned char *) y)
return (0);
} /* compara_uchar */

/*
-----
*/

int
compara_double(void *x, void *y)
{
if (*(double *) x < *(double *) y)
return (-1);
if (*(double *) x > *(double *) y)
return (1);
if (*(double *) x == *(double *) y)
return (0);
} /* compara_uchar */

/*
-----
*/

int
compara_uchar_vesre(void *x, void *y)
{
if (*(unsigned char *) x < *(unsigned char *) y)
return (1);
if (*(unsigned char *) x > *(unsigned char *) y)
return (-1);
if (*(unsigned char *) x == *(unsigned char *) y)
return (0);
}

/*
-----
*/

double
estimates_3(unsigned char *x, unsigned n_filas, unsigned n_colunas)
{
unsigned      i, j, k, v_400, v_310, v_220, v_211;
unsigned char  valores[3];

estimacao_geral(x, n_filas, n_colunas); /* esta funcao calcula a
 * parte comum para os casos
 * com mais de dois estados
 * possiveis; o resultado e
 * guardado em pp[3] */

```

```

v_400 = v_310 = v_220 = v_211 = 0;

for (i = 1; i < n_filas - 1; i++) {
for (j = 1; j < n_colunas - 1; j++) {

for (k = 0; k < 3; k++) {
valores[k] = 0;
}

valores[(unsigned) x[j + (i - 1) * n_colunas]] += 1;
valores[(unsigned) x[j + (i + 1) * n_colunas]] += 1;
valores[(unsigned) x[(j - 1) + i * n_colunas]] += 1;
valores[(unsigned) x[(j + 1) + i * n_colunas]] += 1;

qsort(valores, 3, sizeof(unsigned char),
      compara_uchar_vesre);

if (valores[0] == 4) {
v_400++;
} else if (valores[0] == 3) {
v_310++;
} else if (valores[0] == valores[1]) {
v_220++;
} else {
v_211++;
}
}
}

pp[4] = (double) v_400;
pp[5] = (double) v_310;
pp[6] = (double) v_220;
pp[7] = (double) v_211;

return (rtbis(plikehood_3, pp));
}

/*
 * -----
 */

double
estimates_4(unsigned char *x, unsigned n_class, unsigned n_filas,
            unsigned n_colunas)
{
    unsigned i, j, k, v_400 = 0, v_310 = 0, v_220 = 0, v_211 = 0,
            v_111 = 0;
    unsigned char *valores;

    if ((valores = (unsigned char *) malloc(sizeof(unsigned char) *
        n_class)) == NULL) {
        printf("\n>>ERRO ALOCANDO MEMORIA EM FUNCAO
        \\"estimates_4\\"<<\nPROGRAMA ABORTADO\n\n");
        exit(0);
    }
    estimacao_geral(x, n_filas, n_colunas);

    for (i = 1; i < n_filas - 1; i++) {
    for (j = 1; j < n_colunas - 1; j++) {

    for (k = 0; k < n_class; k++) {
    valores[k] = 0;
    }

    valores[(unsigned) x[j + (i - 1) * n_colunas]] += 1;
    valores[(unsigned) x[j + (i + 1) * n_colunas]] += 1;

```

```

valores[(unsigned) x[(j - 1) + i * n_colunas]] += 1;
valores[(unsigned) x[(j + 1) + i * n_colunas]] += 1;

qsort(valores, n_class, sizeof(unsigned char),
      compara_uchar_vesre);

if (valores[0] == 4) {
v_400++;
} else if (valores[0] == 3) {
v_310++;
} else if (valores[0] == valores[1]) {
v_220++;
} else if (valores[0] == 2) {
v_211++;
} else {
v_111++;
}
}

free((char *) valores);

pp[4] = (double) v_400;
pp[5] = (double) v_310;
pp[6] = (double) v_220;
pp[7] = (double) v_211;
pp[8] = (double) v_111;

return (rtbis(plikehood_4, pp));
}

/*
 * -----
 */

double
estimacao_global(unsigned char *x, unsigned n_class, unsigned n_filas,
                unsigned n_colunas)
{
if (n_class == 2) {
return (estimates_2(x, n_filas, n_colunas));
} else if (n_class == 3) {
return (estimates_3(x, n_filas, n_colunas));
} else {
return (estimates_4(x, n_class, n_filas, n_colunas));
}
}

/*
 * -----
 */

double
plikehood_3(double search, double *pp)
{
double          t1, t2, t3, t4;

t1 = exp(search);
t2 = t1 * t1;
t3 = t2 * t1;
t4 = t2 * t2;
return (pp[3] - ((pp[4] * 4. * t4 / (2. + t4)) +
(pp[5] * ((3. * t3 + t1) / (t3 + t1 + 1.))) +
(pp[6] * 4. * t2 / (2. * t2 + 1.)) + (pp[7] * (2. * t1 + 2.) / (t1 + 2.))));
}

```

```

/*
 * -----
 */

double
plikehood_4(double search, double *pp)
{ /* o numero de classes e passado na posicao
 * "10" do vetor pp */
double      t1, t2, t3, t4;

t1 = exp(search);
t2 = t1 * t1;
t3 = t2 * t1;
t4 = t2 * t2;

return (pp[3] - ((pp[4] * 4. * t4 / ((pp[9] - 1.) + t4)) +
(pp[5] * ((3. * t3 + t1) / (t3 + t1 + (pp[9] - 2.)))) +
(pp[6] * 4. * t2 / (2. * t2 + (pp[9] - 2.))) +
(pp[7] * (2. * t2 + 2. * t1) / (t2 + 2. * t1 + (pp[9] - 3.)) +
(pp[8] * 4. * t1 / (4. * t1 + (pp[9] - 4.))))));
}

/*
 * -----
 */

void
estimacao_geral(unsigned char *x, unsigned n_filas, unsigned n_colunas)
{
/*
 * calculo da parte que nao envolve o valor de beta na equacao de
 * pseudoverossimilhanca
 */
pp[3] = 0.;
unsigned char  v_ij;
unsigned       i, j;

for (i = 1; i < n_filas - 1; i++) {
for (j = 1; j < n_colunas - 1; j++) {
v_ij = x[j + i * n_colunas];
if (x[j + (i - 1) * n_colunas] == v_ij)
pp[3]++;
if (x[j + (i + 1) * n_colunas] == v_ij)
pp[3]++;
if (x[(j - 1) + i * n_colunas] == v_ij)
pp[3]++;
if (x[(j + 1) + i * n_colunas] == v_ij)
pp[3]++;
}
}
}

/*
 * -----
 */

unsigned char
atualiza_posicao(unsigned char *x, unsigned n_class, unsigned n_colunas,
unsigned i, unsigned j, unsigned char *dados_observados, double *backscatters)
{

unsigned char  novo, *valores_vizinhos;
unsigned       k;
double         probab, *novos_valores;

if ((valores_vizinhos = (unsigned char *) calloc(n_class,
sizeof(unsigned char))) == NULL || (novos_valores = (double *)

```

```

    malloc(sizeof(double) * n_class)) == NULL) {
printf("\n>>ERRO ALOCANDO MEMORIA EM
  \'atualiza_posicao\'\nPROGRAMA ABORTADO<<\n");
exit(0);
}
valores_vizinhos[(unsigned) x[j + (i - 1) * n_colunas]] += 1;
valores_vizinhos[(unsigned) x[j + (i + 1) * n_colunas]] += 1;
valores_vizinhos[(unsigned) x[(j - 1) + i * n_colunas]] += 1;
valores_vizinhos[(unsigned) x[(j + 1) + i * n_colunas]] += 1;

for (k = 0; k < n_class; k++) {
novos_valores[k] = pow(backscatters[k], -2.) * exp(beta_est *
(double) valores_vizinhos[k] - pow(dados_observados[j + i * n_colunas] /
backscatters[k], 2.) / 2.);
}

novo = (unsigned char) 0;
probab = novos_valores[0];
for (k = 1; k < n_class; k++) {
if (probab < novos_valores[k]) {
probab = novos_valores[k];
novo = (unsigned char) k;
}
}

free((char *) valores_vizinhos);
free((char *) novos_valores);
return (novo);

}

/*
* -----
*/

```

B.3 Programa que implementa os Filtros Robustos

```

#include <values.h>
#include <limits.h>
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>

int          compara_uchar(void *x, void *y);

/*
* Programa: filrobesp.c
*
* Proposito: cria novas imagens resultantes da aplicacao de tres filtros
* robustos para a estimacao do parametro de uma distribuicao de Raighley: os
* filtros da mediana, o IQR (distancia inter-quartil) e o MAD (Median
* Absolute Deviation). As constantes K_IQR, K_MAD e K_MED foram calculadas
* para esta distribuicao. O algoritmo e de aplicabilidade geral, devendo
* apenas serem mudados estes valores.
*
* A constante sqrt(M_PI / 2.), aplicada as tres constantes, serve para
* relacionar o valor do estimador do basckscatter (que e o que retornam os
* filtros acima mencionados) com o valor da esperanza da distribuicao de
* Raighley (que e o valor desejado para as imagens filtradas; desta forma o
* valor medio em regioes homogneas e mantido).
*
* A janela sobre a qual a filtragem e feita e quadrada e de tamanho variavel.
*
* Na janela de comando escreve-se a linha sobre a qual os filtros estao sendo
* calculados. Caso haja igualdade entre todas as observacoes a serem

```

```

* filtradas, os tres estimadores sao definidos como esse valor, e na linha
* de comando, escreve-se um asterisco por cada ocorrencia deste evento.
*
* INPE: 29/9/92, 30/9/92, 1/10/92, 14/10/92, 15/10/92.
*
* Alejandro C. Frery e Sidnei Sant'Anna.
*
* Modificacao: agora teremos uma opcao para indicar se a detecao e linear ou
* quadratica. Caso seja linear, serao usadas as constantes para a
* distribuicao de Rayleigh; caso contrario serao usadas as constantes
* adequadas para a distribuicao exponencial.
*
* INPE: 27/10/92.
*
* 4/11/92: sao acrescentados os estimadores pelos metodos de maxima
* verossimilhanca e momentos baseado no primeiro momento amostral
*
* 2/12/92: sao acrescentados os estimadores pelos metodos de maxima
* verossimilhanca e momentos baseado no primeiro momento amostral aparados
*
*
*/

```

```

main()
{
time_t      tempo_i, tempo_f;
FILE        *original, *iqr, *mad, *med, *maxver, *mom, *mvt, *mot;
char        arq_original[20], arq_iqr[20], arq_mad[20], arq_med[20],
arq_maxver[20],
arq_mvt[20], arq_mot[20], arq_mom[20];
unsigned char uc, uciqr, ucmad, ucmed, ucmaxver, ucmom, uct, *x,
*obs, uctemp;
double      K_IQR, K_MAD, K_MED, temp, t_maxver, t_mom, tmv_t, tmo_t,
alpha0, rpi2 = sqrt(M_PI / 2.);
int         linhas, colunas, i, j, janela, j1, j2, jq, ele, oi,
oj, nj, si, ss, nap;
unsigned    opcao;
loop1:
printf("\nIngresse o nome do arquivo da imagem original: ");
scanf("%s", arq_original);
while ((original = fopen(arq_original, "rb")) == NULL)
goto loop1;

printf("\nIngresse, na ordem, os numeros de linhas e de colunas da imagem
original: ");
scanf("%d %d", &linhas, &colunas);
x = (unsigned char *) malloc(sizeof(unsigned char) * linhas * colunas);
if (x == NULL) {
printf("\n>>ERRO ALOCANDO MEMORIA: PROGRAMA ABORTADO<<\n");
exit(0);
}
loop_opcao:
printf("\nIndique \t[0] se a detecao for linear (Rayleigh) ou \n\t\t[1] se
for quadratica (Exponencial): ");
scanf("%u", &opcao);
if (opcao != 0 && opcao != 1) {
printf("\nPor favor ingresse uma opcao valida...");
goto loop_opcao;
} else if (opcao == 0) { /* constantes para a distribuicao Rayleigh */
K_IQR = (sqrt(4. * log(2.)) - sqrt(2. * log(4. / 3.))) /
sqrt(M_PI / 2.);
K_MAD = .448453 / sqrt(M_PI / 2.);
K_MED = sqrt(2. * log(2.)) / sqrt(M_PI / 2.);
} else { /* constantes para a distribuicao exponencial */
K_IQR = log(3.);
K_MAD = .481212;
}
}

```

```

K_MED = log(2.);
}

loop_janela:
printf("\nIngresse o tamanho >>IMPAR<< da janela que sera empregada nas
  filtragens: ");
scanf("%d", &janela);
if (janela % 2 == 0) {
goto loop_janela;
}
j1 = janela - 1;
j2 = janela / 2;
jq = janela * janela;
ele = (jq - 1) / 2;
obs = (unsigned char *) malloc(sizeof(unsigned char) * jq);
if (obs == NULL) {
printf("\n>>ERRO ALOCANDO MEMORIA: PROGRAMA ABORTADO<<\n");
exit(0);
}
loop2:
printf("\nIngresse o nome do arquivo da imagem IQR: ");
scanf("%s", arq_iqr);
while ((iqr = fopen(arq_iqr, "wb")) == NULL)
goto loop2;

loop3:
printf("\nIngresse o nome do arquivo da imagem MAD: ");
scanf("%s", arq_mad);
while ((mad = fopen(arq_mad, "wb")) == NULL)
goto loop3;

loop4:
printf("\nIngresse o nome do arquivo da imagem mediana: ");
scanf("%s", arq_med);
while ((med = fopen(arq_med, "wb")) == NULL)
goto loop4;

loop5:
printf("\nIngresse o nome do arquivo da imagem maxver: ");
scanf("%s", arq_maxver);
while ((maxver = fopen(arq_maxver, "wb")) == NULL)
goto loop5;

loop6:
printf("\nIngresse o nome do arquivo da imagem momentos: ");
scanf("%s", arq_mom);
while ((mom = fopen(arq_mom, "wb")) == NULL)
goto loop6;
loop7:
printf("\nIngresse o nome do arquivo da imagem max-ver aparado: ");
scanf("%s", arq_mvt);
while ((mvt = fopen(arq_mvt, "wb")) == NULL)
goto loop7;
loop8:
printf("\nIngresse o nome do arquivo da imagem momentos aparado: ");
scanf("%s", arq_mot);
while ((mot = fopen(arq_mot, "wb")) == NULL)
goto loop8;

printf("\nIngresse a proporcao de delecao: ");
scanf("%lf", &alpha0);
alpha0 /= 2.;

si = (unsigned) floor(alpha0 * (double) jq) + 1;
ss = jq - (unsigned) floor((double) jq * alpha0);
nap = ss - si + 1;
printf("\nSerao aproveitados somente %d dos %d pixels de cada janela para

```

```

as imagens baseadas em estimadores aparados\n", nap, jq);
printf("\nProcessando fila... ");

/* leitura da imagem no vetor x */
for (i = 0; i < linhas; i++)
for (j = 0; j < colunas; j++) {
fread((char *) &uc, sizeof(unsigned char), 1, original);
x[j + i * colunas] = uc;
}

time(&tempo_i);
for (i = 0; i < linhas; i++) {
printf("%d ", i);
for (j = 0; j < colunas; j++) {
if (i >= j2 && j >= j2 && i < linhas - j2 &&
j < colunas - j2) {

t_maxver = t_mom = tmv_t = tmo_t = 0.;
for (oi = -j2; oi <= j2; oi++)
for (oj = -j2; oj <= j2; oj++) {
uctemp = x[j + oj + (i + oi) *
colunas];
obs[(oj + j2) + (oi + j2) *
janela] = uctemp;
t_maxver += (double) (uctemp *
uctemp);
t_mom += uctemp;
}
}
/*
* calculo e escrita dos estimadores maxver e
* momentos
*/

t_maxver = rpi2 * sqrt(t_maxver / (double) (jq *
2));
t_maxver = (t_maxver < 255. ? t_maxver : 255.);
ucmaxver = (unsigned char) floor(t_maxver);
ucmom = (unsigned char) floor(t_mom /
(double) jq);
fwrite((char *) &ucmaxver, sizeof(unsigned char),
1, maxver);
fwrite((char *) &ucmom, sizeof(unsigned char), 1,
mom);

qsort(obs, jq, sizeof(unsigned char),
compara_uchar);

/*
* se o vetor de observacoes tiver todos os
* seus valores iguais, os estimadores sao
* definidos como esse valor
*/
if (obs[0] == obs[jq - 1]) {
printf("**");
uc = x[j + i * colunas];
fwrite((char *) &uc, sizeof(unsigned
char), 1, iqr);
fwrite((char *) &uc, sizeof(unsigned
char), 1, mad);
fwrite((char *) &uc, sizeof(unsigned
char), 1, med);
} else {
/*
* calculo e escrita do estimador IQR
*/

```

```

if (ele % 2 == 0) {
uciqr = (obs[jq - ele / 2] +
obs[jq + 1 - ele / 2] - obs[ele / 2] - obs[ele / 2 + 1]) / 2;
} else {
uciqr = obs[jq + 1 - (ele + 1) /
2] - obs[(ele + 1) / 2];
}
temp = (double) uciqr / K_IQR;
temp = temp < 255. ? temp : 255.;
uciqr = (unsigned char) floor(temp + .5);
fwrite((char *) &uciqr, sizeof(unsigned
char), 1, iqr);

/*
* calculo e escrita do estimador da
* mediana
*/
ucmed = ucmed = obs[jq / 2 + 1];
temp = (double) ucmed / K_MED;
temp = temp < 255. ? temp : 255.;
ucmed = (unsigned char) floor(temp + .5);
fwrite((char *) &ucmed, sizeof(unsigned
char), 1, med);

for (nj = si; nj <= ss; nj++) {

uct = obs[nj];

/*
* calculo do estimador pelo
* max-ver aparado
*/
tmv_t += (double) (uct * uct);

/*
* calculo do estimador pelos
* momentos aparados
*/
tmo_t += (double) uct;
}

/* escrita dos estimadores aparados */

tmv_t = rpi2 * sqrt(tmv_t / (2. * nap));
tmv_t = tmv_t < 255. ? tmv_t : 255;
uct = (unsigned char) floor(tmv_t + .5);
fwrite((char *) &uct, sizeof(unsigned
char), 1, mvt);

tmo_t /= nap;
tmo_t = tmo_t < 255. ? tmo_t : 255;
uct = (unsigned char) floor(tmo_t + .5);
fwrite((char *) &uct, sizeof(unsigned
char), 1, mot);

/*
* calculo e escrita do estimador MAD
*/
for (int k = 0; k < jq; k++)
obs[k] = (unsigned char)
abs(obs[k] - ucmed);
qsort(obs, jq, sizeof(unsigned char),
compara_uchar);
ucmad = obs[jq / 2 + 1];
temp = (double) ucmed / K_MAD;
temp = temp < 255. ? temp : 255.;

```

```

ucmad = (unsigned char) floor(temp + .5);
fwrite((char *) &ucmad, sizeof(unsigned
char), 1, mad);
}

} else {
/* valores nao filtrados permanecem */
uc = x[j + i * columnas];
fwrite((char *) &uc, sizeof(unsigned char), 1,
iqr);
fwrite((char *) &uc, sizeof(unsigned char), 1,
mad);
fwrite((char *) &uc, sizeof(unsigned char), 1,
med);
fwrite((char *) &uc, sizeof(unsigned char), 1,
maxver);
fwrite((char *) &uc, sizeof(unsigned char), 1,
mom);
fwrite((char *) &uc, sizeof(unsigned char), 1,
mvt);
fwrite((char *) &uc, sizeof(unsigned char), 1,
mot);
}
}
}

time(&tempo_f);

free((char *) x);
free((char *) obs);
fclose(original);
fclose(iqr);
fclose(mad);
fclose(med);
fclose(maxver);
fclose(mom);
fclose(mvt);
fclose(mot);

printf("\n0 tempo total de processamento foi de %d segundos.\n",
tempo_f - tempo_i);
printf("\n\n>>FIM<<\n\n");
return (0);
}

/* ////////////////////////////////////// */

int
compara_uchar(void *x, void *y)
{
if (*(unsigned char *) x < *(unsigned char *) y)
return (-1);
if (*(unsigned char *) x > *(unsigned char *) y)
return (1);
if (*(unsigned char *) x == *(unsigned char *) y)
return (0);
}

/* ////////////////////////////////////// */

```

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS
EDITADAS PELO INPE

RELATÓRIOS DE PESQUISA

- Reportam resultados de pesquisa tanto de natureza técnica quanto científica.

PUBLICAÇÕES DIDÁTICAS

- Apostilas, notas de aula e manuais didáticos

NOTAS TÉCNICO-CIENTÍFICAS

- Incluem resultados preliminares de pesquisa, descrição de equipamentos, software, sistemas e experimentos, apresentação de testes, dados e atlas, e documentação de projetos de engenharia.

TESES E DISSERTAÇÕES

- Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

MANUAIS TÉCNICOS

- Descrevem normas, procedimentos, instruções e orientações.

PUBLICAÇÕES SERIADAS

- Periódicos Técnico-Científicos:
Boletim de Sensoriamento Remoto,
Climanálise: Boletim de
Monitoramento e Análise
Climática.
- Anais de Eventos