



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

INPE-14448-TDI/1139

**ÁREAS AGRÍCOLAS EM SENSORES COM RESOLUÇÃO  
ESPACIAL DE 30 M ESTIMADOS A PARTIR DE DADOS  
ORIGINAIS E SIMULADOS MODIS E MÉTRICAS DE PAISAGEM**

Denis de Moura Soares

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto,  
orientada pelos Drs. Antônio Roberto Formaggio e Lênio Soares Galvão, aprovada em  
30 de maio de 2006.

INPE  
São José dos Campos  
2006

528.831.1 (815)

Soares, D. M.

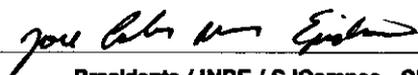
Áreas agrícolas em sensores com resolução espacial de 30 m estimadas a partir de dados originais e simulados MODIS e métricas de paisagem / Denis de Moura Soares. – São José dos Campos: INPE, 2006.

153p. ; (INPE-1448-TDI/1139)

1.Resolução moderada. 2.Padrão espacial. 3.Métrica da paisagem. 4.Análise de regressão. 5.Filtragem de textura. 6.Filtragem de maioria. 7.Escala. 8.MODIS/Terra. I.Título.

Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de **Mestre** em  
**Sensoriamento Remoto**

Dr. José Carlos Neves Epiphânio



Presidente / INPE / SJC Campos - SP

Dr. Antonio Roberto Formaggio



Orientador(a) / INPE / SJC Campos - SP

Dr. Lênio Soares Galvão



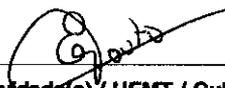
Orientador(a) / INPE / SJC Campos - SP

Dr. Yosio Edemir Shimabukuro



Membro da Banca / INPE / SJC Campos - SP

Dr. Eduardo Guimarães Couto



Convidado(a) / UFMT / Cuiabá - MT

Aluno (a): **Dênis de Moura Soares**

São José dos Campos, 30 de maio de 2006



*“Por enfrentarem o que é penoso que os homens são chamados corajosos.”*

*“A sabedoria prática é a disposição da mente que se ocupa com coisas justas, nobres e boas para o homem.”*

ARISTÓTELES



*À doce Samira Soares*



## AGRADECIMENTOS

Agradeço a todas as pessoas que estiveram presentes e contribuíram de qualquer forma para a conclusão de mais esta etapa da vida.

À Samira Soares, esposa presente e dedicada que também sofreu, venceu e concluiu este trabalho.

Ao Exército Brasileiro, na figura da Diretoria de Serviço Geográfico, por permitir a inteira dedicação aos estudos realizados.

Ao Instituto Nacional de Pesquisas Espaciais (INPE), pela oportunidade de aperfeiçoamento e utilização de todos os seus vastos recursos.

Aos pesquisadores do INPE pela prontidão e dedicação ao compartilhar quaisquer conhecimentos.

Aos meus orientadores Dr. Antônio Roberto Formaggio e Dr. Lênio Soares Galvão, pela experiência trocada, pela orientação, apoio e presença constante na consecução deste trabalho. Sobretudo, amigos e incentivadores, sem os quais a árdua tarefa de dissertar se tornaria intolerável.

À Dr<sup>a</sup>. Corina da Costa Freitas e ao Dr. Camilo Daleles Rennó, pelo apoio a tempo e à hora para desvendar os mistérios da estatística multivariada. Sempre disponíveis e acolhedores.

Aos amigos do Laboratório de Tratamento de Imagens Digitais (LTID): Delano, Marcos Gaúcho, Daniela França, Camila Anjos, Michelle, Pablo, Marcus Catarina, Taise, Tiago. Garantia de incentivo e boas risadas nas horas difíceis.

Aos integrantes da turma de pós-graduação ingressantes em 2004, cujos nomes não serão citados sob pena de muito me alongar, mas que terão sempre um lugar em meus pensamentos.

À sempre pronta Etel, secretária exemplar da pós-graduação em Sensoriamento Remoto e à Cris cuja alegria e disponibilidade irradiam-se pela Divisão de Sensoriamento Remoto (DSR).

A meus pais que jogaram a semente e muito cultivaram para permitir o sucesso que ora se apresenta.

A Deus, o Criador, que fez parecer impossível este desafio e tornou-o um honroso triunfo.



## RESUMO

O agronegócio tem papel de destaque na economia brasileira. Dessa forma, a criação de metodologias para o monitoramento agrícola é fundamental. Nesta linha de raciocínio, a estimativa de área agrícola é uma atividade importante para a previsão de safras e avaliação da disponibilidade de produtos para abastecimento interno e exportações. A utilização de sensores remotos tem se mostrado eficiente na medição de áreas. No entanto, a abundância de nuvens representa um fator crítico para o sucesso de sua aplicação. Sensores com repetitividade quase diária podem ser uma solução para essa limitação, apesar de sua resolução espacial moderada. Assim, este trabalho teve por objetivo avaliar as diferenças obtidas nas estimativas de áreas de culturas agrícolas quando são utilizados sensores de resolução espacial moderada (p.ex, MODIS/Terra, com 250m), ao invés de resolução espacial fina (ETM+/Landsat-7, com 30m), considerando dados originais e simulados, diferentes culturas agrícolas e seu padrão de distribuição espacial (métricas da paisagem). As culturas agrícolas avaliadas foram o milho, a cana-de-açúcar e a soja, na região de Ipuã, Guará e São Joaquim da Barra, no norte paulista. Também foram incluídas na análise as classes temáticas mata, pastagem e solo exposto. Para atingir tal objetivo foi estudada a evolução dos valores das métricas de paisagem em função de degradações sucessivas da imagem ETM+ para obter imagens de 90m, 150m, 210m e 270m, utilizando filtragem espacial de textura e de maioria. Modelos de regressão simples (área) e múltipla (área e métricas) foram elaborados com base em dados originais dos sensores ETM+ e MODIS, considerando todas as classes em conjunto (abordagem geral) e cada classe individualmente (abordagem específica). Os resultados obtidos mostraram que: a) as duas técnicas de simulação afetaram de maneira semelhante o padrão espacial das classes temáticas, sendo a filtragem de textura mais realística na tarefa de representar o sensor MODIS/Terra; b) na abordagem geral para estimativa de área, a regressão simples entre as áreas de classes temáticas obtidas das imagens originais apresentou coeficiente de determinação ( $R^2$ ) de 0,46 e a inclusão dos índices de padrão espacial no modelo de regressão múltipla elevou tal grandeza para 0,49, sendo as métricas *Área*, *LPI*, *LSI*, *TCA*, *PLADJ* e *IJI* as constituintes do modelo; c) na abordagem específica, a criação de modelos estatísticos para cada cultura agrícola elevou bastante o  $R^2$ , atingindo valores de 0,52, 0,67 e 0,87, para o milho, a cana e a soja, respectivamente. O modelo para o milho foi composto pelos índices *LSI*, *CLUMPY*, *IJI*, *MESH* e *NP*, para a cana pelos *Área*, *LSI*, *CLUMPY*, *IJI* e *NLSI*. Por fim, para a soja as métricas relevantes foram: *Área*, *NP*, *NDCA*, *DCAD*, *COHE*. Portanto, os resultados demonstram que sensores de resolução espacial moderada podem ser utilizados para predizer áreas vistas por sensores de resolução mais fina, especialmente para as culturas agrícolas menos fragmentadas (soja e cana) e com a adoção de modelos estatísticos específicos que incorporem métricas de paisagem.



## CROP AREA IN SPATIAL RESOLUTION OF 30m ESTIMATED WITH ORIGINAL AND SIMULATED MODIS DATA AND LANDSCAPE METRICS

### ABSTRACT

The national agribusiness is very important for the Brazilian economy. Thus, the creation of methodologies to monitor the activity is fundamental. In this way, the estimation of crop area is an important tool for prediction of the availability of products for national consumption and exportation. Remote sensors have frequently been used to measure areas, but clouds abundance, mainly in agricultural seasons, represents a huge difficulty. Sensors with almost daily revisit time can be a solution for that limitation, but their spatial resolution is usually poor. Therefore, the aim of this work was to evaluate the differences between crop area estimation from coarse resolution data (e.g. MODIS/Terra, with 250m) and fine resolution data (ETM+/Landsat-7, with 30m), using real and simulated data, different crop types and their spatial pattern (landscape metrics). The analysis was applied for three different crops: corn, sugarcane and soybean. The study area was located close to the Ipuã, Guar and So Joaquim da Barra, cities in the north of the So Paulo state, Brazil. The thematic classes woodland, pasture and exposed soil were also included in the analysis. To reach the goal, the behavior of the landscape metrics was studied as a function of the simulation of different levels of spatial resolution (90m, 150m, 210m and 270m) from ETM+ data using texture and majority filtering. Simple (area) and multiple (area plus landscape metrics) regression models were constructed using original data of the sensors ETM+ and MODIS, considering all classes together (general approach) and each class individually (specific approach). The results showed that: a) both simulation techniques affected similarly the spatial pattern of the thematic classes, but the texture filtering was more realistic to represent the MODIS/Terra sensor; b) in the general approach to estimate crop area, the simple regression between class areas from real data presented low coefficient of determination ( $R^2$  of 0.46). By adding the landscape metrics, this coefficient increased to 0.49. The selected indices for this procedure were *rea*, *LPI*, *LSI*, *TCA PLADJ* and *IJI*; c) in the specific approach, the creation of statistic models for each agricultural class increased the  $R^2$  to 0.52, 0.67 and 0.87 for corn, sugarcane and soybean, respectively. The model for corn was composed of the metrics *LSI*, *CLUMPY*, *IJI*, *MESH* and *NP*. For sugarcane, *rea*, *LSI*, *CLUMPY*, *IJI* and *NLSI* were selected as the most important metrics. Finally, the relevant landscape metrics for soybean were *rea*, *NP*, *NDCA*, *DCAD* and *COHE*. Thus, the results demonstrated that coarse spatial resolution data can be utilized to predict crop area measured from fine spatial resolution data, especially for less fragmented crops (soybean and sugarcane) and with the use of specific crop regression models incorporating landscape metrics.



## SUMÁRIO

Pág.

### LISTA DE FIGURAS

### LISTA DE TABELAS

### LISTA DE SÍMBOLOS

### LISTA DE SIGLAS E ABREVIATURAS

<b>CAPÍTULO 1 - INTRODUÇÃO .....</b>	<b>29</b>
<b>CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>33</b>
2.1 A cultura do Milho .....	33
2.2 A cultura da Soja .....	34
2.3 A Cultura da Cana-de-Açúcar .....	34
2.4 A Estimativa de Área Agrícola no Brasil .....	35
2.5 O Sensoriamento Remoto na Agricultura .....	36
2.6 Os Sistemas Sensores .....	37
2.7 Padrão Espacial da Paisagem e Efeitos de Agregação Espacial.....	40
2.8 Métricas da Paisagem .....	43
2.8.1 Índices de Área/Densidade/Borda (ADB) .....	43
2.8.2 Índices de Área Nuclear ( <i>Core</i> ) .....	46
2.8.3 Índices de Contágio/Espalhamento (CE) .....	47
2.8.4 Índices de Forma e Conectividade .....	51
2.9 Simulação da Resolução Espacial Moderada .....	52
2.9.1 Filtragem Espacial de Maioria .....	53
2.9.2 Filtragem Espacial de Textura .....	54
2.10 Classificação Digital de Imagens .....	56
2.10.1 Classificação Supervisionada .....	58
2.10.2 Classificação Não Supervisionada .....	59
2.11 Análise de Regressão .....	59
2.11.1 Coeficiente de Determinação .....	61
2.11.2 Premissas Fundamentais da Análise de Regressão .....	62
2.11.2.1 Análise de Normalidade .....	63
2.11.2.2 Análise da Variância dos Resíduos .....	65
2.11.3 Identificação de <i>Outliers</i> Influentes .....	66
2.11.4 Seleção das Variáveis do Modelo Regressão Múltipla .....	66
2.11.5 Diagnóstico de Multicolinearidade .....	68
2.11.6 Validação do Modelo de Regressão .....	69
<b>CAPÍTULO 3 - ÁREA DE ESTUDO.....</b>	<b>71</b>
<b>CAPÍTULO 4 - MATERIAIS E MÉTODO .....</b>	<b>75</b>
4.1 Materiais .....	75
4.1.1 Dados de Trabalho .....	75

4.1.2 Recursos de <i>Software</i> .....	76
4.2 Método .....	79
4.2.1 Simulação de Resolução Espacial por Filtragem.....	79
4.2.1.1 Filtragem de Maioria .....	80
4.2.1.2 Filtragem de Textura.....	80
4.2.2 Cálculo e Evolução das Métricas da Paisagem em Função da Resolução Espacial... 83	
4.2.3 Determinação da Técnica de Simulação (Filtragem) mais Realística e Pré-Seleção de Métricas para o Modelo de Regressão Múltipla .....	85
4.2.4 Modelagem Estatística para Predição de Área: Obtenção de Imagem de 30m através de Sensor de 250m de Resolução Espacial .....	85
4.2.4.1 Modelo de Regressão Simples com Dados Originais ( $ÁREA_{ETM+}$ versus $ÁREA_{MODIS}$ ).....	86
4.2.4.2 Modelo de Regressão Múltipla com Dados Originais e Métricas de Paisagem ( $ÁREA_{ETM+}$ versus $ÁREA_{MODIS}$ + Métricas) .....	87
4.2.5 Comparação dos Modelos de Regressão Simples e Múltipla: Abordagens Geral (Todas as Culturas) e Específica (por Cultura).....	88
<b>CAPÍTULO 5 - RESULTADOS E DISCUSSÃO .....</b>	<b>91</b>
5.1 Análise das Classificações dos Produtos de Resolução Espacial Simulada por Filtragem de Textura e Maioria .....	91
5.2 Evolução das Métricas de Paisagem com a Simulação da Resolução Espacial .....	95
5.2.1 Filtragem de Textura.....	95
5.2.1.1 Índices de Área/Densidade/Borda .....	95
5.2.1.2 Índices de Área Nuclear.....	97
5.2.1.3 Índices de Contágio/Espalhamento.....	99
5.2.1.4 Índices de Forma e de Conectividade .....	101
5.2.2 Filtragem de Maioria .....	102
5.2.2.1 Índices de Área/Densidade/Borda .....	102
5.2.2.2 Índices de Área Nuclear.....	104
5.2.2.3 Índices de Contágio/Espalhamento.....	105
5.2.2.4 Índices de Forma e de Conectividade .....	107
5.3 Aspectos Comparativos entre as Filtragens de Maioria e de Textura .....	107
5.4 Determinação da Técnica de Filtragem mais Realística para a Simulação da Resolução Espacial Moderada (MODIS/Terra).....	116
5.5 Confeção do Modelo Estatístico para Predição de Áreas Agrícolas do Sensor de Resolução Espacial Fina através da Moderada e de Métricas da Paisagem.....	118
5.5.1 Pré-Seleção das Métricas para a Modelagem Estatística com Base na Filtragem Espacial.....	118
5.5.2 Análise Exploratória dos Dados .....	119
5.5.2.1 Análise de normalidade da variável dependente .....	119
5.5.2.2 Transformações das Variáveis Preditivas .....	122
5.5.3 Constituição do Modelo de Regressão Simples.....	122
5.5.3.1 Sumário da Regressão Simples.....	124
5.5.3.2 Análise de Variância .....	125
5.5.3.3 Análise de Normalidade .....	126
5.5.3.4 Identificação de <i>Outliers</i> Influentes.....	128
5.5.3.5 Validação da Regressão Simples .....	128

5.5.4 Constituição do Modelo de Regressão Múltipla.....	128
5.5.4.1 Análise de Variância .....	133
5.5.4.2 Análise de Normalidade .....	134
5.5.4.3 Identificação de <i>Outliers</i> Influentes.....	135
5.5.4.4 Efeito de Multicolinearidade .....	135
5.5.4.5 Validação do Modelo de Regressão .....	136
5.6 Análise Comparativa entre a Abordagem Geral versus a Específica Utilizando Modelos de Regressão Simples e Múltipla.....	137
<b>CAPÍTULO 6 - CONCLUSÕES E RECOMENDAÇÕES .....</b>	<b>141</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>145</b>



## LISTA DE FIGURAS

	Pág.
2.1 – Erro devido à agregação espacial na degradação da resolução espacial. ....	41
2.2 – Filtragem de Maioria com dimensão 5 ´ 5.....	54
2.3 – Definição da vizinhança do <i>pixel</i> * .....	55
2.4 – (a) Imagem 4x4, 2 bits; (b) Forma geral de qualquer MDETC; (c)–(f) Todas MDETC para distância unitária. ....	57
2.5 – Espaço multiespectral bidimensional, representando a probabilidade de <i>pixels</i> pertencerem a cada classe – P(x).....	58
2.6 – Simulação (a) da condição inicial; (b), (c) e (d) de três iterações efetuadas pelo algoritmo K-means; e (e) da migração resultante dos vetores das médias.....	60
2.7 – Variável independente (X) predizendo dependente (Y): (a) Perfeita adequação; (b) nuvem de pontos no entorno de uma reta. ....	62
2.8 – Análise da normalidade dos dados: (a) Histograma; (b) Box-plot; (c) Probabilidade normal. ....	64
3.1 – Localização da área de estudo. ....	71
3.2 – Localização dos municípios na área de estudo. ....	72
3.3 – Mapa de uso e ocupação do solo da área de estudo. ....	73
4.1 – Mapa de referência para classificação temática. ....	77
4.2 – Análise de simulações através do uso de filtragens de maioria. ....	81
4.3 – Análise de simulações através do uso de filtragens de textura. ....	82
4.4 – Teste da relação $A_F=f(A_M)$ . ....	87
4.5 – Seleção dos índices espaciais .....	89
5.1 – Classificação não-supervisionada da imagem ETM+/Landsat-7 (1x1 pixel) e das imagens resultantes da filtragem de textura com janelas de 3x3 (90m), 5x5 (150m), 7x7 (210m) e 9x9 (270m) <i>pixels</i> , com os respectivos índices Kappa.....	92
5.2 – Variação do índice Kappa em função do tamanho da janela (resolução espacial) utilizada na filtragem de textura.....	93
5.3 – Classificação não-supervisionada da imagem ETM+/Landsat-7 (1x1 pixel) e das imagens resultantes da filtragem de maioria com janelas de 3x3 (90m), 5x5 (150m), 7x7 (210m) e 9x9 (270m) <i>pixels</i> . com os respectivos índices Kappa.....	94
5.4 – Variação do índice Kappa em função do tamanho da janela (resolução espacial) utiliza da na filtragem de maioria.....	95
5.5 – Comportamento dos índices de área/densidade/borda à variação da dimensão do filtro de textura (resolução espacial). ....	96
5.6 – Comportamento dos índices de área nuclear à variação da dimensão do filtro de textura (resolução espacial). ....	98
5.7 – Comportamento dos índices de contágio/espalhamento à variação da dimensão do filtro de textura (resolução espacial). ....	100
5.8 – Comportamento dos índices de forma (PAFRAC) e de conectividade (COHE) à variação da dimensão do filtro de textura (resolução espacial). .	101
5.9 – Comportamento dos índices de área/densidade/borda à variação da dimensão do filtro de maioria (resolução espacial). ....	103

5.10 – Comportamento dos índices de área nuclear à variação da dimensão do filtro de maioria (resolução espacial). .....	105
5.11 – Comportamento dos índices de contágio/espalhamento à variação da dimensão do filtro de maioria (resolução espacial).....	106
5.12 – Comportamento dos índices de forma (PAFRAC - Dimensão fractal Perímetro-Área) e de conectividade (COHE - Coesão de Área) à variação da dimensão do filtro de maioria (resolução espacial). .....	108
5.13 – Evolução da área das diversas culturas, em função da dimensão do filtro de textura e Maioria aplicado (resolução espacial).....	109
5.14 – Regressões entre as áreas das culturas cana e soja, em hectares, na imagem ETM+ e nas imagens resultantes dos processos de filtragem.....	111
5.15 – Regressões entre as áreas das classes solo e pastagem, em hectares, na imagem ETM+ e nas imagens resultantes dos processos de filtragem. ....	112
5.16 – Regressões entre as áreas das classes milho e mata, em hectares, na imagem ETM+ e nas imagens resultantes dos processos de filtragem. ....	113
5.17 – Evolução do $R^2$ das regressões entre as áreas obtidas dos dados simulados e aquelas extraídas da imagem do sensor ETM+/Landsat-7, para as diversas culturas.....	115
5.18 – Comparação entre a área (em hectares), obtida através da imagem MODIS/Terra (250m) e aquela extraída das simulações da resolução espacial efetuadas por filtragem de textura e maioria (270m), por classe temática. ....	117
5.19 – Comparação entre a área (em hectares) obtida da imagem MODIS/Terra (250m) e os valores de áreas resultantes das simulações de resolução espacial efetuadas por filtragem de textura e maioria (270m), para todas as classes de cobertura analisadas.....	118
5.20 – Estudo da normalidade da variável dependente (Área ETM+): (a) Histograma; (b) Probabilidade normal; (c) Box-plot. ....	120
5.21 – Estudo da normalidade da variável dependente transformada (raiz de Área ETM+): (a) Histograma; (b) Probabilidade normal; (c) Box-plot. ....	121
5.22 – Normalização dos resíduos brutos após transformação da variável dependente. ....	124
5.23 – Relação dos resíduos brutos com variável dependente.....	126
5.24 – Gráfico de probabilidade normal dos resíduos brutos.....	127
5.25 – Histograma de frequência dos resíduos brutos.....	127
5.26 – Seleção dos índices espaciais: Critério $C_p$ .....	130
5.27 – Seleção dos índices espaciais: Critério $R_p^2$ .....	131
5.28 – Seleção dos índices espaciais: Critério $R_a^2$ .....	131
5.29 – Relação dos resíduos brutos com as variáveis independentes <i>Área, raiz LPI, PLADJ</i> <sup>2</sup> . ....	133
5.30 – Análise da normalidade dos resíduos: a) Gráfico de probabilidade normal; b) Histograma; c) <i>Box-plot</i> . ....	135

## LISTA DE TABELAS

	Pág.
2.1 – Área, produção e produtores de milho no Brasil.....	33
2.2 – Características do sensor ETM+/Landsat-7.....	39
2.3 – Características do sensor MODIS.....	39
3.1 – Área cultivada, em hectares, por cultura, para o ano de 2002. ....	74
3.2 – Tamanho mínimo, médio e máximo das UPA, em hectares. ....	74
4.1 – Modelagem do banco de dados geográfico. ....	75
4.2 – Recursos de <i>software</i> empregados.....	79
4.3 – Métricas de paisagem utilizadas no trabalho. ....	84
5.1 – Comparação do Número de Polígonos ( <i>NP</i> ) para as classes milho e mata nas diferentes fitragens.....	104
5.2 – Comparação do Número de Polígonos ( <i>NP</i> ) para as classes soja e cana nas diferentes fitragens.....	104
5.3 – Matriz de correlação dos índices de padrão espacial. ....	123
5.4 – Parâmetros da regressão simples.....	125
5.5 – Análise da variância (ANOVA). ....	125
5.6 – Parâmetros da regressão simples ponderada. ....	126
5.7 – Intervalos de confiança da estimativa de parâmetros.....	126
5.8 – Parâmetros da regressão da amostra de validação. ....	128
5.9 – Parâmetros da regressão múltipla. ....	129
5.10 – Soma do quadrado dos resíduos ( <i>SSE</i> ). ....	129
5.11 – Parâmetros do modelo elaborado pelo critério $C_p$ .....	130
5.12 – ANOVA do modelo elaborado pelo critério $C_p$ .....	131
5.13 – Parâmetros do modelo elaborado pelos critérios $R_p^2$ e $R_a^2$ , com dados estatisticamente significantes indicados em vermelho. ....	132
5.14 – Soma do quadrado dos resíduos ( <i>SSE</i> ). ....	132
5.15 – Parâmetros da regressão múltipla ponderada. ....	134
5.16 – Intervalos de confiança da estimativa de parâmetros.....	134
5.17 – VIF do modelo de regressão múltipla. ....	136
5.18 – Parâmetros da regressão da amostra de validação. ....	136
5.19 – Parâmetros da regressão simples específica para o milho. ....	137
5.20 – Parâmetros da regressão simples específica para a soja. ....	137
5.21 – Parâmetros da regressão simples específica para a cana-de-açúcar.....	137
5.22 – Parâmetros da regressão múltipla específica para o milho. ....	138
5.23 – Parâmetros da regressão múltipla específica para a soja. ....	138
5.24 – Parâmetros da regressão múltipla específica para a cana-de-açúcar. ....	138
5.25 – Coeficientes de determinação ( $R^2$ ) para as regressões simples e múltiplas com abordagem geral e específica. ....	138



## LISTA DE SÍMBOLOS

- a*** - Nível de Significância
- b*** - Parâmetro da Análise de Regressão
- e*** - Resíduo da Análise de Regressão
- B*** - Estimador do Parâmetro da Análise de Regressão
- C<sub>p</sub>*** - Critério de Mallow para Seleção de Variáveis
- F*** - Valor da Distribuição *F* de Fisher
- F\**** - Valor do Teste Estatístico *F* Calculado
- ha** - Hectares
- R*** - Coeficiente de Correlação
- R<sup>2</sup>*** - Coeficiente de Determinação
- R<sub>p</sub><sup>2</sup>*** - Critério *R<sup>2</sup>* para Seleção de Variáveis
- R<sub>a</sub><sup>2</sup>*** - Critério *R<sup>2</sup>* Ajustado para Seleção de Variáveis
- t*** - Valor da Distribuição *t-student*
- t\**** - Valor do Teste Estatístico *t-student* Calculado



## LISTA DE SIGLAS E ABREVIATURAS

1/raiz4 MESH	- Um dividido pela raiz quarta da Tamanho de Trama Efetiva
ADB	- Área/Densidade/Borda
AEB	- Agência Espacial Brasileira
AI	- Agregação
ANOVA	- Análise de Variância
ARD	- Ajuste à Reta Diagonal
AVHRR	- <i>Advanced Very High Resolution Radiometer</i>
CATI	- Coordenadoria de Assistência Técnica Integral
CBERS	- Satélite Sino-Brasileiro de Recursos Terrestres
CCD	- <i>Charge Coupled Device</i>
CE	- Contágio/Espalhamento
CENBIO	- Centro Nacional de Referência de Biomassa
CEPAGRO	- Comissão Especial de Planejamento, Controle e Avaliação das Estatísticas Agropecuárias
CLUMP	- Agrupamento
CNA	- Confederação da Agricultura e Pecuária do Brasil
COHE	- Coesão de Área
COHE <sup>4</sup>	- Quarta potência Coesão de área
CPLAND	- Porcentagem Nuclear de Cenário
DCAD	- Densidade de Área Nuclear Disjunta
Desv. Pad.	- Desvio Padrão
df	- Graus de Liberdade
DFFITS	- Diferença Padronizada na Estimativa da <i>iésima</i> Observação

DIV	- Divisão de Cenário
Embrapa	- Empresa Brasileira de Pesquisa Agropecuária
ENVI	- <i>Environment for Visualizing Images</i>
EOS	- <i>Earth Observing System</i>
ETM+	- <i>Enhanced Thematic Map Plus</i>
IBGE	- Instituto Brasileiro de Geografia e Estatística
IDL	- <i>Interactive Data Language</i>
IJI	- Interposição e Justaposição
INPE	- Instituto Nacional de Pesquisas Espaciais
LPI	- Maior Área
LSI	- Forma de Classe
LSPA	- Levantamento Sistemático de Produção Agrícola
MCT	- Ministério de Ciência e Tecnologia
MDETC	- Matriz de Dependência Espacial dos Tons de Cinza
MESH	- Tamanho de Trama Efetiva
MISR	- <i>Multi-angle Imaging Spectroradiometer</i>
MMQ	- Método dos Mínimos Quadrados
MODIS	- <i>Moderate Resolution Imaging Spectroradiometer</i>
MRT	- MODIS Reprojection Tool
MSPR	- Erro Quadrático Médio de Predição
NASA	- <i>National Aeronautics and Space Administration</i>
NASS	- <i>National Agricultural Statistics Service</i>
NDCA	- Número de Áreas Nucleares Disjuntas
NLSI	- Forma de Classe Normalizada
NOAA	- <i>National Oceanic and Atmospheric Administration</i>

NP	- Número de Polígonos
PAFRAC	- Dimensão Fractal Perímetro-Área
PARA	- Razão Perímetro-Área
PD	- Densidade de Área
PIB	- Produto Interno Bruto
PLADJ	- Porcentagem de Adjacência
PLADJ <sup>2</sup>	- Segunda potência da Porcentagem de Adjacência
PLAND	- Porcentagem de Cenário
PNAE	- Programa Nacional de Atividades Espaciais
PRODES	- Programa de Monitoramento de Desmatamento da Amazônia
REM	- Radiação Eletromagnética
S. J. Barra	- São Joaquim da Barra
SAD-69	- <i>South American Datum - 1969</i>
SPLIT	- Fragmentação
SPRING	- Sistema para Processamento de Informações Georreferenciadas
SSE	- Soma dos Quadrados dos Resíduos
TCA	- Área Nuclear Total
TM	- <i>Thematic Mapper</i>
UPA	- Unidades de Produção Agropecuária
USDA	- <i>United States Department of Agriculture</i>
USGS	- <i>United States Geological Survey</i>
UTM	- Universal Transversa de Mercator
VIF	- Fator de Inflação da Variância
WFI	- <i>Wide Field Imager</i>



## CAPÍTULO 1

### INTRODUÇÃO

A produção agrícola nacional tem posição de destaque no cenário econômico, sendo responsável por grande parte das exportações nacionais. Com um Produto Interno Bruto (PIB) de R\$ 524,5 bilhões, correspondendo a 30,4% do PIB total da economia, o setor agropecuário tem posição de destaque no panorama econômico brasileiro (CNA, 2005). O Brasil, um país de dimensões continentais e clima favorável, desponta como potência mundial no setor agrícola, sendo apontado como um dos poucos países mundiais que ainda têm disponibilidade de terras para aumentar a extensão de seu cultivo.

Fica patente, então, a importância do agronegócio para o Brasil, o que exige esforços governamentais no controle do que se planta, e de como, onde e quanto é plantado. Tais informações são fundamentais para o desenvolvimento de políticas públicas para exportações, bem como para abastecimento interno.

Dessa forma, o Instituto Brasileiro de Geografia e Estatística (IBGE) desenvolve mensalmente o Levantamento Sistemático da Produção Agrícola (LSPA), obtendo informações nos municípios sobre previsão e acompanhamento de safras agrícolas, com estimativas de produção, rendimento médio e áreas plantadas e colhidas (IBGE, 2002). Entretanto, as informações obtidas são de caráter subjetivo, uma vez que se baseiam em questionários aplicados a produtores, associações de classe e cooperativas, sem, contudo, haver medições físicas (IBGE, 2002; Pino, 2001).

O uso de técnicas de sensoriamento remoto, portanto, apresenta-se como alternativa para tornar mais objetiva a estimativa de áreas agrícolas. As imagens do sensor ETM+/Landsat-7 ou do CCD/CBERS-2 são de resolução espacial intermediária, permitindo, assim, uma escala de representação compatível com os alvos agrícolas de interesse. No entanto, esses sensores apresentam uma repetitividade temporal muito baixa (16 dias para o ETM+ e 26 dias para o CCD), o que dificulta sobremaneira a

obtenção de imagens livres de nuvens. Pax-Lenney e Woodcock (1997) afirmam, ainda, que sensores de resolução fina são úteis para representação e monitoramento do detalhe de pequenas áreas, no entanto, apresentam restrições para uso em escala regional ou global, uma vez que sua área de abrangência é pequena e a cobertura temporal limitada.

Entretanto, existem sensores como o *Moderate Resolution Imaging Spectroradiometer* (MODIS), *Multi-angle Imaging SpectroRadiometer* (MISR), ambos da plataforma Terra, e o *Advanced Very High Resolution Radiometer* (AVHRR), da plataforma *National Oceanic and Atmospheric Administration* (NOAA), que podem apresentar uma resolução temporal diária, sendo, assim, ideais para possibilitar a obtenção de imagens livres de cobertura de nuvens. Estas imagens, porém, têm uma resolução espacial que pode variar de 250m (MODIS/Terra) a 1100m (AVHRR/NOAA-17 e MISR/Terra), portanto, muito inferior àquela do CCD/CBERS-2 (20m) e do ETM+/Landsat-7 (30m). Tal disparidade sugere a necessidade de estudos para avaliação da habilidade de sensores dotados de resolução espacial moderada em detectar os elementos agrícolas de interesse.

Fazendo um paralelo ao Programa Nacional de Atividades Espaciais (PNAE), há um forte interesse na implantação de um sistema de monitoramento eficiente, que forneça um acesso rápido às informações que denotam a dinâmica de uso do solo (AEB, 2005). Dessa forma, faz-se necessária a determinação da configuração ótima de sensores remotos para utilização nas diversas aplicações de um programa de satélites de sensoriamento remoto.

Turner et al. (1989) estudaram o efeito da mudança de escala na extração de informações de alvos terrestres. Eles afirmam que o desenvolvimento de métodos que preservem informação através das escalas ou quantifiquem sua perda na mudança de escala tem se tornado muito importante, uma vez que diversos estudos requerem a extrapolação de medidas feitas sobre dados de resolução fina para análises globais ou regionais.

Diversos trabalhos analisaram, ainda, o desempenho de imagens de resolução espacial moderada na estimativa de área para o cálculo de desmatamento. Malingreau e Belward

(1992), Mayaux e Lambin (1995), Moody e Woodcock (1995), Mayaux e Lambin (1997), Moody (1998), Ponzoni et al. (2002), Millington et al. (2003), Frohn e Hao (2006) são exemplos de estudos desse tema. Entretanto, apenas um pequeno número de trabalhos abordou a questão da influência da resolução espacial na estimativa de áreas agrícolas. Por exemplo, Pax-Lenney e Woodcock (1997) discutiram o impacto da resolução espacial na estimativa de áreas agrícolas do Egito, usando um algoritmo de convolução para degradar a resolução espacial de imagens TM/Landsat-5 (30m) para diferentes tamanhos de pixels (120m, 240m, 480m e 960m). Segundo estes autores, a degradação da resolução espacial produziu uma sub-estimativa do cálculo de áreas agrícolas que variou de 4,3% (120m) a 9,4% (960m). Entretanto, os próprios autores questionaram a eficiência de procedimentos de degradação em reproduzir, com fidelidade, sensores de resolução espacial moderada como o MODIS/Terra. Eles também ressaltaram que a precisão das estimativas pode ter sido influenciada pelas características da paisagem (p.ex., tamanho das fazendas).

É nesse contexto que se insere o presente trabalho. A representação de alvos agrícolas em imagens produzidas por sensores com baixa resolução espacial pode estar condicionada ao padrão espacial das culturas, sendo o efeito de agregação espacial um aspecto relevante na estimativa das proporções dos tipos de cobertura do solo (Moody e Woodcock, 1994; Mayaux e Lambin, 1995).

O objetivo deste trabalho foi, portanto, avaliar as diferenças obtidas nas estimativas de áreas de culturas agrícolas quando são utilizados sensores de resolução espacial mais moderada (p.ex, MODIS/Terra, com 250m), ao invés de resolução mais fina (ETM+/Landsat-7, com 30m), considerando dados originais e simulados, diferentes culturas agrícolas e seu padrão de distribuição espacial (métricas da paisagem). As culturas abordadas foram soja, cana-de-açúcar e milho, que são produtos de interesse nacional e apresentam padrões espaciais distintos. Além dessas culturas, foram ainda consideradas na análise as classes temáticas mata, pastagem e solo exposto para compor um modelo estatístico mais abrangente e facilitar a compreensão geral da evolução das métricas da paisagem. A região correspondente aos municípios de Ipuã, Guará e São Joaquim da Barra, norte paulista, foi o foco do estudo, sendo utilizado um banco de

dados previamente organizado por Sanches (2004). O padrão das parcelas agrícolas nesta área é considerado representativo das condições encontradas, predominantemente, na agricultura paulista e do centro-sul brasileiro.

Para atender ao objetivo supracitado, usando imagens reflectância de superfície ETM+/Landsat-7 e MODIS/Terra obtidas na mesma data (05 de Janeiro de 2002), buscou-se atingir os seguintes objetivos específicos:

- estudar a evolução dos valores das métricas de paisagem em função da simulação ou degradação sucessiva de resolução espacial da imagem ETM+/Landsat-7 (30m), para valores de tamanho de pixel de 90m, 150m, 210m e 270m, utilizando duas técnicas distintas de filtragem espacial (Maioria e Textura);
- avaliar a eficiência das duas técnicas de filtragem aplicadas aos dados ETM+/Landsat-7 para simular produtos de sensores com a resolução espacial moderada desejada, pela comparação de valores de áreas das culturas agrícolas obtidos a partir de dados simulados (270m) e originais (MODIS/Terra, 250m);
- verificar as relações de regressão simples entre as estimativas de áreas obtidas com os dados originais de resolução espacial de 30m (ETM+/Landsat-7) e 250m (MODIS/Terra) para as seis classes temáticas (soja, cana-de-açúcar, milho, mata, pastagem e solo exposto) estudadas em conjunto e separadamente;
- construir um modelo de regressão múltipla que incorpore informação sobre o arranjo espacial das classes temáticas para melhorar a precisão da estimativa de suas áreas com dados originais de baixa resolução espacial (MODIS/Terra, 250m), selecionando as métricas de paisagem mais adequadas para tal finalidade.
- Traçar um paralelo entre o desempenho de modelos específicos (para cada classe, separadamente) e a performance de modelos gerais (abrangendo todas as classes).

## CAPÍTULO 2

### FUNDAMENTAÇÃO TEÓRICA

#### 2.1 A cultura do Milho

Para Duarte (2005), o milho é a mais importante planta comercial com origem nas Américas, sendo uma das culturas mais antigas do mundo. Suas diversas formas de utilização determinam sua importância econômica. Seu uso vai desde a alimentação animal até a indústria de alta tecnologia, passando pelo consumo humano.

O milho brasileiro é cultivado basicamente por um grande número de pequenos produtores, que plantam para subsistência, dando um aspecto de alta fragmentação espacial. O IBGE apresenta no censo agropecuário de 1996 essa característica, na qual apenas 2,4% dos produtores são responsáveis por 43,91% da área, com uma produção de 60,08% de todo o milho colhido no Brasil. No entanto, a grande maioria de 94,3% dos produtores de milho responde por 30% da produção, ocupando 45,63% das terras onde se cultiva o cereal. A Tabela 2.1 apresenta um quadro resumo da situação de ocupação de solo pela cultura de milho no Brasil. Duarte (2005) faz menção, ainda, à dualidade existente no cultivo do milho com muitos pequenos produtores com interesses particulares e poucos grandes produtores, com lavouras tecnológicas e visão comercial.

TABELA 2.1 – Área, produção e produtores de milho no Brasil.

Área plantada em (ha)	Área		Produção		Informantes	
	(1000 ha)	(%)	(1000 t)	(%)	(X 1000)	(%)
<b>Menos de 10</b>	4842	45,63	7654	30,00	2395	94,3
<b>10 a 20</b>	1110	10,46	2531	9,92	84	3,3
<b>20 a 100</b>	1951	18,38	5544	21,73	51	2,0
<b>Acima de 1000</b>	2709	25,53	9783	38,35	10	0,4
<b>Total</b>	10612		25512		2540	

FONTE: IBGE (2005a).

## **2.2 A cultura da Soja**

A soja é uma das principais commodities mundiais, sendo seu preço estipulado pela negociação do grão nas principais bolsas de mercadoria do mundo. É um grão de múltiplas utilidades, o que determina sua forte demanda de consumo, atingindo patamares mundiais superiores a 180 milhões de toneladas. Os Estados Unidos ocupam a primeira posição entre os países produtores, respondendo por 78 milhões de toneladas. O Brasil é o segundo maior produtor de soja e, na safra 2003, produziu cerca de 50 milhões de toneladas (Embrapa, 2005a).

A soja brasileira é produzida, de maneira geral, em todo o país. Os principais Estados produtores são o Mato Grosso, o Paraná e o Rio Grande do Sul. Até a década de 80, os Estados do Sul do Brasil dominavam a produção brasileira. No entanto, a partir da década de 90, o Centro-Oeste tornou-se a principal região produtora, sendo atualmente responsável por cerca de 40% do total da produção brasileira (Embrapa, 2005b).

É uma monocultura de larga escala com larga utilização de tecnologia agrícola. Normalmente é cultivada em grandes propriedades, sendo comum haver rotação com outras culturas como o milho safrinha, o sorgo e o milheto (Gürtler, 2003).

## **2.3 A Cultura da Cana-de-Açúcar**

Historicamente, a cana-de-açúcar é um dos principais produtos agrícolas do Brasil, sendo cultivada desde a época da colonização. Os produtos gerados de seu processo de industrialização são: o açúcar, nas suas mais variadas formas e tipos; o álcool, anidro e hidratado; o vinhoto e o bagaço (CENBIO, 2005). Devido à grande importância que assumiu para a economia nacional com a produção de álcool etílico, a cultura da cana expandiu-se por todos os Estados brasileiros, principalmente em São Paulo, Pernambuco e Alagoas, ocupando extensas áreas e com predomínio da atividade de grandes empresas agroindustriais (MCT, 2005).

A cana-de-açúcar é cultivada numa extensa área territorial, compreendida entre os paralelos 35° de latitude Norte e Sul do Equador, apresentando melhor comportamento

nas regiões quentes. O clima ideal é aquele que apresenta duas estações distintas, uma quente e úmida, para proporcionar a germinação, perfilhamento e desenvolvimento vegetativo, seguido de outra fria e seca, para promover a maturação e conseqüente acúmulo de sacarose nos colmos (Agrobyte, 2003).

Solos profundos, pesados, bem estruturados, férteis e com boa capacidade de retenção são os ideais para a cana-de-açúcar que, devido à sua rusticidade, se desenvolve satisfatoriamente também em solos arenosos e menos férteis, como os de cerrado. Solos rasos, isto é, com camada impermeável, superficial ou mal drenados, não devem ser indicados para a cana-de-açúcar (Agrobyte, 2003).

#### **2.4 A Estimativa de Área Agrícola no Brasil**

Conforme mencionado anteriormente, o IBGE realiza mensalmente um levantamento acerca das safras agrícolas nacionais: o LSPA. Neste trabalho, que foi implantado em novembro de 1972, o Instituto vai além da estimativa de área, pesquisando também a produção e o rendimento médio de 35 culturas temporárias e permanentes, de acordo com o Plano Único de Estatísticas Agropecuárias estabelecido pela Comissão Especial de Planejamento, Controle e Avaliação das Estatísticas Agropecuárias (CEPAGRO). Os procedimentos adotados no processo de investigação são específicos para adaptação às características diversas de cada cultura (IBGE, 2002).

De forma geral, durante a fase anterior ao plantio são levantadas as demandas por insumos agrícolas através de produtores, associações de classe e cooperativas. Assim faz-se a primeira estimativa de área plantada, bem como de produção esperada. Nas fases seguintes do cultivo são feitas, mensalmente, verificações para aferir a estimativa *a priori* realizada, levando-se em consideração condições climáticas e/ou fitossanitárias. Para todos os levantamentos são aplicados questionários, que serão traduzidos em números capazes de resultar na estimativa total desejada (IBGE, 2002).

Portanto, o sistema de coleta fundamenta-se no acompanhamento permanente da evolução da produção e na sua avaliação sempre atualizada, através de resultados de levantamentos diretos, bem como pelas informações complementares obtidas nos registros administrativos. Tais registros, mantidos por entidades públicas e privadas que atuam no setor, incluem informações sobre meteorologia, ação dos agentes climáticos adversos, incidência de pragas e doenças, suporte de crédito e financiamentos concedidos, comercialização, industrialização e demanda de insumos tecnológicos (sementes fiscalizadas, corretivos, fertilizantes) (IBGE, 2002).

Pode-se notar que a técnica utilizada é bastante abrangente e busca um acompanhamento durante todo o ano, de forma a aperfeiçoar os indicadores coletados. Tenta-se levantar diversas variáveis que podem ser importantes e ter um impacto direto nas previsões e cálculos realizados. No entanto, apresenta caráter subjetivo, uma vez que é realizada por pesquisas feitas junto às pessoas ligadas ao meio agrícola.

## **2.5 O Sensoriamento Remoto na Agricultura**

Jensen (2000) define sensoriamento remoto como sendo a arte e ciência de obter informação sobre um alvo sem entrar em contato direto com ele. Acrescenta, ainda, ser uma tecnologia científica que pode ser utilizada para medir e monitorar importantes características e atividades biofísicas da Terra.

Podem, então, ser identificados quatro elementos fundamentais nas atividades de sensoriamento remoto, quais sejam: a radiação eletromagnética (REM); a fonte, de onde provém a REM; o alvo, do qual se deseja obter informações e, finalmente, o sensor, que é responsável pela interceptação da REM e respectivo processamento de forma a transformá-la em informação interpretável. Assim, após a interação da REM com o alvo e conseqüente coleta pelo sensor, é formada a imagem, a partir da qual serão realizados os trabalhos.

Segundo Luiz et al. (2002), a área de cada classe de cobertura do solo em uma região pode ser estimada a partir de levantamentos por sensoriamento remoto, nos quais a cobertura do solo é interpretada em fotografias aéreas ou em imagens de satélite.

Epiphany et al. (2001) afirmam, ainda, que de posse de imagens de satélite de uma região, é possível atingir um razoável nível de precisão de interpretação do uso do solo.

O Serviço Nacional de Estatísticas Agrícolas (*National Agricultural Statistics Service - NASS*), do Departamento Norte Americano de Agricultura (*United States Department of Agriculture - USDA*), utiliza imagens de sensores orbitais desde o lançamento do primeiro satélite da série Landsat em 1972. Há três principais aplicações no programa estatístico da USDA, quais sejam: confecção do quadro de amostragem de área, estimativa de área agrícola e diagnóstico das condições das culturas (Hale et al., 1999).

Nas duas primeiras aplicações são utilizadas imagens do sensor *Thematic Mapper (TM)*, sendo a primeira o principal trabalho realizado, com a utilização de mapas digitais do Serviço Geológico Norte Americano (*United States Geological Survey - USGS*), enquanto a segunda aplicação se utiliza do quadro de amostragem e das imagens para melhorar a precisão da estimativa, em um nível estadual. Já na terceira prática, são utilizados índices de vegetação calculados do sensor *Advanced Very High Resolution Radiometer (AVHRR/NOAA-17)* para monitoramento de vastas áreas de culturas, nos diversos estágios de desenvolvimento (Hale et al., 1999).

## **2.6 Os Sistemas Sensores**

Os sensores são instrumentos capazes de captar a REM proveniente de um alvo e transformar o pulso elétrico gerado em número digital, que gera novos dados e informação através de técnicas de processamento visual e/ou digital de imagens. Tais equipamentos podem obter características muito específicas de objetos da superfície da Terra (como o diâmetro da copa de uma árvore) ou a extensão geográfica de um fenômeno (todo o contorno poligonal de uma floresta) (Jensen, 2000).

Os diversos sensores remotos existentes podem ser classificados de acordo com sua resolução espacial, que é, como afirma Jensen (2000), a medida da menor separação angular ou linear entre dois objetos capaz de ser resolvida pelo sensor. Dois grupos de sensores que são freqüentemente utilizados em estudos agrícolas são:

- Sensores com resolução espacial intermediária, muitas vezes mencionada como resolução espacial fina (Jensen, 2000), que operam numa faixa de 20m a 30m (p. ex., sensores ETM+/Landsat-7 e CCD/CBERS-2).
- Sensores com resolução espacial baixa, também mencionada como moderada. Estes instrumentos operam na faixa de centenas de metros, até quilômetros. Exemplos desses sensores são: o MODIS/Terra, AVHRR/NOAA-17, o MISR/Terra e o *Wide Field Imager* (WFI/CBERS-2).

As características espectrais e espaciais do sensor ETM+/Landsat-7 são apresentadas na Tabela 2.2. Maiores informações podem ser obtidas em USGS (2006a) e NASA (2006).

O MODIS, por sua vez, é um sensor a bordo das plataformas Terra e Aqua. A plataforma Aqua voa na direção ascendente e cruza o Equador pela tarde (13h30min.), enquanto a plataforma Terra tem movimento descendente e passa sobre o Equador pela manhã (10h30min.). Os dois satélites recobrem o globo terrestre a cada um ou dois dias. Conforme visto na Tabela 2.3, o MODIS apresenta uma configuração de 36 bandas e uma resolução espacial variável de acordo com a posição das bandas (NASA, 2005a).

Pode-se notar da inspeção da Tabela 2.3 que o MODIS apresenta duas bandas com resolução espacial de 250m (vermelho e infravermelho próximo), cinco bandas com resolução de 500m e as 29 restantes com resolução de 1000m.

Rizzi (2004) ressaltou as características positivas da resolução temporal deste sensor. Com sua larga faixa de imageamento (2230 km), o MODIS é capaz de observar diariamente praticamente toda a superfície da Terra, permitindo, assim, a geração de imagens livres de nuvens por meio da composição de várias imagens obtidas em certo período de tempo. Esta vantagem é bastante importante na aplicação em agricultura, uma vez que os meses agrícolas são bastante nebulosos, o que dificulta a obtenção de boas imagens.

TABELA 2.2 – Características do sensor ETM+/Landsat-7.

<b>Bandas</b>	<b>Comprimento de Onda (micrômetros)</b>	<b>Resolução Espacial (metros)</b>
Banda 1	0,45-0,52	30
Banda 2	0,53-0,61	30
Banda 3	0,63-0,69	30
Banda 4	0,78-0,90	30
Banda 5	1,55-1,75	30
Banda 6	10,40-12,50	60
Banda 7	2,09-2,35	30
Banda 8	0,52-0,90	15

FONTE: Adaptada de USGS (2005b).

TABELA 2.3 – Características do sensor MODIS.

<b>Bandas</b>	<b>Comprimento de Onda (µm)</b>	<b>Resolução Espacial (m)</b>	<b>Bandas</b>	<b>Comprimento de Onda (µm)</b>	<b>Resolução Espacial (m)</b>
1	0,620 – 0,670	250	19	0,915 – 0,965	1000
2	0,841 – 0,876		20	3,660 - 3,840	
3	0,459 – 0,479	500	21	3,929 - 3,989	
4	0,545 – 0,565		22	3,929 - 3,989	
5	1,230 – 1,250		23	4,020 - 4,080	
6	1,628 – 1,652		24	4,433 - 4,498	
7	2,105 – 2,155		25	4,482 - 4,549	
8	0,405 – 0,420	1000	26	1,360 - 1,390	
9	0,438 – 0,448		27	6,535 - 6,895	
10	0,483 – 0,493		28	7,175 - 7,475	
11	0,526 – 0,536		29	8,400 - 8,700	
12	0,546 – 0,556		30	9,580 - 9,880	
13	0,662 – 0,672		31	10,780 - 11,280	
14	0,673 – 0,683		32	11,770 - 12,270	
15	0,743 – 0,753		33	13,185 - 13,485	
16	0,862 – 0,877		34	13,485 - 13,785	
17	0,890 – 0,920		35	13,785 - 14,085	
18	0,931 – 0,941	36	14,085 - 14,385		

FONTE: Adaptada de NASA (2005b).

Outra característica ímpar do sensor MODIS é a disponibilização, por parte do *Earth Observing System* (EOS) – projeto que reúne diversos países com operação de diversas plataformas para o monitoramento terrestre – de 44 produtos para aplicação em áreas específicas, subdividas em *Terrestres*, *Oceânicas* e *Atmosféricas*. Para aplicação em agricultura podem ser obtidos 10 produtos distintos, entre eles o MOD09, que corresponde à imagem de reflectância de superfície, que é um dado corrigido para os efeitos de absorção e espalhamento atmosféricos para as bandas de 1 a 7 (NASA, 2005c; Vermote, 2005).

Para maiores informações acerca do sensor MODIS e de seus produtos sugere-se uma visita ao sítio oficial do programa - <http://modis.gsfc.nasa.gov/> - que apresenta descrições completas, especificações técnicas e muito material relevante para melhor conhecimento do projeto. Salomonson et al. (1989), Justice et al. (1998) e Masuoka et al. (1998) apresentam descrições que podem ser consultadas para melhor aprofundamento no assunto.

## **2.7 Padrão Espacial da Paisagem e Efeitos de Agregação Espacial**

Bailey e Gatrell (1995) definem que a análise espacial de dados deve ser utilizada quando se tem dados espacialmente distribuídos e considera-se explicitamente a importância do arranjo espacial na análise ou interpretação dos resultados.

A abordagem de informações referenciadas à localização geográfica pode ser de difícil implementação, mas é capaz de agregar muito valor às conclusões obtidas. A análise da informação de forma a extrair as vantagens da dimensão espacial consiste em definir padrões espaciais e discutir suas implicações.

Nessa linha de raciocínio, quando se analisa o efeito do padrão espacial dos elementos em sua representação em diversas escalas, Moody e Woodcock (1994) acrescentam que a estimativa de área através de sensores de baixa resolução apresenta uma deficiência devida ao efeito de agregação espacial. Segundo estes autores, a extensão desse erro depende do nível de abstração do mapa, da proporção das classes e do arranjo espacial da cobertura do solo na resolução fina. A Figura 2.1 apresenta as discrepâncias

determinadas por Mayaux e Lambin (1997) para diferentes arranjos espaciais, sendo  $P_F$  e  $P_M$  a proporção de uma cultura na escala fina e moderada, respectivamente, e  $E$  o erro devido à agregação espacial, que pode ser definido pela Equação 2.1:

$$E = (P_M - P_F) / P_F \quad (2.1)$$

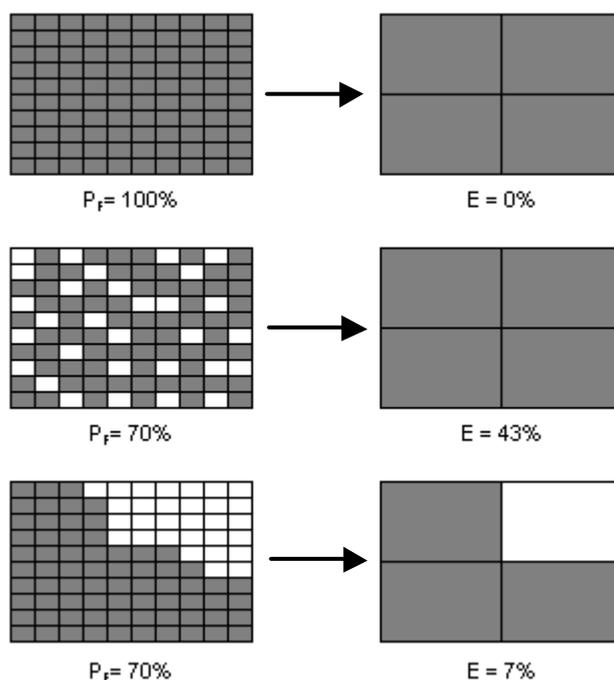


FIGURA 2.1 – Erro devido à agregação espacial na degradação da resolução espacial.  
 FONTE: Adaptada de Mayaux e Lambin (1997).

Pode-se observar na Figura 2.1 que, para valores iguais de  $P_F$ , é possível que sejam obtidos erros de estimativa de área distintos, dependendo do padrão de fragmentação da paisagem estudada. Dessa forma, fica claro que a estimativa de  $E$  não pode ser associada exclusivamente ao tamanho das áreas, devendo sua fragmentação ser considerada. Nelson e Holben (1986) corroboram esse raciocínio ao afirmarem que não é só a dimensão das feições que é relevante para detectá-las, mas também seu padrão espacial.

Diversos estudos, tais como os de Henderson-Sellers et al. (1985), Turner et al. (1989), Moody e Woodcock (1994), Moody e Woodcock (1995), Ponzoni et al. (2002), Ponzoni e Rezende (2002), Wu et al. (2002) e Frohn e Hao (2006), abordaram o tema. No entanto, estes autores basearam-se em uma classificação de uso do solo proveniente,

originalmente, de uma imagem de maior resolução (p.ex., TM/Landsat-5), degradando-a para outros níveis de resolução espacial por algoritmos ou procedimentos variados. Segundo Mayaux e Lambin (1995), quando a degradação é feita após a geração do mapa temático, há uma simplificação do problema de comparação das classificações obtidas por diferentes sensores com resoluções espaciais distintas. Eles afirmam que nessa situação há mudanças no tamanho e na forma de certas áreas e perda de pequenos objetos (abaixo de certo limiar).

Ponzoni et al. (2002) e Ponzoni e Rezende (2002), especificamente, utilizaram o procedimento de simulação de resolução espacial proposto por Banon (1990), obtendo novas imagens com as resoluções espaciais de 60m, 100m, 120m, 200m e 250m. Nesses estudos, as imagens com 200m de tamanho de elemento resolutivo apresentaram maiores discrepâncias, e as categorias que dominavam a cena, em termos espectrais e espaciais, aumentaram à medida que a resolução espacial foi degradada.

Frohn e Hao (2006), por sua vez, aplicaram processos de degradação da imagem original e da classificação temática, o que garantiu subsídio para análise da eficiência das duas metodologias. Os autores obtiveram resultados muito semelhantes para ambos os métodos de degradação, sendo tal afirmação calcada na utilização de índices que denotam o padrão espacial de culturas.

Com relação ao padrão espacial, Mayaux e Lambin (1995) afirmaram que ele pode ser quantificado por índices que utilizam diversos conceitos tais como: indicadores geométricos (fragmentação, distância simples); fractais (perímetro/área); análises espectrais (transformada de Fourier); estatística espacial (teste de auto-correlação, vizinho mais próximo); geoestatística (semi-variograma); operações de janela móvel (média, desvio padrão) ou morfologia matemática. Para exemplos de aplicação dessas técnicas para medida de padrão espacial de culturas, sugere-se a consulta aos trabalhos de Haralick (1979), Legendre e Fortrin (1989), Cullinan e Thomas (1992), Jeanjean et al. (1994) e Frohn e Hao (2006).

Qualquer técnica, no entanto, necessitará da definição rigorosa de três conceitos fundamentais que estarão sempre tangenciando todas as análises, quais sejam: a)

polígono, o qual pode ser entendido como a unidade de área que forma o mosaico da região de estudo; b) classe, que congrega todos os polígonos dotados de mesma característica física e, por fim, c) cenário, representando a totalidade, ou o universo dos elementos contidos nos estudos realizados. Como se observa, há uma relação direta de continência entre os três conceitos apresentados.

## 2.8 Métricas da Paisagem

Métricas de paisagens são valores calculados que têm a aptidão de quantificar a composição e configuração espacial de um cenário, o qual pode ser apresentado por um mapa ou uma imagem de sensoriamento remoto (Frohn e Hao, 2006).

A seguir será apresentada uma breve descrição dos conceitos envolvidos, sendo indicado como referência McGarigal e Marks (1995) e Coppedge et al. (2001).

### 2.8.1 Índices de Área/Densidade/Borda (ADB)

As principais métricas de área, densidade e borda que quantificam a composição do cenário estudado são descritas a seguir.

- 1) Área: a soma da área de todos os polígonos pertencentes a uma classe, em hectares. É uma métrica muito importante, sendo utilizada na composição de diversos índices. Seu domínio é o dos números reais positivos. Sua expressão matemática pode ser elaborada da seguinte forma:

$$Área = \sum_{j=0}^n a_{ij} \text{ [ha]} \quad (2.2)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono e  $a_{ij}$  é a área do polígono  $j$  pertencente à classe  $i$ .

- 2) Número de polígonos (NP): medida simples da quantidade de polígonos que forma cada uma das classes, denotando, assim, sua fragmentação. Portanto, não traz consigo nenhuma informação da dimensão da área, densidade ou

distribuição. Apresenta valores superiores à unidade. A Equação 2.3 declara sua formulação:

$$NP = n_i \quad (2.3)$$

Sendo:  $i$  denotando cada classe temática e  $n$  o número de polígonos da classe associada.

- 3) Porcentagem de cenário (PLAND): definida pelo somatório das áreas de todos os polígonos pertencentes a uma classe, dividido pela área total do cenário estudado e multiplicado por 100, denotando, dessa forma, a porcentagem que a classe representa em relação ao todo. Sua gama de valores possíveis vai de tão próximo de 0 quanto possível, para existência muito limitada da classe, a 100, quando todo o cenário é composto por apenas uma classe. Sua equação é:

$$PLAND = \frac{\sum_{j=0}^n a_{ij}}{A_{cen}} \times 100 \quad [\%] \quad (2.4)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono;  $a_{ij}$  é a área do polígono  $j$  pertencente à classe  $i$  e  $A_{cen}$  é a área total do cenário.

- 4) Densidade de área (PD): expressa pela razão entre o número de polígonos (NP) e a área total do cenário, multiplicado por 100. Sua diferença, quando comparado a NP, é que PD é normalizado pela área total, sendo útil quando se deseja confrontar índices calculados para cenários de diferentes dimensões. A densidade será sempre maior que zero e limitada a um valor que é função do tamanho do elemento de resolução. Esta restrição pode ser mais bem compreendida ao se analisar que o maior valor de PD será atingido quando, hipoteticamente, cada *pixel* representar um polígono de uma classe. Desta forma, o índice seria proporcional ao inverso da segunda potência do tamanho do *pixel*. A Equação 2.5 descreve, matematicamente, a métrica.

$$PD = \frac{n_i}{A_{cen}} \times 100 \quad [\text{número por } 100 \text{ ha}] \quad (2.5)$$

Sendo:  $i$  denotando cada classe temática;  $n$  o número de polígonos da classe associada e  $A_{cen}$  a área total do cenário.

- 5) Maior área (LPI): divisão da área do maior polígono de determinada classe pela área total do cenário, multiplicando-se por 100 para transformar em porcentagem, conforme apresenta a Equação 2.6. Este índice denota a porcentagem da área total que é recoberta pelo maior elemento da classe em tela, sendo superior a zero e menor ou igual a 100, quando um polígono domina completamente.

$$LPI = \frac{\max_{j=1}^a(a_{ij})}{A_{cen}} \times 100 \quad [\%] \quad (2.6)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono;  $a_{ij}$  é a área do polígono  $j$  pertencente à classe  $i$  e  $A_{cen}$  é a área total do cenário.

- 6) Forma de classe (LSI): igual ao perímetro da classe (somatório do comprimento, em *pixels*, das bordas de todos os polígonos pertencentes à classe) dividido pela dimensão mínima (em *pixels*) que essa grandeza assumiria caso seu padrão espacial fosse de um polígono compacto. É uma medida da complexidade da forma da classe, uma vez que compara o arranjo real com a unidade mínima de agregação. Sua característica lhe confere valores não inferiores à unidade, inexistindo limite superior, conforme se pode concluir da Equação 2.7.

$$LSI = \frac{e_i}{\min(e_i)} \quad (2.7)$$

Onde:  $e_i$  é o comprimento de borda (perímetro) da classe  $i$ , em *pixels*, e  $\min(e_i)$  é o perímetro mínimo (em *pixels*), caso o padrão espacial da classe fosse de um polígono compacto.

- 7) Forma de classe normalizada (NLSI): O conceito aqui envolvido é o mesmo já mencionado para o LSI, com a diferença que os dados são escalados de forma a

apresentar seu domínio no intervalo fechado de 0 a 1. Na formulação apresentada na Equação 2.8 é inserido agora o conceito de perímetro máximo, que traz significado oposto ao supramencionado  $min(e_i)$ .

$$NLSI = \frac{e_i - min(e_i)}{max(e_i) - min(e_i)} \quad (2.8)$$

Sendo:  $e_i$  o comprimento de borda (perímetro) da classe  $i$ , em pixels;  $min(e_i)$  o perímetro mínimo (em pixels), caso o padrão espacial da classe fosse de um polígono compacto e  $max(e_i)$  o perímetro máximo (em pixels), caso o padrão espacial da classe fosse tão disperso quanto possível.

### 2.8.2 Índices de Área Nuclear (Core)

A área nuclear, que em inglês é denominada *core*, é definida por Millington et al. (2003) como o interior do polígono que compõe a área, não sendo, portanto, influenciada pelo efeito de borda causado pela interação com os elementos adjacentes. Métricas que se apóiam neste conceito são apresentadas a seguir:

- 1) Área nuclear total (TCA): a soma da área nuclear (core) de todos os polígonos pertencentes a uma classe, em hectares, podendo ser apresentada conforme a Equação 2.9.

$$TCA = \sum_{j=1}^n a_{ij}^c \text{ [ha]} \quad (2.9)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono e  $a_{ij}^c$  é a área nuclear do polígono  $j$  pertencente à classe  $i$ .

- 2) Porcentagem nuclear de cenário (CPLAND): divisão da área nuclear total da classe pela totalidade do cenário estudado, multiplicado por 100 para transformar em porcentagem, o que confere valores de 0 a 100 à métrica (Equação 2.10). Esta métrica traz a normalização pela área total, o que facilita nas análises comparativas entre cenários de dimensões distintas.

$$CPLAND = \frac{\sum_{j=1}^n a_{ij}^c}{A_{cen}} \times 100 \text{ [%]} \quad (2.10)$$

Sendo:  $i$  denotando cada classe temática;  $j$  equivalente a cada polígono;  $a_{ij}^c$  a área nuclear do polígono  $j$  pertencente à classe  $i$  e  $A_{cen}$  é a área total do cenário.

- 3) Número de áreas nucleares disjuntas (NDCA): quantificação simples do número de áreas nucleares disjuntas presentes no cenário, podendo assumir quaisquer valores não negativos. É uma alternativa ao índice número de polígonos (NP), quando se fizer necessária a eliminação do efeito de adjacência.

$$NDCA = \sum_{j=1}^n n_{ij}^c \quad (2.11)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono e  $n_{ij}^c$  o número de polígonos nucleares da classe associada.

- 4) Densidade de área nuclear disjunta (DCAD): divisão do NDCA pela área total do cenário, em hectares, e multiplicação por 100. Traz a vantagem da normalização do dado. Sua formulação é expressa na Equação 2.12.

$$DCAD = \frac{\sum_{j=1}^n n_{ij}^c}{A_{cen}} \times 100 \text{ [%]} \quad (2.12)$$

Sendo:  $n_{ij}^c$  o número de polígonos nucleares da classe associada e  $A_{cen}$  a área total do cenário.

### 2.8.3 Índices de Contágio/Espalhamento (CE)

Espalhamento denota a existência de mistura de polígonos pertencentes a classes diferentes em dada vizinhança. Contágio, além desse conceito, indica também a tendência de classes formarem conglomerados, apresentando-se espacialmente

agregadas. Tendo em vista que essas métricas avaliam a distribuição espacial da informação, elas descrevem aspectos de textura. Segue a apresentação de índices dessa categoria:

- 1) Porcentagem de adjacência (PLADJ): mede o grau de agregação da classe estudada, através do uso de uma medida chamada auto-adjacência, que mede a quantidade de vezes que aparece um encontro entre *pixels* de mesma classe. Dessa forma, PLADJ expressa a relação entre a auto-adjacência e o número total de células de adjacência que envolvem a classe estudada (Equação 2.13). A multiplicação por 100 gera a unidade de medida em porcentagem.

$$PLADJ = \left( \frac{g_{ii}}{\left( \sum_{k=1}^m g_{ik} \right)} \right) \times 100 \text{ [%]} \quad (2.13)$$

Onde:  $g_{ii}$  é a auto-adjacência da classe  $i$  e  $g_{ik}$  é o número de adjacências entre as classes  $i$  e  $k$ .

- 2) Agregação (AI): calcula uma medida da vizinhança existente entre células de mesma classe. Para tal, apresenta-se como a razão entre o número de adjacências existentes entre *pixels* de mesma classe e o valor máximo possível de contigüidade, que é obtido quando a classe se caracteriza por uma única área compacta. Esse valor é multiplicado por 100, para que sua unidade seja a porcentagem (Equação 2.14). Conforme se nota, esta métrica calcula uma medida interna da classe estudada, não fazendo alusão às demais e seu valor denota o quão compacta ela se apresenta.

$$AI = \left( \frac{g_{ii}}{\text{máx} \rightarrow g_{ii}} \right) \times 100 \text{ [%]} \quad (2.14)$$

Onde:  $g_{ii}$  é a auto-adjacência da classe  $i$  e  $\text{máx} \rightarrow g_{ii}$  é o valor máximo possível de contigüidade, que é obtido quando a classe se caracteriza por uma única área compacta.

3) Interposição e Justaposição (IJI): definindo tipo de borda como o limite entre duas classes específicas  $i$  e  $k$ , o IJI pode ser enunciado como o oposto ao somatório do comprimento de cada tipo de borda único, envolvendo a classe em questão, dividido pela dimensão total das bordas que a classe estudada faz com as demais, multiplicado pelo logaritmo natural deste mesmo quociente; dividido pelo logaritmo neperiano de 1 menos o número de classes. Para transformação em porcentagem, multiplica-se, ainda, pela centena. A Equação 2.15 ajudará a melhor elucidar o conceito trazido pela métrica, que é uma medida de diversidade e é baseada na extensão da adjacência entre áreas de classes diferentes.

$$IJI = \frac{-\sum_{k=1}^m \left[ \left( \frac{e_{ik}}{\sum_{k=1}^m e_{ik}} \right) \cdot \ln \left( \frac{e_{ik}}{\sum_{k=1}^m e_{ik}} \right) \right]}{\ln(m-1)} \times 100 \text{ [%]} \quad (2.15)$$

Sendo:  $e_{ik}$  o comprimento de borda, em metros, entre as classes  $i$  e  $k$ , em pixels;  $\min(e_i)$  o perímetro mínimo (em *pixels*), caso o padrão espacial da classe fosse de um polígono compacto e  $\max(e_i)$  o perímetro máximo (em *pixels*), caso o padrão espacial da classe fosse tão disperso quanto possível.

4) Agrupamento (CLUMP): Do inglês, *clumpiness*, este índice é uma medida da vizinhança entre *pixels* de mesma classe. Para defini-lo faz-se necessária a aceção da variável proporção de adjacência ( $G_i$ ), conforme Equação 2.16. Assim, CLUMP mede o desvio proporcional de  $G_i$  do valor que esta deveria apresentar num caso de uma distribuição espacial aleatória. Isso é feito comparando a proporção de adjacência com a proporção do cenário que é ocupado pela classe em tela ( $P_i$ ). A Equação 2.17 enuncia esta métrica, que assume valores do intervalo fechado [-1,1]. Valores próximos a -1 serão obtidos quando a classe estudada apresentar máxima desagregação; próximos a 0 quando houver aleatoriedade; e próximos a 1 para máxima agregação.

$$G_i = \left( \frac{g_{ii}}{\left( \sum_{k=1}^m g_{ik} \right) - \text{mín}(e_i)} \right) \quad (2.16)$$

$$CLUMP = \left[ \begin{array}{l} \frac{G_i - P_i}{P_i} \text{ para } G_i < P_i \text{ e } P_i < 0.5, \text{ ou} \\ \frac{G_i - P_i}{1 - P_i} \end{array} \right] \quad (2.17)$$

Onde:  $g_{ii}$  é a auto-adjacência da classe  $i$ ;  $g_{ik}$  é o número de adjacências entre as classes  $i$  e  $k$ ;  $P_i$  é a proporção do cenário que é ocupado pela classe  $i$  e  $\text{mín}(e_i)$  o perímetro mínimo (em *pixels*), caso o padrão espacial da classe fosse de um polígono compacto

- 5) Divisão de cenário (DIV): interpretado como a probabilidade de dois *pixels* escolhidos aleatoriamente não se situarem na mesma área da classe em foco. Matematicamente isto é obtido subtraindo-se da unidade a razão entre o somatório das áreas da classe pela área total do cenário. A Equação 2.18 formaliza o conceito de heterogeneidade trazido pelo índice, que se torna maior à medida que o cenário se divide mais finamente.

$$DIV = \left[ 1 - \sum_{j=1}^n \left( \frac{a_{ij}}{A_{cen}} \right)^2 \right] \quad (2.18)$$

Onde:  $i$  denota a cada classe temática;  $j$  equivale a cada polígono;  $a_{ij}$  é a área do polígono  $j$  pertencente à classe  $i$  e  $A_{cen}$  é a área total do cenário.

- 6) Fragmentação (SPLIT): número de polígonos, de igual tamanho, da classe estudada, necessários para compor todo o cenário, indicando, portanto, o grau de subdivisão apresentado pela classe. Esta métrica é obtida pela divisão do quadrado da área do cenário pelo somatório das áreas quadráticas dos polígonos pertencentes à mesma classe, conforme mostrado pela Equação 2.19. *SPLIT* pode assumir um valor mínimo de 1, que é quando o cenário é formado por um

único polígono, até um limite superior igual ao quadrado do número de *pixels* que compõe a imagem, situação em que a classe é composta por apenas uma célula. Note-se a relação existente com o índice Divisão de Cenário, que é igual a 1 menos o inverso de *SPLIT*.

$$SPLIT = \frac{A_{cen}^2}{\sum_{j=1}^n a_{ij}^2} \quad (2.19)$$

Sendo: *i* denotando a cada classe temática; *j* equivalente a cada polígono; *a<sub>ij</sub>* a área do polígono *j* pertencente à classe *i* e *A<sub>cen</sub>* a área total do cenário.

- 7) Tamanho de trama efetiva (*MESH*): tamanho dos polígonos necessários para efetuar a operação sugerida pela métrica *SPLIT*. Obtém-se pela divisão da soma dos quadrados das áreas que compõem determinada classe pela área total do cenário (Equação 2.20).

$$MESH = \frac{\sum_{j=1}^a a_{ij}^2}{A_{cen}} \quad [\text{ha}] \quad (2.20)$$

Onde: *i* denota a cada classe temática; *j* equivale a cada polígono; *a<sub>ij</sub>* é a área do polígono *j* pertencente à classe *i* e *A<sub>cen</sub>* a área total do cenário.

#### 2.8.4 Índices de Forma e Conectividade

- 1) Dimensão fractal perímetro-área (*PAFRAC*): é uma métrica muito interessante, pois reflete a complexidade da forma através de certo intervalo de tamanhos de polígonos. Tal propriedade é relevante uma vez que o índice tradicional de forma, que é o razão perímetro-área (*PARA*), é muito sensível à variação de área. Por exemplo, aumentando-se a área de um polígono sem alterar a forma, resultaria na redução do *PARA*. Em sua formulação *PAFRAC* é igual a 2 dividido pela inclinação da reta da regressão entre o logaritmo neperiano da área e o logaritmo natural do perímetro (Equação 2.21), podendo assumir, portanto, o

intervalo: 1 = PAFRAC = 2. Valores próximos do limite superior indicam alta complexidade de forma.

$$PAFRAC = \frac{2 \left[ n_i \sum_{j=1}^n (\ln p_{ij} \cdot \ln a_{ij}) \right] - \left[ \sum_{j=1}^n (\ln p_{ij}) \cdot \sum_{j=1}^n (\ln a_{ij}) \right]}{\left( n_i \sum_{j=1}^n \ln p_{ij}^2 \right) - \left( \sum_{j=1}^n \ln p_{ij} \right)^2} \quad (2.21)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono;  $a_{ij}$  é a área do polígono  $j$  pertencente à classe  $i$ ;  $p_{ij}$  é o perímetro do polígono  $j$  pertencente à classe  $i$  e  $n_i$  é o número de polígonos pertencentes à classe  $i$ .

- 2) Coesão de área (COHE): medida de conectividade, o COHE aumenta quando a configuração espacial dos polígonos é mais agregada, ou seja, fisicamente conectados. O índice varia de 0 (inclusive) a 100 (exclusive), sendo o menor valor atingido quando a classe é composta por um único *pixel* isolado, crescendo à medida que se formam porções conectas do cenário. Sua expressão matemática é apresentada pela Equação 2.22.

$$COHE = \left( I - \frac{\sum_{j=1}^n p_{ij}}{\sum_{j=1}^n p_{ij} \sqrt{a_{ij}}} \right) \cdot \left( I - \frac{I}{\sqrt{A_{cen}}} \right)^{-1} \times 100 [\%] \quad (2.22)$$

Onde:  $i$  denota cada classe temática;  $j$  equivale a cada polígono;  $a_{ij}$  é a área do polígono  $j$  pertencente à classe  $i$ ;  $p_{ij}$  é o perímetro do polígono  $j$  pertencente à classe  $i$  e  $A_{cen}$  a área total do cenário.

## 2.9 Simulação da Resolução Espacial Moderada

Pax-Lenney e Woodcock (1997) associaram a resolução espacial à variância dos dados. Assim, imagens com resolução espacial moderada tendem a apresentar valores reduzidos para essa estatística. Eles alertam, no entanto, que o método utilizado para

degradação dos dados para outros níveis de resolução espacial, também influenciará na relação entre a variabilidade dos dados e a resolução espacial. Henderson-Sellers e Pitman (1992), Malingreau e Belward (1992), Belward e Lambin (1990) apresentaram experimentos com diferentes tipos de degradação espacial, dos quais se pode concluir que a perda de variabilidade para resoluções moderadas está associada ao método utilizado para geração da imagem.

Portanto, uma forma simples de obter a resolução espacial moderada é pela degradação da resolução fina. Dessa forma, é possível a simulação de qualquer tipo de resolução espacial, permitindo, assim, efetuarem-se análises em um ambiente controlado. Justice et al. (1989), Saura (2004), Frohn e Hao (2006) aplicaram filtragens espaciais em imagens de resolução espacial fina como forma de simular a resolução moderada desejada, sendo as técnicas de filtragem de maioria e de textura bastante utilizadas nesse contexto.

### **2.9.1 Filtragem Espacial de Maioria**

O método mais comum de comparação do padrão espacial da paisagem com relação à agregação dos dados é pela transformação da escala de mapas temáticos por intermédio de filtros com regra de maioria (Frohn e Hao, 2006).

O conceito aqui envolvido é bastante simples. Uma janela móvel de dimensão  $n \times m$  (definida pelo usuário) percorre toda a imagem, sendo aplicado ao *pixel* central da janela o valor mais freqüente no *kernel* utilizado. Dessa forma, áreas pouco representativas serão removidas como ruído. A Figura 2.2 simula a operação desse filtro, cujo efeito de suavização é proporcional ao tamanho da janela.

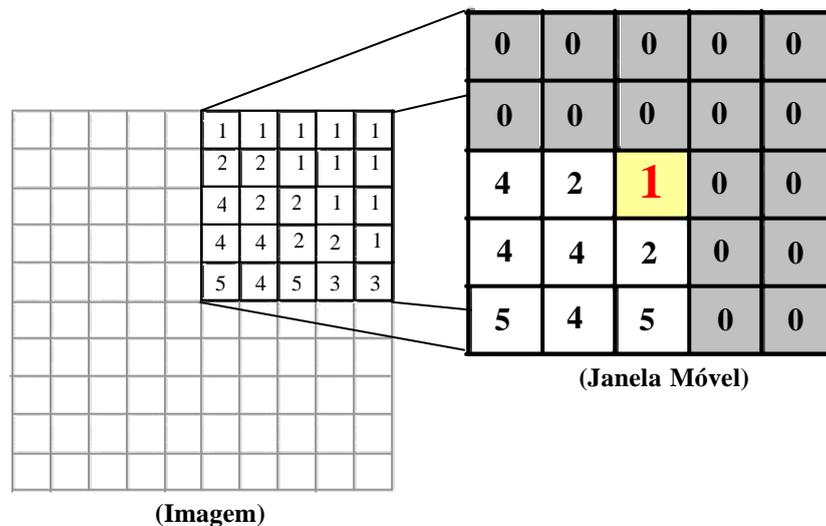


FIGURA 2.2 – Filtragem de Maioria com dimensão  $5 \times 5$ .

Pode-se observar da Figura 2.2 que, no recorte de imagem correspondente à janela móvel, o *pixel* central foi substituído pelo valor de maior ocorrência. Todos os demais *pixels* mantêm o valor original, com exceção da borda da imagem cujos elementos não apresentarão informação.

Segundo alguns autores (p.ex., Saura, 2004), a simplicidade desse método traz, no entanto, o inconveniente de produzir mapas menos fragmentados do que os que seriam normalmente produzidos por um sensor de mesma resolução espacial. Assim, a filtragem de textura pode ser uma alternativa à filtragem de maioria em estudos da influência da resolução espacial na estimativa de áreas agrícolas.

### 2.9.2 Filtragem Espacial de Textura

A textura está relacionada à distribuição espacial (estatística) dos tons de cinza e é uma das características importantes usadas na identificação de objetos ou regiões de interesse em uma imagem. Seu uso é muito importante, uma vez que ela traz informações relevantes para discriminação de feições da imagem (Haralick et al., 1973).

Para descrever este conceito, faz-se necessário definir algumas variáveis. Suponha uma imagem  $I$  com  $N_t$  possíveis tons de cinza, que seja formada por  $N_x$  células horizontais e  $N_y$  verticais. Isso permite a definição dos seguintes conjuntos:  $C_x = \{1, 2, 3, \dots, N_x\}$ ,  $C_y =$

$\{1,2,3,\dots,N_y\}$  e  $T = \{1,2,3,\dots,N_t\}$ , que representam, respectivamente, o domínio espacial horizontal e vertical e o domínio dos tons de cinza (nível de quantização da imagem). Assim,  $I$  pode ser definida pela seguinte função:

$$I: C_y \times C_x \textcircled{R} T \quad (2.23)$$

Uma vez que o conceito de textura está diretamente ligado ao conceito de adjacência, definem-se os oito vizinhos de um *pixel* segundo a notação da Figura 2.3, com as células 1 e 5, 4 e 8, 3 e 7, 2 e 6 sendo, respectivamente, os vizinhos  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ .

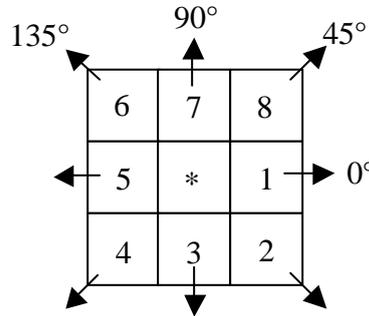


FIGURA 2.3 – Definição da vizinhança do *pixel* \*.

FONTE: Haralick et al. (1973, p. 612).

A informação de textura pode ser entendida, então, como a freqüência  $F_{ij}$  com que ocorrem dois elementos de resolução com valores de quantização  $i$  e  $j$ , separados por uma distância  $d$  (definidor da adjacência), em uma das quatro direções definidas. Fica caracterizada, portanto, a distribuição espacial dos níveis de cinza em uma vizinhança.

Pode-se definir, a partir de então, o modelo rigoroso da freqüência  $F$  para as quatro direções:

$$F(i,j,d,0^\circ) = Q\{((k,l),(m,n)) \in (C_y \times C_x) \times (C_y \times C_x) \mid k - m = 0, \\ |l - n| = d, I(k,l) = i, I(m,n) = j\} \quad (2.24)$$

$$F(i,j,d,45^\circ) = Q\{((k,l),(m,n)) \in (C_y \times C_x) \times (C_y \times C_x) \mid (k - m = d, \\ l - n = -d) \text{ ou } (k - m = -d, l - n = d), I(k,l) = i, I(m,n) = j\} \quad (2.25)$$

$$F(i,j,d,90^\circ) = Q\{((k,l),(m,n)) \in (C_y \times C_x) \times (C_y \times C_x) \mid |k - m| = d,$$

$$l - n = 0, \mathbf{I}(k,l) = i, \mathbf{I}(m,n)=j \quad (2.26)$$

$$F(i,j,d,135^\circ) = Q\{((k,l),(m,n)) \in (C_y \times C_x) \times (C_y \times C_x) \mid (k - m = d, \\ l - n = d) \text{ ou } (k - m = -d, l - n = -d) \mathbf{I}(k,l) = i, \mathbf{I}(m,n)=j\} \quad (2.27)$$

Onde  $Q$  representa a quantidade de elementos no conjunto e  $F$  é denominada Matriz de Dependência Espacial dos Tons de Cinza (MDETC).

Para exemplificar o conceito da MDETC, a Figura 2.4 apresenta uma imagem hipotética com nível de quantização de 2 bits, bem como as matrizes para as quatro direções discutidas.

Pode-se observar, analisando a Figura 2.4, o processo para confecção da MDETC para qualquer direção. Tomando por base a  $F_{0^\circ}$ , nota-se que  $Q(0,0) = 4$ , pois, no universo da imagem, ocorre 4 vezes a dupla de tons de cinza 0-0 na direção horizontal;  $Q(0,1) = 2$ , uma vez que o encontro 0-1 aparece 2 vezes na direção  $0^\circ$ ; e assim por diante.

A definição precisa da MDETC é fundamental, pois é ela que traz toda a informação de textura de uma imagem. Assim, com base nela, são criadas medidas estatísticas de textura.

Haralick et al. (1973) e Haralick (1979) descreveram de maneira pormenorizada os conceitos e formulação envolvida na filtragem de textura, bem como expuseram diversas medidas estatísticas usadas para sua caracterização. Portanto, recomenda-se a leitura destes trabalhos para um aprofundamento do assunto.

## 2.10 Classificação Digital de Imagens

A classificação temática é o processo responsável pela reprodução, a partir de imagens digitais, da distribuição espacial de feições da superfície terrestre, o que permite a confecção de mapas temáticos (Schowengerdt, 1997). Por meio desse procedimento, áreas que apresentam características semelhantes são agrupadas em categorias chamadas classes.

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

(a)

		Tons de Cinza			
		0	1	2	3
Tons de Cinza	0	$Q(0,0)$	$Q(0,1)$	$Q(0,2)$	$Q(0,3)$
	1	$Q(1,0)$	$Q(1,1)$	$Q(1,2)$	$Q(1,3)$
	2	$Q(2,0)$	$Q(2,1)$	$Q(2,2)$	$Q(2,3)$
	3	$Q(3,0)$	$Q(3,1)$	$Q(3,2)$	$Q(3,3)$

(b)

$$F_{0^\circ} = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

(c)

$$F_{90^\circ} = \begin{pmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

(d)

$$F_{135^\circ} = \begin{pmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

(e)

$$F_{45^\circ} = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

(f)

FIGURA 2.4 – (a) Imagem 4×4, 2 bits; (b) Forma geral de qualquer MDETC; (c)–(f) Todas MDETC para distância unitária.

FONTE: Adaptada de Haralick et al. (1973, p. 613).

Mather (1999) divide o processo de classificação em dois estágios: o primeiro é o reconhecimento de categorias de objetos do mundo real, como, por exemplo, área agrícola, corpos d'água, florestas entre outras; o segundo estágio é responsável pela rotulação de cada *pixel* com um valor numérico, o qual será associado a uma das classes definidas. Dessa forma, há uma transformação de um espaço discreto, formado pelos elementos de resolução, para um contínuo, composto pelos polígonos da classificação.

Pode-se distinguir duas metodologias de extração das classes de uma imagem, quais sejam a classificação supervisionada e a não-supervisionada, das quais será exposta uma breve descrição.

### 2.10.1 Classificação Supervisionada

É um método que exige pré-conhecimento das classes que estão contidas na imagem, uma vez que há necessidade de coleta de amostras de treinamento que serão utilizadas como parâmetro para atribuição da classe a que cada *pixel* pertence.

Richards (1986) ressalta que as principais técnicas de classificação supervisionada adotadas em sensoriamento remoto supõem que cada classe espectral pode ser descrita como uma distribuição de probabilidade num espaço multiespectral, a qual descreve a chance de determinado *pixel* pertencer a cada classe. A Figura 2.5 apresenta um espaço multiespectral bidimensional, formado pela integração de hipotéticas bandas 1 e 2, podendo-se observar a distribuição estatística das classes, com  $P(x)$  representando a probabilidade de *pixels* pertencerem a cada classe. Os limites de decisão correspondem ao lugar geométrico dos pontos com igual probabilidade de pertencer a duas classes.

Para formar os agrupamentos apresentados na Figura 2.5, é necessário que o operador selecione, na imagem, regiões que tenham comportamento semelhante ao das classes que se deseja mapear. As áreas coletadas são chamadas de amostras de treinamento e é a partir da estatística delas que se extrapolará a classificação para toda a imagem.

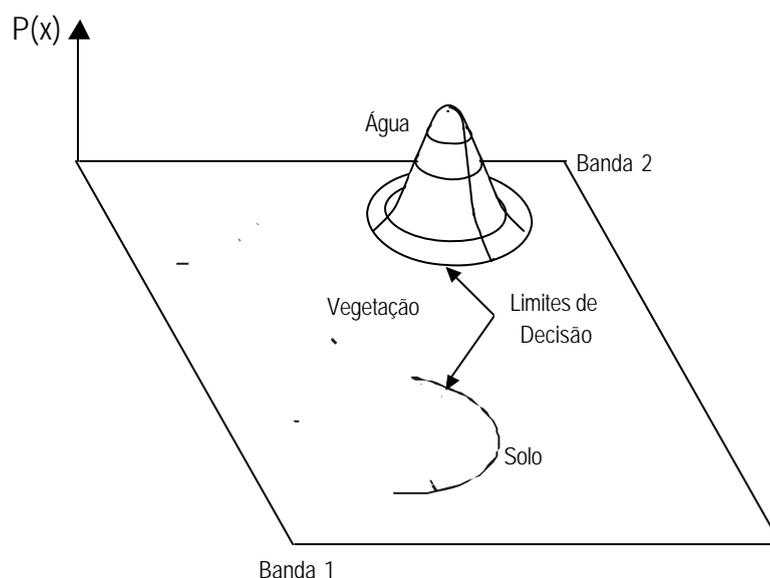


FIGURA 2.5 – Espaço multiespectral bidimensional, representando a probabilidade de *pixels* pertencerem a cada classe –  $P(x)$ .  
FONTE: Adaptada de Richards (1986, p. 80).

### 2.10.2 Classificação Não-Supervisionada

A classificação não-supervisionada, em oposição ao que foi supra-exposto, não exige prévia seleção das classes que deverão estar presentes no mapa temático. Dessa forma, o algoritmo procura nos dados os padrões existentes, agregando áreas com semelhante comportamento espectral na mesma categoria. Richards (1986) considera que, dado que muitas vezes o usuário não tem nenhum conhecimento das categorias presentes na cena, tal procedimento é útil para determinar a composição espectral das classes que compõem a imagem antes de análises mais detalhadas pelos métodos supervisionados de classificação.

O procedimento de importantes algoritmos, como o *K-means*, por exemplo, é encontrar uma divisão ótima do espaço de atributos considerando o número de subconjuntos especificados (Schowengerdt, 1997). Na Figura 2.6 é apresentada a aplicação do algoritmo *K-means* em um conjunto de dados bidimensional.

Como injunção *a priori*, um vetor de médias iniciais é especificado arbitrariamente para cada um dos  $K$  agrupamentos especificados (Figura 2.6a). Um *pixel* será alocado a uma classe quando seus vetores das médias forem o mais próximos quanto possível. Assim, forma-se a primeira partição do espaço de atributos. As médias para as novas classes são calculadas e, mais uma vez, os *pixels* serão mapeados de acordo com a proximidade dos vetores das médias. Esse processo iterativo ocorrerá até que um critério de interrupção seja atingido (Figuras 2.6b, c, d). Como exemplos de condições para encerrar o processo, podem-se citar o número de iterações e a variação dos vetores das médias (Figura 2.6e) (Schowengerdt, 1997; Mather, 1999).

### 2.11 Análise de Regressão

Segundo Neter et al. (1996), a análise de regressão é uma ferramenta estatística que utiliza a relação existente entre duas ou mais variáveis, de forma que uma variável pode ser predita a partir de outra (ou outras). Dessa forma, pode-se enunciar a Equação 2.28, que relaciona uma variável dependente e um conjunto de variáveis explicativas (independentes):

$$Y = \sum_{i=1}^p \beta_i \cdot X_i + e \quad (2.28)$$

Onde  $Y$  é a variável dependente, a qual se deseja ser explicada;  $X_i$  são as variáveis independentes;  $\beta_i$  são os  $p$  parâmetros da regressão; e, por fim,  $e$  é o termo relativo aos resíduos.

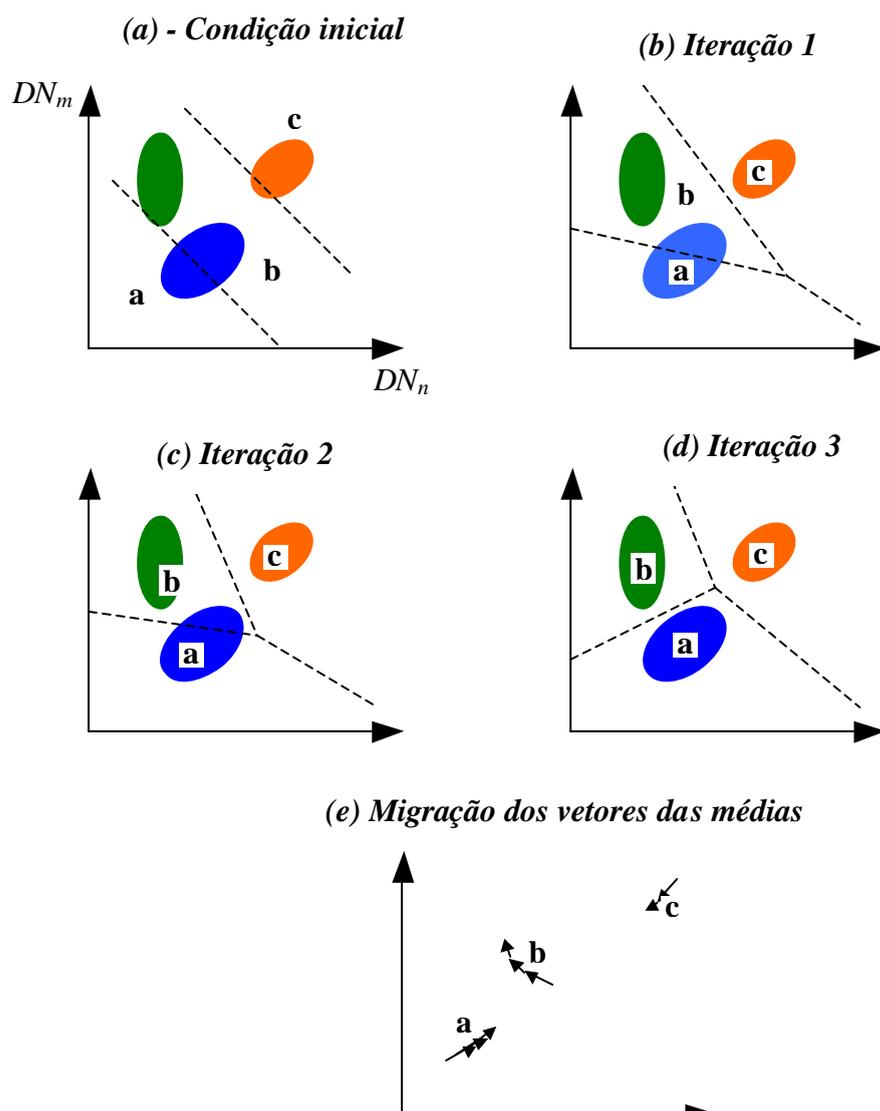


FIGURA 2.6 – Simulação (a) da condição inicial; (b), (c) e (d) de três iterações efetuadas pelo algoritmo K-means; e (e) da migração resultante dos vetores das médias.

FONTE: Adaptada de Schowengerdt (1997, p. 405).

Fonseca et al. (1985) conceituaram, ainda, a análise de regressão como um conjunto de técnicas e métodos para o estabelecimento de fórmulas empíricas que interpretem a relação funcional entre variáveis, com boa aproximação.

Para compor uma análise de regressão é necessário um conjunto de observações para a variável independente e suas correspondentes para as dependentes. Ao projetar esses pontos em um espaço  $n+1$  dimensional, onde  $n$  é o número de variáveis independentes e cada variável compõe um eixo coordenado, é possível avaliar o quão bem o arranjo de pontos se ajusta, ou seja, a habilidade das variáveis independentes em explicar a dependente.

Por caráter puramente didático, muitas argumentações, análises e exemplos apresentados neste trabalho foram desenvolvidos com o uso da regressão simples (apenas uma variável independente), valendo ressaltar, no entanto, que todas as assertivas são válidas para modelos de regressão múltipla.

A utilização da análise de regressão em qualquer aplicação requer uma série de estudos preliminares dos dados, de forma que se possa identificar e, se possível, corrigir eventuais inconsistências. Análise da significância de variáveis, estudo da adequação dos dados a premissas básicas, pesquisa sobre a existência de observações espúrias, exame da correlação entre as variáveis independentes são exemplos de tarefas a serem executadas para obter a melhor performance do modelo estatístico elaborado. Ademais, faz-se necessária uma cuidadosa escolha das variáveis a serem incorporadas na análise, bem como um julgamento da validade da equação de regressão resultante. Uma sucinta explanação acerca desses temas é apresentada nesta seção.

### **2.11.1 Coeficiente de Determinação**

O modelo de regressão perfeito deve permitir que o conjunto de pares coordenados se acumule sobre uma linha reta, conforme mostrado na Figura 2.7a. Assim, para medir a capacidade de a variável dependente prever a independente, ou seja, o quão longe da linha a massa de pontos se apresenta, basta calcular a dispersão dos pontos em relação à

reta que melhor se ajusta à nuvem de pares de observações, conforme ilustra a Figura 2.7b.

O coeficiente de determinação ( $R^2$ ), também conhecido como coeficiente de explicação, é uma medida do agrupamento dos pontos no entorno da reta de regressão e pode ser interpretado como a redução proporcional da variação total das observações da variável dependente, associada com o uso da variável preditiva  $X$  (Neter et al., 1996). A Equação 2.29 formaliza o conceito, sendo  $\hat{Y}_i$  o valor predito pela equação de regressão para a variável dependente,  $Y_i$  o valor observado da variável dependente e, por fim,  $\bar{Y}$  a média das observações da variável dependente.

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad (2.29)$$

O coeficiente de determinação assume valores de 0 a 1 e à medida que se observa o conjunto de pares coordenados mais próximos da reta de regressão, há uma tendência para um aumento do  $R^2$ .

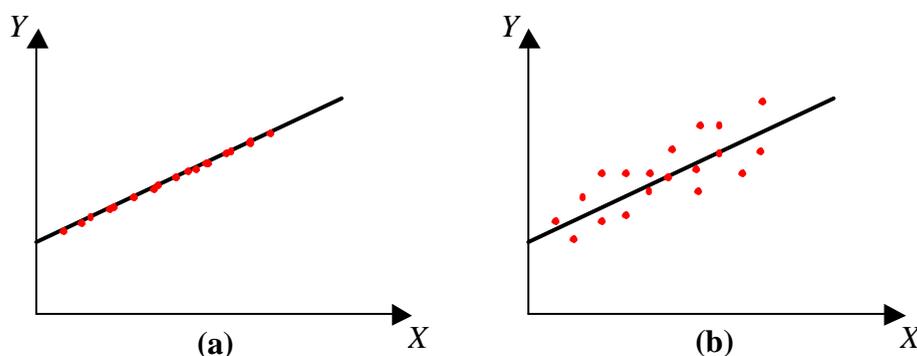


FIGURA 2.7 – Variável independente ( $X$ ) predizendo dependente ( $Y$ ): (a) Perfeita adequação; (b) nuvem de pontos no entorno de uma reta.

### 2.11.2 Premissas Fundamentais da Análise de Regressão

Segundo Neter et al. (1996) é bastante importante ressaltar que o valor de  $R^2$  não deve ser analisado de maneira isolada, tendo em vista que há algumas premissas básicas sobre as quais é edificada a teoria de regressão linear. Dessa forma, deve-se observar a

normalidade da distribuição estatística da variável dependente, normalidade dos resíduos, homocedasticidade (constância da variância) dos resíduos, independência dos resíduos e, por fim, linearidade da função de regressão. A não observância de um desses princípios pode causar instabilidade no cálculo dos coeficientes da regressão ou limitação em seus intervalos de confiança.

Como os dados utilizados nas análises estatísticas são, normalmente, frutos de coletas de informações reais de algum fenômeno, é comum pelo menos uma das injunções iniciais não serem respeitadas. Assim, são necessárias transformações nas variáveis a fim de suprir as exigências e, deste modo, criar modelos robustos.

### **2.11.2.1 Análise de Normalidade**

Em procedimentos de regressão é necessário fazer duas análises básicas de normalidade. A primeira, que deve dar início ao trabalho de exploração dos dados, é a análise da normalidade da variável dependente; a segunda é o estudo de normalidade dos resíduos da regressão.

Três diagramas básicos podem trazer informação importante sobre a distribuição estatística dos dados. São eles:

- **Histograma:** É um dos mais conhecidos diagramas para estudo da distribuição de um conjunto de dados. Ele divide o domínio do conjunto estudado em intervalos e exprime o número de observações contidas em cada um deles. A Figura 2.8a apresenta um exemplo deste tipo de gráfico;
- **Box-plot:** Tukey (1977) apresenta o conceito que expressa o intervalo de variação de uma variável, através de quartis ou desvios padrões, bem como sua tendência central, por meio da média ou mediana. Este diagrama pode ser utilizado, também, para identificação de *outliers*, sendo exemplificado na Figura 2.8b;
- **Probabilidade normal:** Baseado em um teste de correlação, esta técnica realiza um procedimento de regressão simples cuja variável explicativa é composta

pelas observações brutas, e a independente representa seu valor normal esperado, conforme ilustra a Figura 2.8c. O critério para atingir uma conclusão acerca da normalidade é o ajustamento dos pontos à reta de regressão, o qual é medido pelo coeficiente de correlação (D'Agostino e Stephens, 1986).

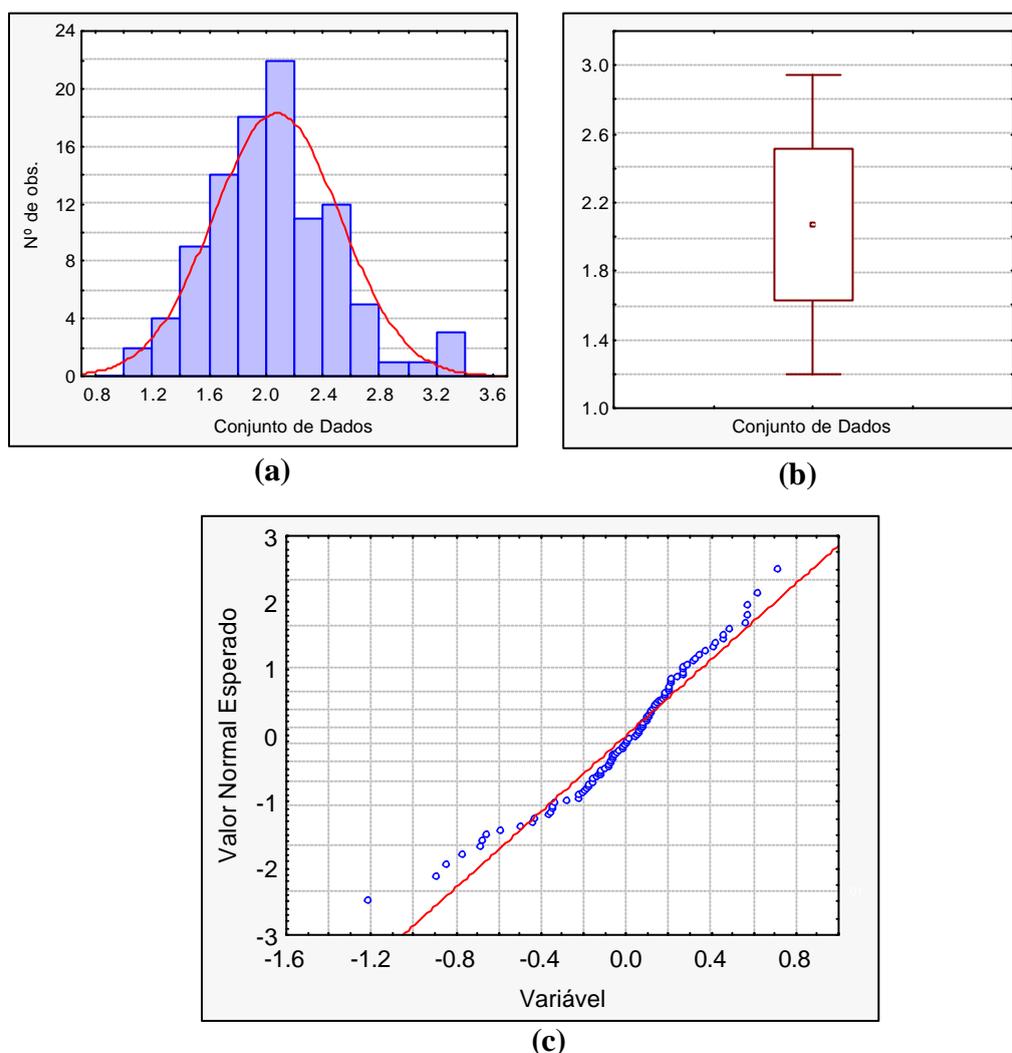


FIGURA 2.8 – Análise da normalidade dos dados: (a) Histograma; (b) Box-plot; (c) Probabilidade normal.

Por fim, a normalidade pode ser testada, também, através do teste de Shapiro-Wilk, elaborado por Shapiro e Wilk (1965). Se a estatística  $W$  for significativa, deve-se rejeitar a hipótese de normalidade dos dados. Comparado com diversos outros testes, o Shapiro-Wilk apresenta propriedades que o habilitam como o mais importante para análise da normalidade de um conjunto de dados (D'Agostino e Stephens, 1986).

### 2.11.2.2 Análise da Variância dos Resíduos

Para estimativa dos parâmetros de uma regressão é utilizado o método dos mínimos quadrados, que calcula o valor dos estimadores que minimizam o quadrado médio dos resíduos. Para garantir tal propriedade é necessário que a variância dos resíduos seja constante (Neter et al., 1996).

Um teste simples para avaliar a estabilidade da variância é o de Levene modificado. Para conduzi-lo deve-se separar o conjunto de dados em dois grupos, ordenados em função do valor da variável independente ( $X$ ). Dessa forma, um grupo deve ser composto por elementos com baixos valores de  $X$ , em oposição ao outro que consiste de elementos com alto valor de  $X$ . O teste de Levine modificado determina se a média dos desvios absolutos de cada grupo são estatisticamente diferentes (Neter et al., 1996).

O Levine modificado baseia-se em um teste de hipóteses, que utiliza a distribuição *t-student*. O valor calculado ( $t_L^*$ ) é definido pela Equação 2.30:

$$t_L^* = \frac{\bar{d}_1 - \bar{d}_2}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.30)$$

Onde  $\bar{d}_1$  e  $\bar{d}_2$  representam a média do desvio das observações, respectivamente, do grupo 1 e 2, em relação a suas medianas;  $s$  é a grandeza definida pela Equação 2.31;  $n_1$  e  $n_2$  são o número de elementos de cada grupo, cuja soma é  $n$ .

$$s = \sqrt{\frac{\sum (d_{i1} - \bar{d}_1)^2 + \sum (d_{i2} - \bar{d}_2)^2}{n - 2}} \quad (2.31)$$

Sendo que  $d_{i1}$  e  $d_{i2}$  denotam o desvio em relação à mediana, do *iésimo* elemento de cada grupo.

Para aceitar a hipótese de variância constante, o valor absoluto de  $t_L^*$  deve ser menor ou igual ao valor assumido pela distribuição *t-student*, com determinado nível de confiança e  $n-2$  graus de liberdade.

Para o caso de ser diagnosticada a heterocedasticidade dos dados, faz-se necessário o uso do método dos mínimos quadrados ponderado, em substituição ao tradicional. A ponderação é feita atribuindo-se às observações pesos iguais ao inverso da variância.

### 2.11.3 Identificação de *Outliers* Influentes

Afora as premissas fundamentais que devem ser respeitadas, por diversas vezes existem observações que se destacam do que a massa de dados sugere. Elas são chamadas de *outliers* e podem influenciar negativamente nos resultados, limitando o desempenho da modelagem estatística. Como forma de identificar os *outliers* influentes, o *DFFITs* mede a diferença entre o valor estimado para determinada observação, considerando todos os elementos do modelo, e sua estimativa ao removê-la da modelagem. A Equação 2.32 formaliza o conceito:

$$DFFITs_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\mathbf{s}_i} \quad (2.32)$$

Onde  $\hat{Y}_i$  representa o valor estimado para a observação  $i$ ;  $\hat{Y}_{i(i)}$  expressa a estimativa de  $i$  ao retirá-la do modelo e  $\mathbf{s}_i$  exprime o desvio padrão estimado para  $\hat{Y}_i$ .

O valor máximo aceitável para o *DFFITs* é de 1, para pequenos ou médios conjuntos de dados, e  $2 \cdot \sqrt{p/n}$ , para grandes volumes de observações, sendo  $p$  o número de parâmetros da regressão e  $n$  o número de observações (Neter et al., 1996).

### 2.11.4 Seleção das Variáveis do Modelo Regressão Múltipla

Quando existe mais de uma variável candidata a explicar a variável dependente, elas precisam ser filtradas de forma a manter apenas as variáveis significativas. Segundo afirmam Neter et al. (1996), não faria sentido a constituição de um modelo estatístico

com variáveis desnecessárias, uma vez que, em número excessivo, elas podem prejudicar a manutenção do modelo. Além disso, segundo os autores, a variância dos valores estimados para a variável dependente ( $\hat{Y}_i$ ) tende a crescer à medida que variáveis não relacionadas a ela são incorporadas.

A primeira providência para seleção de variáveis é conduzir um teste estatístico que defina quais delas são significativas para o modelo. Isto é possível por meio de uma análise simples do *p-valor*, que é a probabilidade de a distribuição *t-student* assumir valor superior ao *t calculado* ( $t^*$ ), definido pelo quociente entre o parâmetro estimado e seu desvio padrão. A Equação 2.33 exhibe o conceito:

$$p - \text{valor} = P\{t(1 - \alpha/2; n - 2) > t^*\} \quad (2.33)$$

Onde  $P$  é a função probabilidade;  $t$  é a distribuição *t-student*;  $\alpha$  é o nível de significância,  $n$  é o número de observações e  $t^*$  é o valor calculado para o teste *t-student* (razão entre o parâmetro estimado e seu desvio padrão).

Para um *p-valor* inferior ao nível de significância, deve-se rejeitar a hipótese de nulidade do parâmetro e, portanto, incorporar, ao modelo, a variável associada.

Os parâmetros considerados nulos devem, ainda, passar por um teste de hipótese para verificar se todos eles podem ser descartados concomitantemente. Esse procedimento é implementado pelo cálculo da Soma dos Quadrados dos Resíduos (SSE) para o modelo com todas as variáveis (completo) e para o modelo com apenas as variáveis significativas (reduzido). O cálculo do valor crítico ( $F^*$ ) para o teste é apresentado na Equação 2.34:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \quad (2.34)$$

Onde  $SSE(R)$  refere-se ao modelo reduzido;  $SSE(F)$  alude ao modelo completo;  $df_R$  e  $df_F$  são o número de graus de liberdade, respectivamente, para o modelo reduzido e completo, sendo  $df_R = n - 1$  e  $df_F = n - 2$ .

O critério para rejeitar a hipótese de nulidade de todos os parâmetros testados é  $F^* > F(1-\alpha; df_R - df_F; df_F)$ , onde  $F$  é a distribuição estatística F (Fisher) e  $\alpha$  o nível de significância.

Concluindo-se que não se podem excluir todas as variáveis do modelo, um procedimento que calcule todas as regressões possíveis, com o conjunto de dados, deve ser organizado. Diferentes critérios para comparação dos modelos de regressão podem ser utilizados, tais como  $C_p$ ,  $R_p^2$ ,  $R_a^2$ .

O  $C_p$  é executado pela montagem de um gráfico de dispersão no qual o eixo das ordenadas é composto pelos valores do indicador, enquanto as abscissas assumem o valor do número de variáveis do subconjunto avaliado ( $p$ ). Os pares ordenados localizados sobre a reta  $C_p = p$  são indicados como melhores subconjuntos de variáveis.

Os critérios  $R_p^2$  e  $R_a^2$  também se utilizam do gráfico de dispersão, no entanto a escolha do subconjunto se baseia nos pares ordenados que representam pontos de máximo valor da função.

### **2.11.5 Diagnóstico de Multicolinearidade**

Quando se observa a existência de correlação entre as diversas variáveis independentes, sobre as quais se estão desenvolvendo os estudos, diz-se que há multicolinearidade entre elas.

Assim como a falta de normalidade dos resíduos e a heterocedasticidade, a multicolinearidade não afeta de maneira decisiva estimativas pontuais. Seu efeito está relacionado com os parâmetros, que se tornam instáveis, comprometendo a interpretabilidade dos coeficientes e seus respectivos intervalos de confiança (Neter et al., 1996).

Uma forma de medida da multicolinearidade é pelo Fator de Inflação da Variância (do inglês, *Variance Inflation Factor* – VIF), que é obtido por intermédio da matriz de correlação das variáveis. Neter et al. (1996) ponderam que um valor crítico para a

grandeza não deve superar 10 unidades para cada variável. Observam, no entanto, que a média aritmética do índice superior a 1, considerando todas as variáveis, é indicativa de influência da multicolinearidade na estimativa dos parâmetros.

### 2.11.6 Validação do Modelo de Regressão

Por fim, faz-se necessário testar, em uma análise de regressão, a validade do modelo gerado. Dessa forma, deve-se separar, a priori, um subconjunto das observações, podendo-se realizar dois estudos simples, que comparam o modelo estatístico estabelecido com um composto pela amostra de validação.

O primeiro teste é o cálculo do Erro Quadrático Médio de Predição (*MSPR*), definido pela Equação 2.35 e que é uma medida da diferença entre o valor observado e o calculado com a equação de regressão, para o conjunto de validação. Sendo  $\hat{Y}_i$  o valor predito (estimado) para a *i*ésima observação do conjunto de validação,  $Y_i$  o valor correspondente observado da variável dependente e  $n^*$  o número de observações do subconjunto de validação.

$$MSPR = \frac{\sum_{i=1}^{n^*} (Y_i - \hat{Y}_i)^2}{n^*} \quad (2.35)$$

Um valor de *MSPR* que atesta a validade do modelo gerado é, segundo Neter et al. (1996), aquele que se aproxima do Quadrado Médio dos Resíduos (*MSE*).

O segundo estudo, que segue a mesma linha do primeiro, será denominado neste trabalho como Ajuste à Reta Diagonal (ARD). Ele avalia, para o subconjunto de validação, se a projeção dos pares ordenados compostos pela variável dependente, calculada pelo modelo, e seu valor observado recaem sobre uma reta de coeficiente angular igual à unidade (inclinação de 45°).

Para operacionalizar o ARD são feitos dois testes de hipóteses simples (testes estatísticos *t-student*). Assim, se o modelo fosse perfeito, a regressão  $Y_i \times \hat{Y}_i$  deveria

resultar em um  $R^2=1$  e sua reta de tendência apresentar inclinação de  $45^\circ$ . Em outras palavras, espera-se que o intercepto da regressão supracitada seja estatisticamente nulo e o coeficiente de inclinação igual à unidade. Estas injunções se transformam nas hipóteses nulas, que serão aceitas se o valor do teste estatístico ( $t^*$ ) for menor que o valor tabelado para a distribuição *t-student*.

Para um aprofundamento mais específico na teoria de análise multivariada, sugere-se o estudo de Neter et al. (1996), que faz uma abordagem ampla e aplicada do tema.

## CAPÍTULO 3

### ÁREA DE ESTUDO

A área de estudo está localizada no norte do Estado de São Paulo e corresponde aos municípios de Ipuã, Guará e São Joaquim da Barra. A Figura 3.1 apresenta graficamente a localização da área.

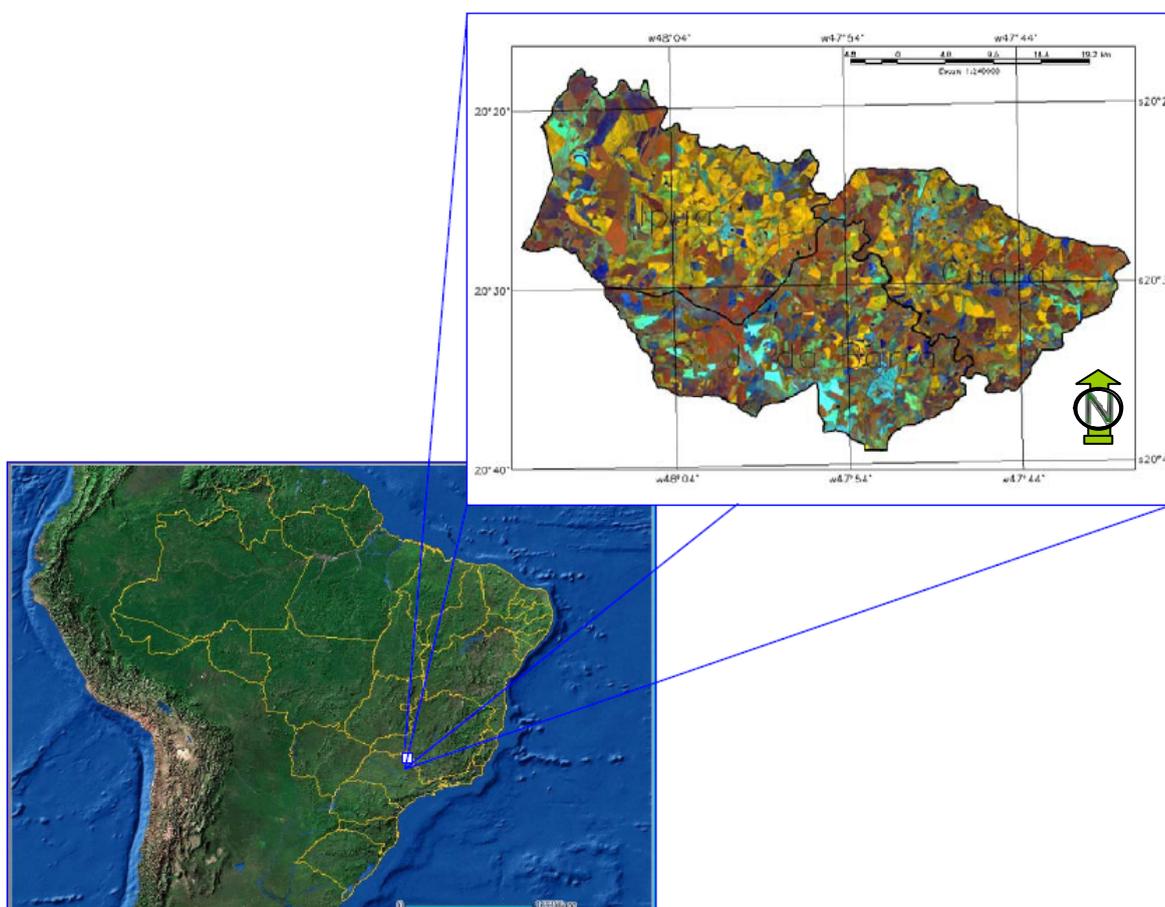


FIGURA 3.1 – Localização da área de estudo.

FONTE: Adaptada de Sanches (2004, p. 43).

A cidade de Ipuã apresenta uma população estimada, para 2004, de 12.647 habitantes, cobrindo uma área de 46.600 ha. São Joaquim da Barra, por sua vez, tem 44.472 habitantes em uma dimensão de 41.200 ha. Por fim, Guará é o menor município, com 36.300 ha e população de 20.226 pessoas (IBGE, 2005).

A Figura 3.2 destaca numa imagem do sensor ETM+/Landsat-7, composição colorida R5G4B3, a localização dos municípios de interesse na área de estudo, podendo-se observar seus limites em traçado contínuo preto, bem como, polígonos com preenchimento branco, que representam as respectivas manchas urbanas.

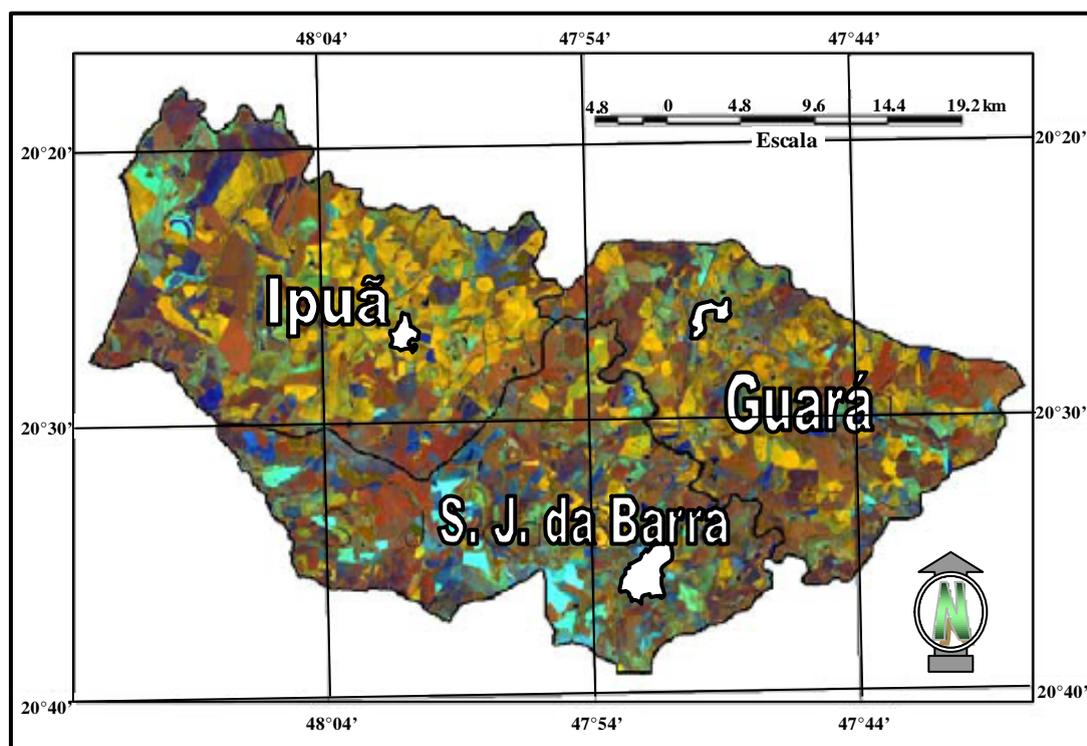


FIGURA 3.2 – Localização dos municípios na área de estudo.

A economia da região é eminentemente agrícola, com predomínio de solos férteis (Latosolos Vermelhos), com relevo suave ondulado a plano. O clima é classificado como Cw. A agricultura é tecnicada, havendo várias áreas com pivô central. As principais culturas são a soja, o milho e a cana-de-açúcar. Praticam-se comumente dois ciclos agrícolas por ano, com milho e sorgo após o período normal de safra. Observa-se, ainda, a ocorrência de rotação entre cana-de-açúcar e uma cultura anual por ocasião da reforma do canavial em alguns talhões (Epiphanyo et al., 2001).

A escolha da área de estudo levou em consideração o fato de mais de 75% da região dos três municípios serem cultivadas, ao menos uma vez ao ano, com as grandes culturas de interesse, quais sejam, cana-de-açúcar, milho e soja (Luiz et al., 2002). As diversas formas de uso do solo presentes foram fundamentais para possibilitar as avaliações propostas neste trabalho.

A Figura 3.3 confirma as afirmações acerca da tendência agrícola da região. Com classificação elaborada por Gürtler (2003), o mapa de uso e ocupação do solo mostra a grande hegemonia da classe área agrícola. A Tabela 3.1, por sua vez, sintetiza as informações em números, para o ano de 2002, explicitando que mais de 90% do cultivo é composto pelo grupo milho, cana e soja.

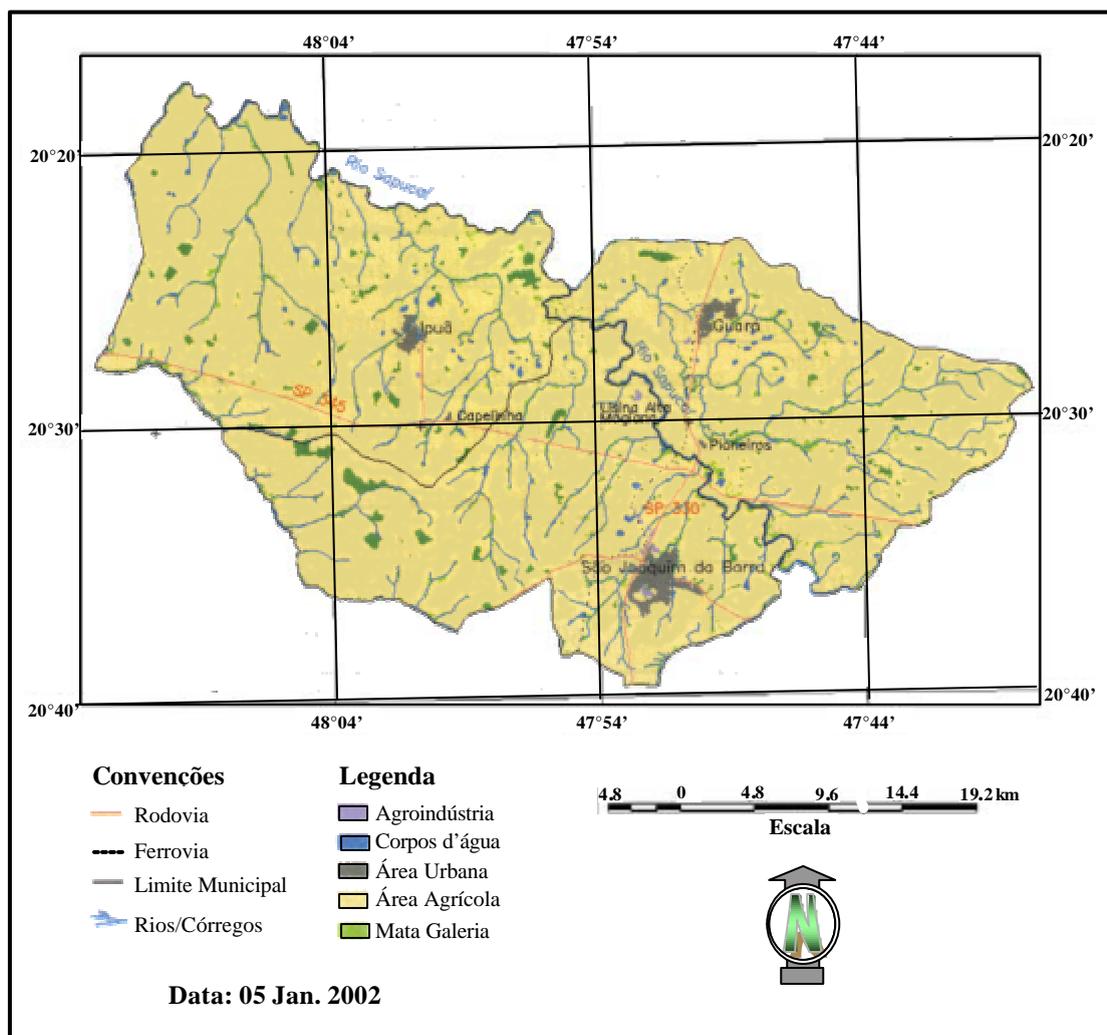


FIGURA 3.3 – Mapa de uso e ocupação do solo da área de estudo.

TABELA 3.1 – Área cultivada, em hectares, por cultura, para o ano de 2002.

	Guará		S. J. Barra		Ipuã	
	Área	%	Área	%	Área	%
<b>Cana-de-açúcar</b>	16.501	65,14	24.000	81,73	25.000	55,11
<b>Milho (em grão)</b>	3.000	11,84	1.244	4,24	4.500	9,92
<b>Soja (em grão)</b>	4.442	17,54	3.800	12,94	13.000	28,66
<b>Soma das 3 culturas</b>	23.943	94,52	29.044	98,91	42.500	93,69
<b>Total de área plantada</b>	25.331	100	29.364	100	45.360	100

FONTE: IBGE (2006).

Fazendo conexão com os objetivos do trabalho, a Tabela 3.2 apresenta uma estatística descritiva das dimensões das Unidades de Produção Agropecuária (UPA) das culturas presentes na região de interesse. Observa-se que a soja e a cana-de-açúcar atingem unidades maiores, da ordem de até 2000ha e 1400ha, respectivamente. O milho, entretanto, é composto por menores UPA, alcançando área máxima de 532ha.

TABELA 3.2 – Tamanho mínimo, médio e máximo das UPA, em hectares.

Cultura	Guará			S. J. Barra			Ipuã		
	Mín.	Méd.	Máx.	Mín.	Méd.	Máx.	Mín.	Méd.	Máx.
<b>Soja</b>	2	100	765	1	202	2150	4	122	1244
<b>Cana-de-açúcar</b>	4	122	561	1	70	1000	0	237	1433
<b>Milho</b>	1	46	249	2	40	245	1	62	532

FONTE: CATI (2006).

## CAPÍTULO 4

### MATERIAIS E MÉTODO

A estrutura do trabalho realizado pode ser dividida em duas partes bem definidas. Num primeiro momento, buscou-se, com processos de filtragem espacial, imitar o dado de resolução espacial moderada, de forma a estudar a viabilidade do uso de imagens simuladas em substituição a sensores de resolução moderada. Após isso, através do uso das imagens MODIS e ETM+, buscou-se determinar a relação existente entre a baixa e a alta resoluções, no tocante à estimativa de áreas agrícolas. Nesta fase, os índices espaciais foram utilizados para caracterização do padrão espacial das culturas, bem como incorporados a um modelo estatístico para refinar a estimativa das áreas.

Para implementação de todas as tarefas, aplicaram-se diversas ferramentas e produtos, os quais passam a ser discutidos neste capítulo.

#### 4.1 Materiais

##### 4.1.1 Dados de Trabalho

Foi utilizado um banco de dados previamente organizado por Sanches (2004), cuja modelagem geográfica encontra-se apresentada na Tabela 4.1.

TABELA 4.1 – Modelagem do banco de dados geográfico.

<b>Banco de Dados Geográfico</b>	
<b>Projeção</b>	UTM
<b>Zona</b>	23
<b>Datum</b>	SAD-69
<b>Retângulo envolvente</b>	Canto inferior esquerdo: X1,Y1 = (162000, 7710000) Canto superior direito: X2,Y2 = (228000, 7755000)

Para a consecução do trabalho, quanto ao dado de resolução fina, foi utilizada a imagem do sensor ETM+/Landsat 7, de órbita/ponto 220/74 e data de 05 de janeiro de 2002. O produto foi corrigido quanto aos efeitos atmosféricos pela técnica desenvolvida por Chavez (1988).

Quanto à imagem de resolução moderada, utilizou-se o produto MOD09 (reflectância de superfície) do sensor MODIS/Terra, de 05 de janeiro de 2002, da área do retângulo envolvente apresentado na Tabela 4.1.

Somente as bandas 1 e 2 (vermelho e infravermelho próximo, respectivamente) do sensor MODIS, com resolução espacial de 250m, foram utilizadas na análise dos dados. Para garantir uma equivalência espectral aproximada entre os sensores MODIS e ETM+/Landsat-7, apenas as bandas 3 e 4 do sensor ETM+ foram consideradas na análise.

De forma a atingir uma equivalência das classificações temáticas com a realidade, utilizou-se um mapa de referência confeccionado por Luiz (2003), elaborado a partir de visita a campo realizada no mês de janeiro de 2002 e adaptado na Figura 4.1. A existência deste produto foi fator considerado na escolha da data das imagens que foram utilizadas nesta pesquisa.

#### **4.1.2 Recursos de Software**

Tendo em vista a especificidade das tarefas executadas para atingir os objetivos estabelecidos, foi necessário o uso de diversos programas, conforme é apresentado a seguir.

Buscando a adequação das imagens para uma visualização global, as imagens MODIS são disponibilizadas na projeção cartográfica sinusoidal, cujos parâmetros não são definidos nos principais programas de processamento digital de imagens. Portanto, fez-se necessário o uso do módulo MODIS *Reprojection Tool* (MRT), em sua versão 3.2a. Esse programa permite a leitura dos arquivos da imagem e subsequente transformação para os principais sistemas de projeção utilizados para representação de dados espaciais.

Para estudo pormenorizado das potencialidades do programa, recomenda-se USGS (2004).

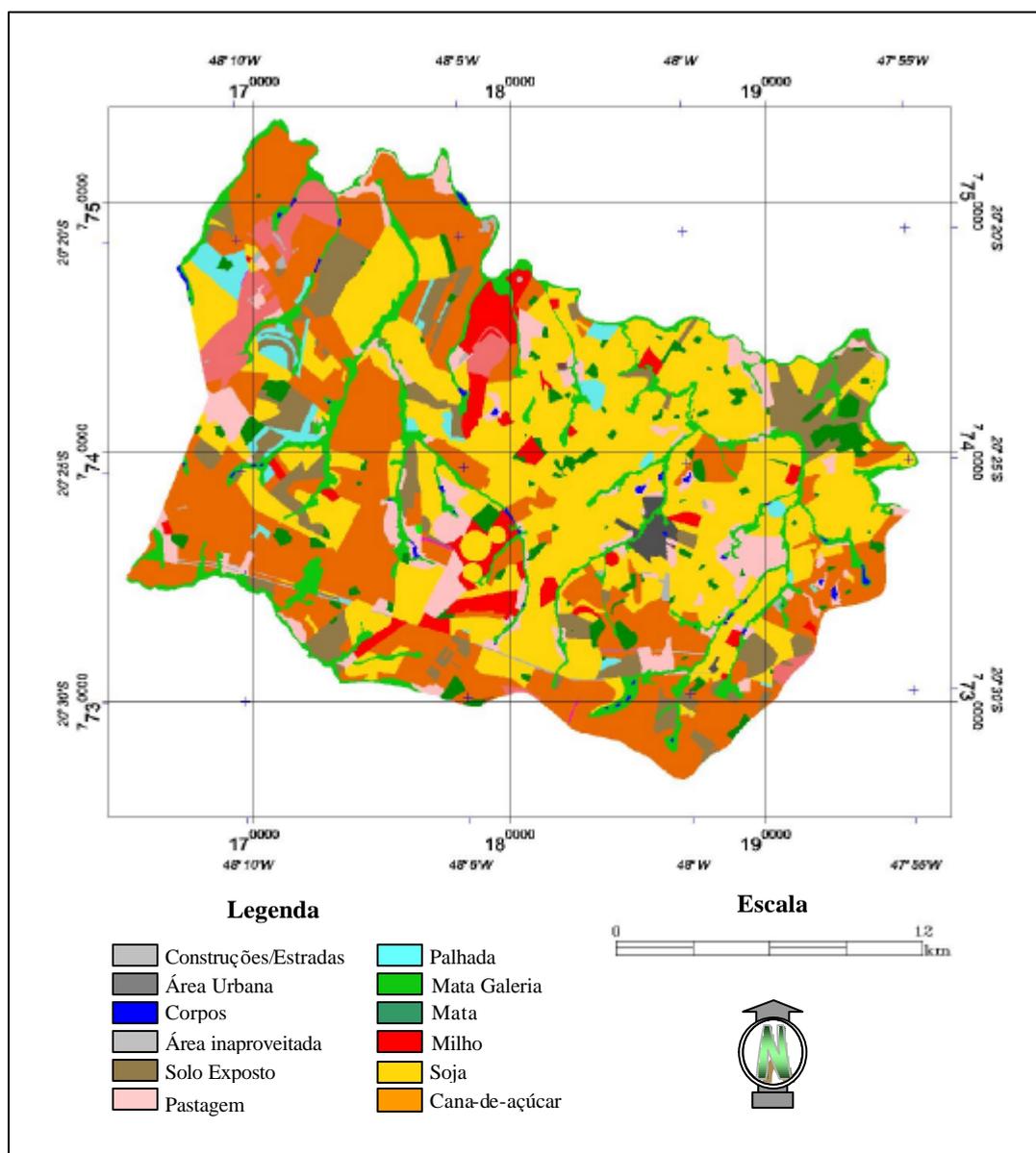


FIGURA 4.1 – Mapa de referência para classificação temática.

A manipulação das informações geográficas, bem como organização das informações do banco de dados foi efetuada no Sistema para Processamento de Informações Georreferenciadas (SPRING), versão 4.1. Este programa foi desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE) e é capaz de processar informações matriciais, vetoriais e dados literais integrando-os numa base de dados, permitindo realização de consultas e integração de dados de diferentes fontes e naturezas.

Muitos processos exigiram, ainda, a utilização de técnicas específicas de processamento digital de imagens, o que determinou a escolha de um programa desta natureza. Assim, o *Environment for Visualizing Images* (ENVI), versão 4.0, desenvolvido pela *Research Systems Inc.* (RSI), apresentou-se como solução viável por ser dotado de ferramentas implementadas capazes de otimizar sobremaneira as operações demandadas.

O *Interactive Data Language* (IDL) é uma linguagem de programação ideal para a implementação de algoritmos que operam com imagens, uma vez que ele é orientado para matrizes, além de possuir uma extensa biblioteca matemática. Baseando-se em tais potencialidades, sua versão 6.0 foi explorada no escopo metodológico desta pesquisa.

Um objetivo claro definido para este trabalho foi a caracterização do padrão espacial das culturas existentes na área de estudo. Para atingi-lo, foi empregado o *software* de domínio público *Fragstats*, versão 3.3, que efetua análises de padrão espacial com cálculo de diversas métricas de paisagem.

Por fim, fez-se mandatória a estruturação das informações de forma a permitir a execução de análises estatísticas que dessem sustentação teórica à dissertação. Dessa forma, tarefas menos complexas como coleção de tabelas, seleção amostral, diagramas de dispersão foram realizadas no *Excel 2003*, da *Microsoft*.

Um modelo rigoroso exige, todavia, estudos mais aprofundados, o que determinou a utilização do *Statistica 6.0*, elaborado pela *StatSoft Inc.*. Este programa foi decisivo para efetuar os cálculos e transformação que compõem a metodologia desta obra.

A TABELA 4.2 apresenta de maneira concisa os principais recursos de *softwares* que foram empregados neste trabalho, trazendo uma breve descrição de suas potencialidades e uso.

TABELA 4.2 – Recursos de *software* empregados.

<b>Programa</b>	<b>Uso</b>	<b>Potencialidade</b>
MRT	Transformação de sistema de projeção.	Manipulação simples; indispensável para utilização das imagens MODIS.
SPRING	Aquisição dos dados.	Integração de informações de natureza distinta; SIG.
ENVI	Processamento digital de imagens.	Diversas ferramentas implementadas.
IDL	Implementação de rotinas.	Linguagem de programação orientada a matrizes.
Fragstats	Cálculo de métricas.	Extração rápida e simples de índices de padrão espacial.
Excel	Análise estatística.	Permite fácil geração de medidas estatísticas básicas.
Statistica	Análise estatística.	Pacote completo para profunda abordagem estatística

## 4.2 Método

### 4.2.1 Simulação de Resolução Espacial por Filtragem

A configuração de um sensor para recursos terrestres passa obrigatoriamente pela seleção da resolução espacial. Assim, simulações realísticas são ferramentas poderosas no processo de avaliação da melhor resolução espacial para uma aplicação específica. Nesta linha, a adequação da simulação de resolução espacial foi estudada neste trabalho, pela aplicação de dois métodos de filtragem: o de maioria e o de textura.

No presente trabalho, a simulação de diferentes níveis de resolução espacial, pela filtragem espacial de dados ETM+/Landsat-7, teve os seguintes objetivos: (a) verificar o comportamento das relações de estimativas de áreas de cada cultura agrícola em função da resolução espacial de 30m e das resoluções simuladas de 90m, 150m, 210m e 270m; (b) avaliar o comportamento das métricas da paisagem em função da resolução espacial e da cultura agrícola sob análise; (c) possibilitar a pré-seleção de métricas potencialmente úteis para compor um modelo de regressão múltipla, visando à predição de área de culturas agrícolas em dados de resolução espacial de 30m a partir de dados de resolução espacial moderada (MODIS/Terra, 250m).

#### 4.2.1.1 Filtragem de Maioria

A Figura 4.2 apresenta o esquema geral dessa operação. Sobre as bandas 3 e 4 do sensor ETM+/Landsat-7, foi executada uma classificação digital não-supervisionada *k-means* com 15 classes e 20 iterações, como propõe Frohn e Hao (2006). Sobre os resultados de classificação, foram aplicados filtros de maioria nas dimensões 3×3 (90m), 5×5 (150m), 7×7 (210m) e 9×9 (270m) *pixels*, o que permitiu obter uma degradação gradual do dado original (30m).

Como forma de medição da qualidade das classificações, foi calculado o índice kappa para cada uma delas, utilizando o mapa de referência de Luiz (2003) (Figura 4.1).

#### 4.2.1.2 Filtragem de Textura

Na mesma linha de raciocínio apresentada para o processo de simulação por análise de maioria, foram utilizados filtros de dimensões 3×3, 5×5, 7×7 e 9×9 *pixels* na filtragem por textura (Figura 4.3).

Apesar das semelhanças em termos de generalização espacial, existem algumas diferenças significativas entre filtragem de textura e de maioria. Primeiramente, conforme se pode constatar na análise comparativa das Figuras 4.2 e 4.3, a filtragem de maioria é aplicada sobre a imagem classificada, enquanto na análise de textura o processo é realizado sobre a imagem original, dessa forma, gerando um novo dado que será classificado. Além disso, a filtragem de textura degrada a partir da informação trazida por uma característica física da imagem, sem, contudo, alterar a resolução espacial; já a análise de maioria opera diretamente sobre os valores numéricos da imagem. Apesar de não modificar o tamanho do *pixel*, de fato há mudança na resolução espacial a cada agregação operada pela janela móvel (Frohn e Hao, 2006).

Como forma de medição da qualidade das classificações geradas após a filtragem de textura, foi calculado o índice kappa para cada uma delas, utilizando o mapa de referência de Luiz (2003) (Figura 4.1).

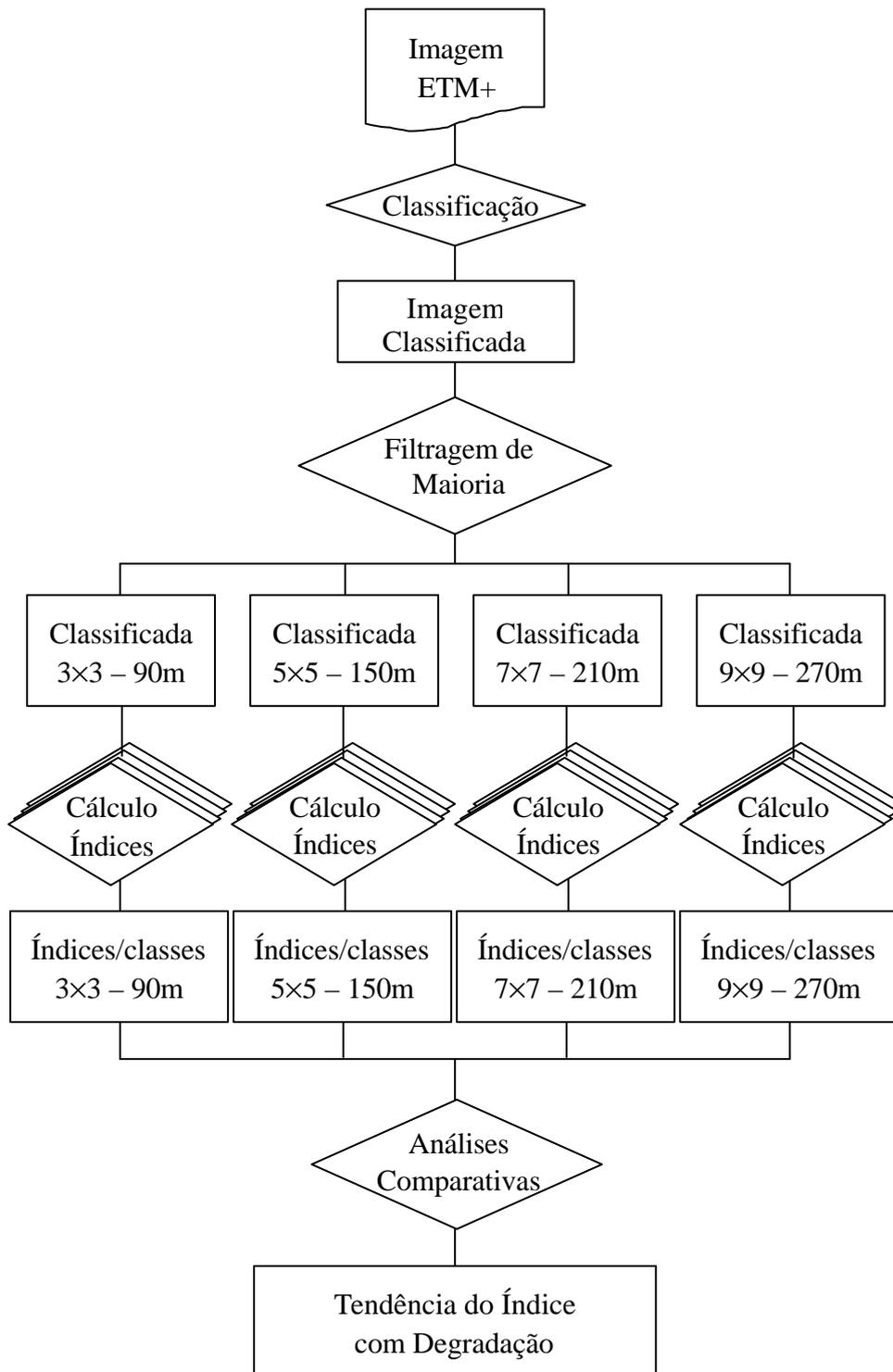


FIGURA 4.2 – Análise de simulações através do uso de filtragens de maioria.

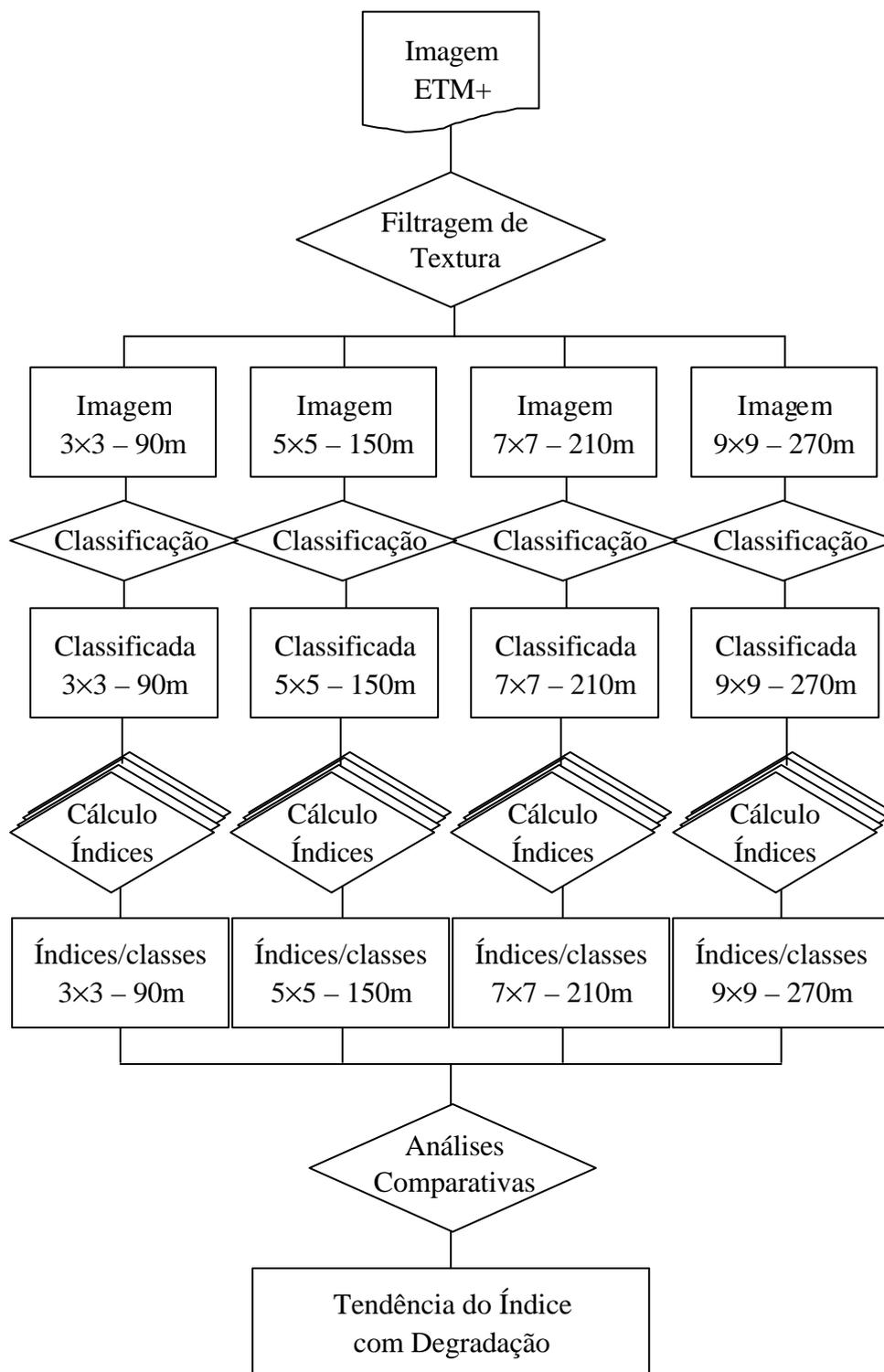


FIGURA 4.3 – Análise de simulações através do uso de filtragens de textura.

## 4.2.2 Cálculo e Evolução das Métricas da Paisagem em Função da Resolução Espacial

Após a realização das filtragens de maioria (Figura 4.2) e textura (Figura 4.3), procedeu-se ao cálculo dos 20 índices (métricas da paisagem) relacionados na Tabela 4.3. As equações utilizadas no cálculo destes índices foram discutidas no Capítulo 2 (Fundamentação Teórica).

Para cada uma das imagens filtradas resultantes, o cálculo das 20 métricas de paisagem possibilitou a caracterização do padrão espacial das classes agrícolas de interesse. Para a variável Índice de Área Nuclear (*Core*), considerou-se o isolamento da vizinhança ao se reservar uma faixa de 2 *pixels* da borda. Tal critério levou em consideração apenas fatores de formação da imagem, não havendo uma abordagem voltada à dinâmica da paisagem. Afinal, o que se objetiva é uma análise comparativa entre imagens de resoluções espaciais distintas, devendo-se, assim, manter as análises tangenciando o conceito inserido no elemento de resolução.

A confecção de gráficos permitiu a análise da sensibilidade dos índices à variação do tamanho da janela de filtragem. Sobretudo a área, que é o principal foco desta pesquisa, foi examinada, ainda, com o uso de histogramas, bem como de regressões simples, nas quais a variável independente foi a área obtida através do sensor ETM+/Landsat-7 (dado original de 30m). A variável dependente foi a área obtida a partir de cada uma das imagens simuladas (imagens classificadas de 90m, 150m, 210m e 270m de resolução espacial).

Portanto, todos os estudos realizados sobre os resultados da filtragem de maioria também foram procedidos para a análise de textura. O objetivo foi garantir a possibilidade de uma profunda comparação entre os métodos e assim permitir a escolha do método mais realístico de simulação de resolução espacial moderada (MODIS/Terra).

TABELA 4.3 – Métricas de paisagem utilizadas no trabalho.

<b>Métrica</b>	<b>Variável</b>	<b>Categoria</b>	<b>Medida</b>
Área	Área	Área/Densidade/Borda	Composição do cenário
Número de polígonos	NP		Subdivisão da classe
Porcentagem de cenário	PLAND		Composição proporcional do cenário
Densidade de área	PD		Subdivisão proporcional da classe
Maior área	LPI		Domínio da classe
Forma de classe	LSI		Agregação
Forma de classe normalizada	NLSI		Agregação (normalizada)
Área nuclear total	TCA	Área nuclear	Composição do cenário
Porcentagem nuclear de cenário	CPLAND		Composição proporcional do cenário
Número de áreas nucleares disjuntas	NDCA		Vizinhança espacial de área nuclear
Densidade de área nuclear disjunta	DCAD		Subdivisão proporcional da classe
Porcentagem de adjacência	PLADJ	Contágio/Espalhamento	Agregação
Agregação	AI		Agregação
Interposição e Justaposição	IJI		Mistura de classes
Agrupamento	CLUMP		Auto-adjacência
Divisão de cenário	DIV		Heterogeneidade
Fragmentação	SPLIT		Fragmentação
Tamanho de trama efetiva	MESH		Homogeneidade
Dimensão fractal perímetro-área	PAFRAC	Forma	Complexidade da forma
Coesão de área	COHE	Conectividade	Conectividade

#### **4.2.3 Determinação da Técnica de Simulação (Filtragem) mais Realística e Pré-Seleção de Métricas para o Modelo de Regressão Múltipla**

Para tal finalidade, foi desenvolvida uma rotina em IDL capaz de recortar uma imagem em parcelas iguais. Para este trabalho foram gerados 216 subconjuntos (blocos) de dimensão  $125 \times 125$  *pixels* para a imagem simulada e  $15 \times 15$  *pixels* para a imagem MODIS. Este processo permitiu a realização de um estudo estatístico robusto na comparação entre os dados originais do sensor MODIS/Terra (250m de resolução espacial) e os resultantes da simulação de 270m ( $9 \times 9$  *pixels*) nas filtragens de maioria e textura. Procedeu-se, assim, uma análise de regressão entre áreas (MODIS original versus MODIS simulado), por cultura, utilizando cada um dos 216 recortes (blocos) de ambas as imagens, para a determinação da técnica de filtragem mais adequada para fins de simulação de resolução espacial.

Com base nesta indicação, procedeu-se a pré-seleção de métricas potencialmente úteis para compor o modelo de regressão múltipla, visando à predição de área de culturas agrícolas em dados de resolução espacial fina (p.ex., ETM+/Landsat-7, 30m) a partir de dados de resolução espacial moderada (p.ex., MODIS/Terra, 250m). Segundo Frohn (1998), uma métrica de paisagem efetiva é aquela que é insensível ou previsivelmente sensível à amostragem do dado, além de dever ser influenciada pelo padrão espacial associado. Com essa visão, as análises das simulações foram utilizadas para eliminar os índices ou métricas que se mostraram inadequados para caracterizar o padrão espacial das culturas agrícolas, ou apresentaram forte correlação.

#### **4.2.4 Modelagem Estatística para Predição de Área: Obtenção de Imagem de 30m a partir de Sensor de 250m de Resolução Espacial**

Foi confeccionado um modelo estatístico capaz de, a partir da imagem de resolução espacial de 250m, predizer a área de classes temáticas extraídas com o uso de um sensor de 30m.

Os modelos de regressão simples e múltipla, descritos a seguir, foram precedidos por uma análise exploratória dos dados, envolvendo a análise da normalidade da variável

dependente (p.ex., teste de normalidade Shapiro-Wilk) e a necessidade de eventuais transformações matemáticas das variáveis preditivas.

#### **4.2.4.1 Modelo de Regressão Simples com Dados Originais ( $\text{ÁREA}_{\text{ETM+}}$ versus $\text{ÁREA}_{\text{MODIS}}$ )**

A Figura 4.4 apresenta o esquema geral dessa operação, que teve como objetivo avaliar o desempenho de um modelo de predição de área baseado somente em medidas de área, sem considerar, portanto, o padrão de distribuição espacial das diferentes culturas (métricas da paisagem). A idéia final foi comparar o desempenho deste modelo com o de regressão múltipla, verificando o ganho da incorporação de métricas da paisagem na predição de áreas de sensores de resolução espacial fina a partir de dados de resolução espacial moderada.

Conforme já mencionado, foi desenvolvida uma rotina em IDL capaz de recortar uma imagem em parcelas iguais, sendo gerados 216 subconjuntos (blocos) de dimensão  $125 \times 125$  *pixels* para a imagem ETM+, e  $15 \times 15$  *pixels* para a imagem MODIS. Assim, com o recorte das imagens classificadas, originou-se um total de 1296 observações, uma vez que foram consideradas seis classes: cana-de-açúcar, soja, milho, solo exposto, mata e pastagem. Do quantitativo total, foram separadas 240 observações (40 para cada classe) para validação da equação de regressão simples.

Como forma de análise inicial dos dados, as áreas das classes extraídas entraram, então, num procedimento de regressão, para determinar a relação existente entre as medidas obtidas nas imagens de resolução espacial fina (ETM+/Landsat-7) e moderada (MODIS/Terra).

O coeficiente de explicação obtido da regressão simples foi um parâmetro de comparação com o modelo de regressão múltipla descrito a seguir.

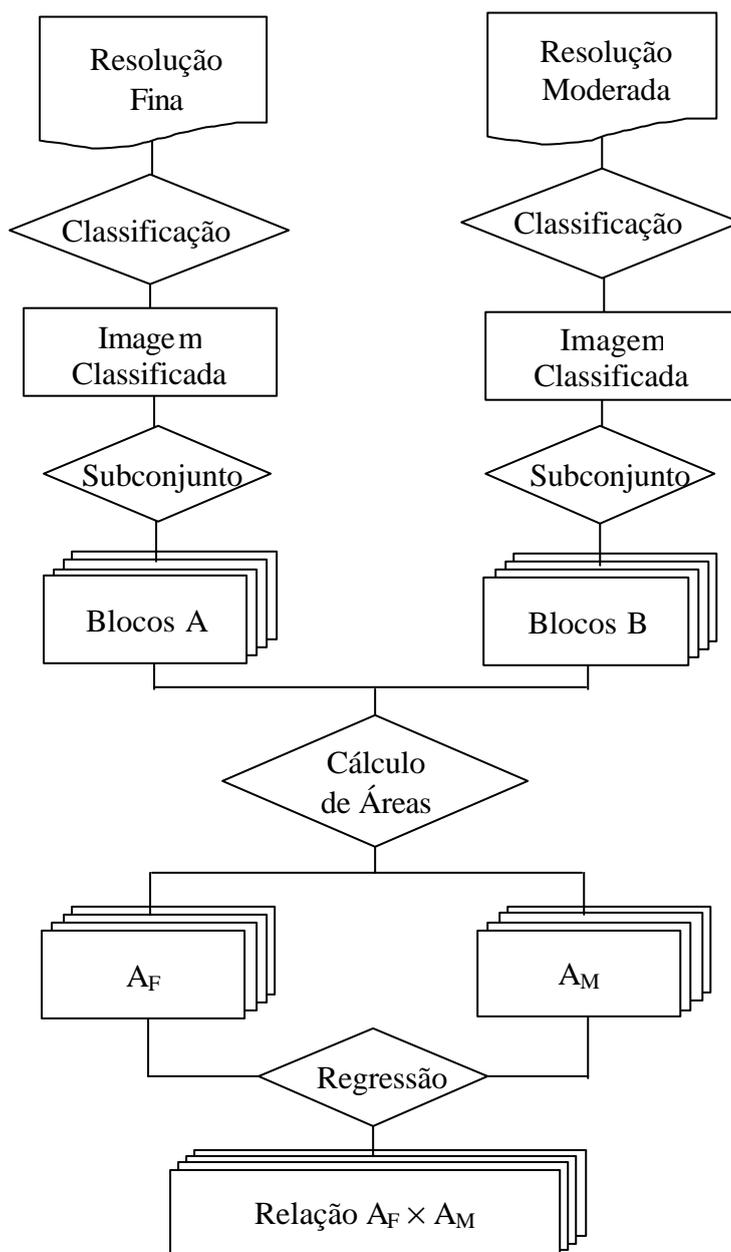


FIGURA 4.4 – Teste da relação  $A_F=f(A_M)$ .

#### 4.2.4.2 Modelo de Regressão Múltipla com Dados Originais e Métricas de Paisagem (ÁREA<sub>ETM+</sub> versus ÁREA<sub>MODIS</sub> + Métricas)

A importância da teoria da regressão linear está na possibilidade de sua aplicação em diversas áreas de conhecimento. Moody e Woodcock (1995) afirmam que a análise de regressão pode ser aplicada para investigar a influência do padrão espacial de culturas na estimativa de área com sensores de resolução espacial moderada. Para tanto, faz-se

necessária a criação de um modelo que relacione a estimativa real (variável dependente), com os diversos padrões que se julgue necessários, além da estimativa extraída da imagem estudada (variáveis independentes).

Dessa forma, foi desenvolvido um modelo de regressão múltipla para estimativa da área de classes temáticas obtidas com sensor de 30m de resolução espacial (ETM+/Landsat-7), a partir de dados de área medidos com imagem de 250m (MODIS/Terra), considerando-se o uso potencial das métricas de paisagem pré-selecionadas na etapa de filtragem, denotando o padrão de distribuição espacial das culturas agrícolas na cena MODIS. A Figura 4.5 apresenta o fluxograma de trabalho adotado nesta fase.

Semelhante ao procedido para a regressão simples foram inseridas na modelagem 1056 observações, deixando-se separadas as 240 restantes para validação do modelo. As métricas de paisagem pré-selecionadas na etapa de filtragem foram consideradas para fins de inclusão/exclusão do modelo de regressão múltipla (seleção dos índices na Figura 4.5) de acordo com os critérios  $C_p$ ,  $R_p^2$ ,  $R_a^2$ , discutidos em detalhe no Capítulo 2 (Fundamentação Teórica).

#### **4.2.5 Comparação dos Modelos de Regressão Simples e Múltipla: Abordagens Geral (Todas as Culturas) e Específica (por Cultura)**

A elaboração de modelos de regressão simples e múltipla teve por objetivo avaliar o benefício causado pela adição do padrão espacial, para estimativa de área de classes temáticas com sensores de resolução moderada.

Adicionalmente, foi feita uma análise comparativa entre modelos estatísticos multivariados estabelecidos, especificamente, para cada cultura agrícola e aquele obtido para classes temáticas genéricas, descrito na seção 4.2.4.2. Tal procedimento teve por fim a discussão sobre o procedimento mais adequado para se obter área de culturas agrícolas com sensores de resolução moderada.

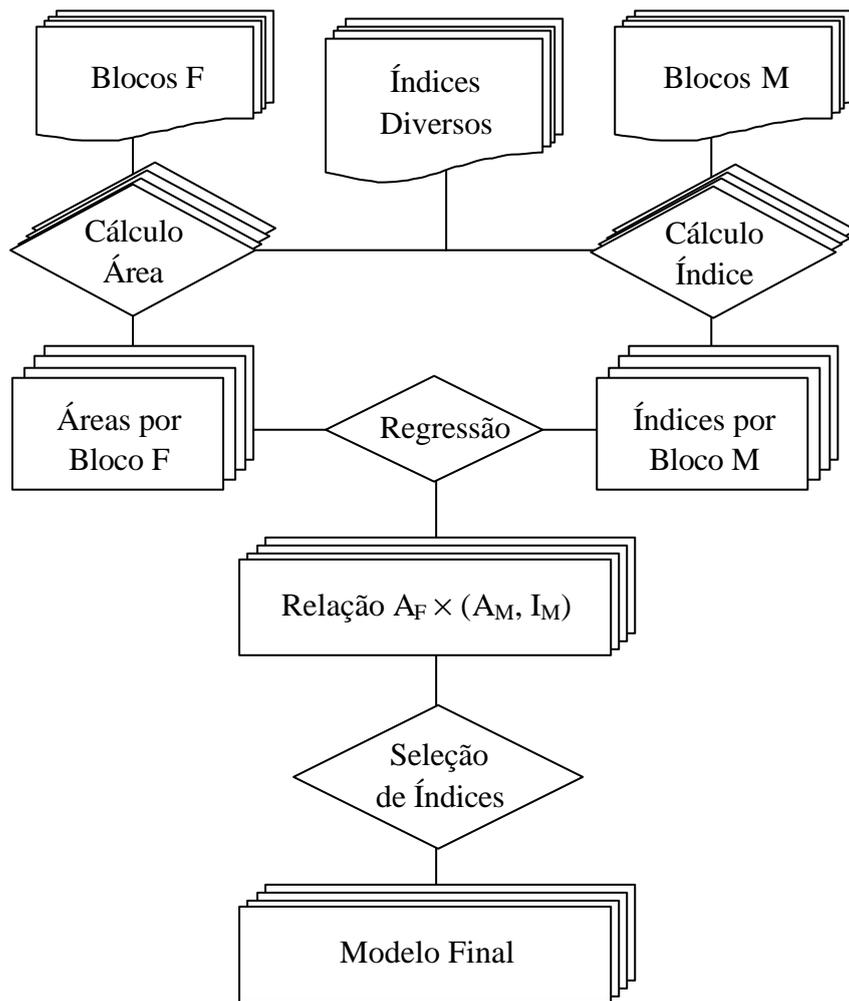


FIGURA 4.5 – Seleção dos índices espaciais.



## CAPÍTULO 5

### RESULTADOS E DISCUSSÃO

#### 5.1 Análise das Classificações dos Produtos de Resolução Espacial Simulada por Filtragem de Textura e Maioria

A Figura 5.1 apresenta os resultados da classificação não-supervisionada (K-Médias) das imagens geradas pela variação do tamanho da janela de filtragem de textura. Nesta figura, também estão indicados os valores do índice Kappa, calculado para medir a qualidade da classificação de cada imagem em relação ao mapa de referência de Luiz (2003), mostrado na Figura 4.1.

Na Figura 5.1, pode-se observar nitidamente que o objetivo da aplicação do filtro foi perfeitamente atingido, visto que há uma redução da complexidade do cenário com a degradação da resolução espacial de 30m para 270m. Muitos polígonos agregaram classes vizinhas para se tornarem mais compactos, enquanto outros gradativamente vão reduzindo-se até, em alguns casos, desaparecerem. Áreas que se apresentam como faixas estreitas, que trazem, portanto, informação de fragmentação mais forte, são exemplos de agregações ocorridas.

Faz-se importante notar, também, a ocorrência isolada de mudanças na própria classificação de alguns polígonos, conforme ressaltado pelas setas na Figura 5.1 (d, e). Tal evento poderá implicar em descontinuidade de alguma tendência, quando se estiver efetuando análises da evolução, através das resoluções, da área de classes específicas.

A Figura 5.2 mostra, por sua vez, a evolução do índice Kappa com a filtragem de textura, que denota sua tendência de queda com a degradação da resolução espacial. Na resolução espacial de 30m o Kappa indicou, de acordo com a classificação de Landis e Koch (1977), uma excelente concordância entre os dados de Luiz (2003) e os resultados da classificação não-supervisionada, convergindo para uma concordância moderada à medida que se degrada mais a resolução e atingindo o limite de 0,65 para 270m.

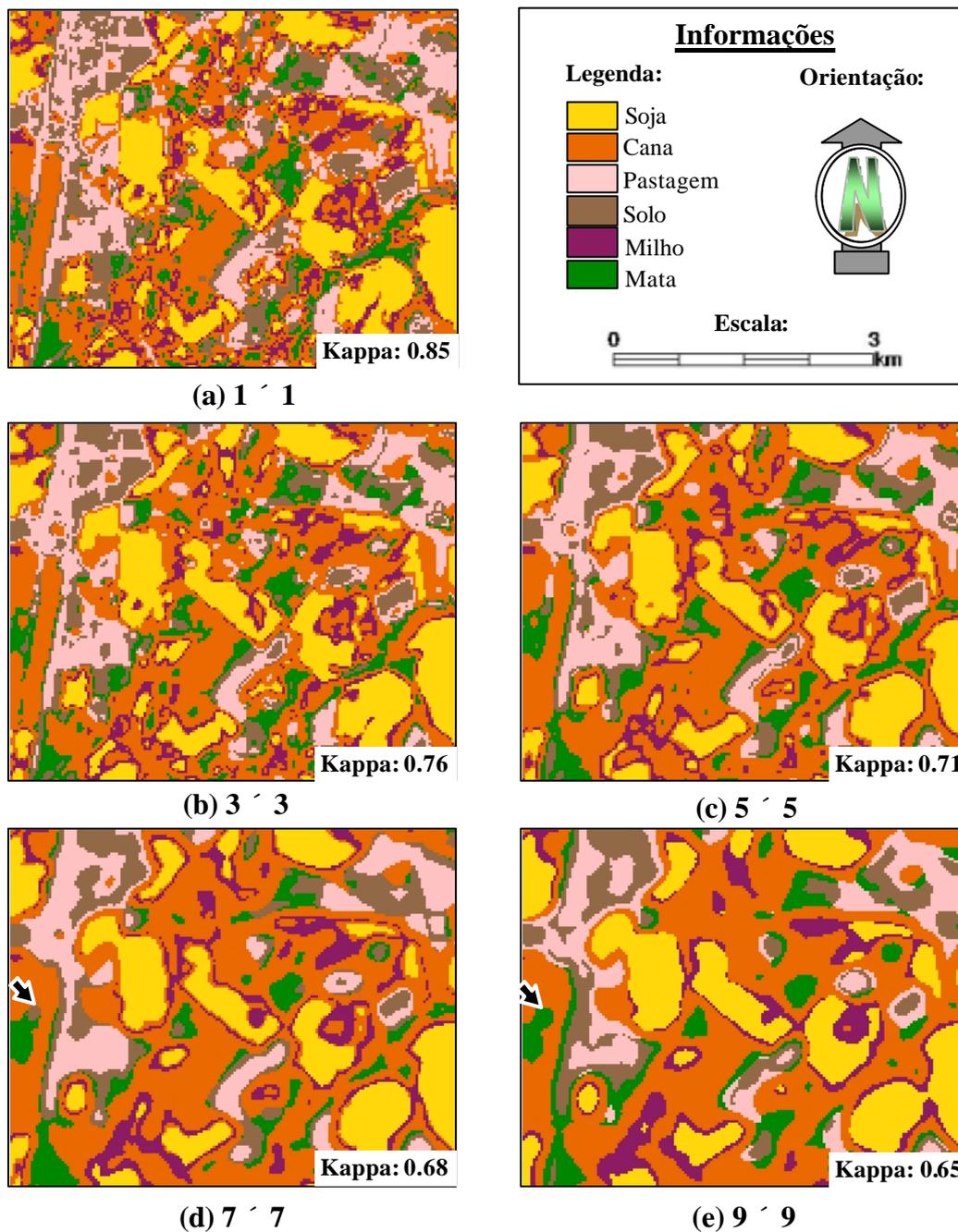


FIGURA 5.1 – Classificação não-supervisionada da imagem ETM+/Landsat-7 (1x1 pixel) e das imagens resultantes da filtragem de textura com janelas de 3x3 (90m), 5x5 (150m), 7x7 (210m) e 9x9 (270m) *pixels*, com os respectivos índices Kappa.

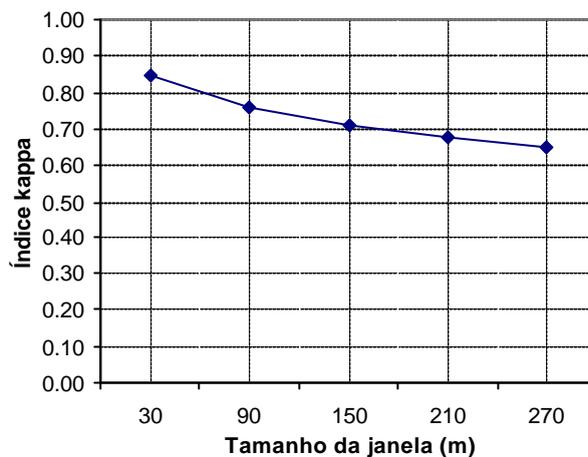


FIGURA 5.2 – Variação do índice Kappa em função do tamanho da janela (resolução espacial) utilizada na filtragem de textura.

Os resultados da filtragem de maioria são apresentados na Figura 5.3 com o mesmo objetivo comparativo realizado para a análise da filtragem de textura. Observa-se a suavização latente das feições que compõem o mapa. Tendo em vista que o processo é efetuado sobre a imagem temática, há um maior controle sobre as alterações que são causadas devido à variação do tamanho da janela de filtragem. Portanto, os estudos da variação das classes temáticas ao longo das escalas, realizados com imagens filtradas por maioria, devem mostrar uma evolução bastante coerente, sem apresentar eventuais descontinuidades de tendência.

Comparativamente à filtragem de textura (Figura 5.1), no entanto, pode-se perceber que a agregação com a filtragem de maioria (Figura 5.3) é muito mais severa. As classes mais fragmentadas tendem, portanto, a sofrer maiores transformações.

Com respeito à variação dos valores do índice Kappa com a filtragem de maioria (Figura 5.4), percebe-se uma tendência praticamente estacionária com ligeira tendência ao crescimento do indicador. Essa constatação, contrária à observada para a filtragem de textura (Figura 5.2), encontra justificativa na forte agregação realizada pela análise de maioria, que elimina eficientemente classes espúrias nas amostras coletadas que são comparadas ao mapa de verdade de campo.

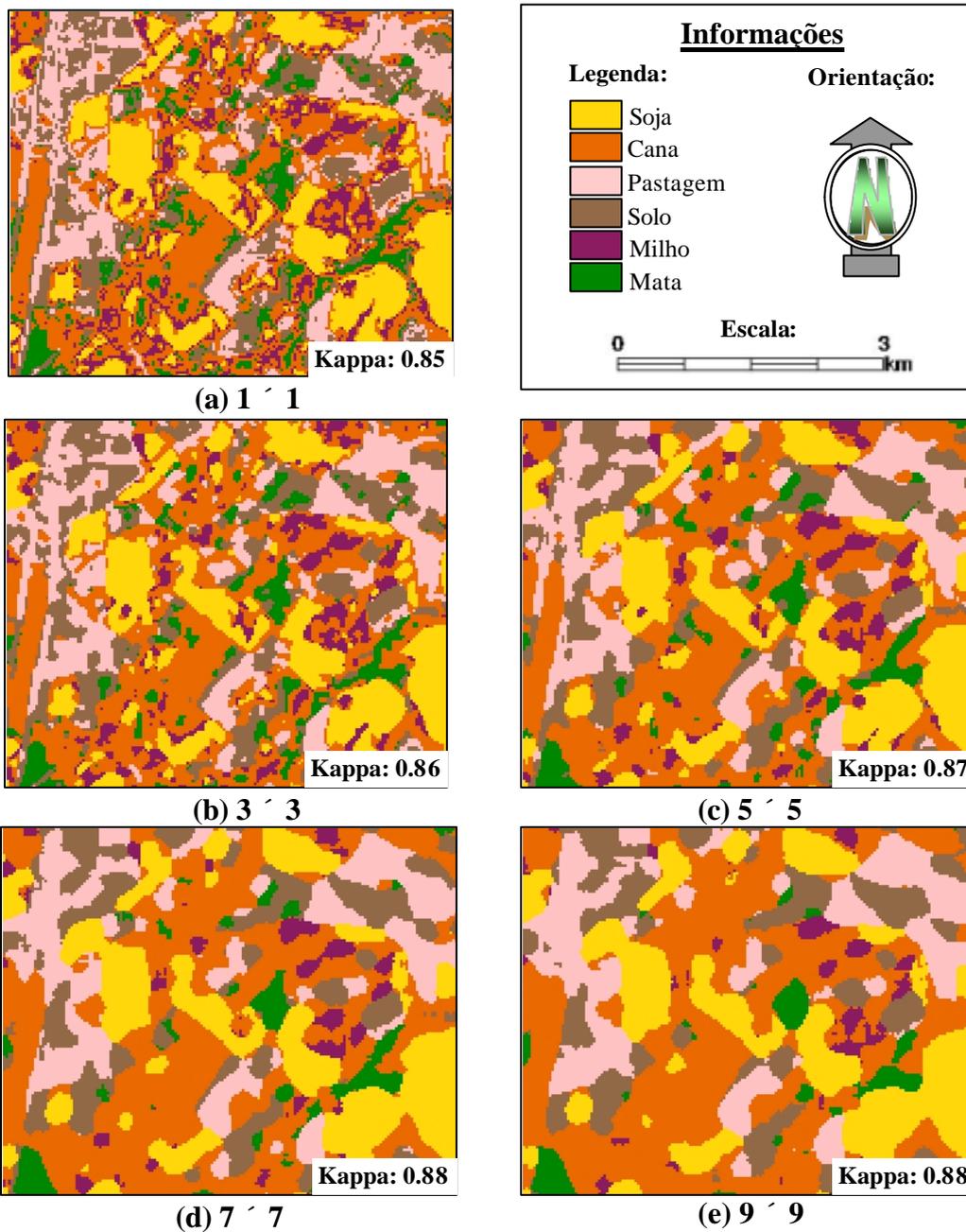


FIGURA 5.3 – Classificação não-supervisionada da imagem ETM+/Landsat-7 (1x1 pixel) e das imagens resultantes da filtragem de maioria com janelas de 3x3 (90m), 5x5 (150m), 7x7 (210m) e 9x9 (270m) *pixels*. com os respectivos índices Kappa.

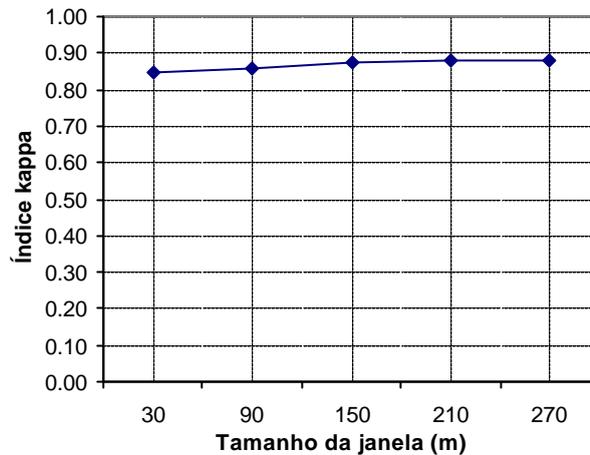


FIGURA 5.4 – Variação do índice Kappa em função do tamanho da janela (resolução espacial) utilizada na filtragem de maioria.

Pela comparação das Figuras 5.1 e 5.3 observa-se, ainda, que as classes cana e soja dominam o cenário, apresentando áreas maiores e mais compactas, tendendo a agregar as demais classes à medida que se aumenta o tamanho da janela de filtragem. Em sentido oposto, o milho e a mata são as classes mais fragmentadas, sendo compostas por polígonos menores e mais dispersos, sofrendo, portanto, muito mais severamente o efeito de degradação da resolução espacial.

## 5.2 Evolução das Métricas de Paisagem com a Simulação da Resolução Espacial

### 5.2.1 Filtragem de Textura

#### 5.2.1.1 Índices de Área/Densidade/Borda

A Figura 5.5 apresenta a evolução dos índices de área/densidade/borda, de acordo com a variação do tamanho da janela de filtragem de textura, para cada uma das classes envolvidas nas análises.

Na Figura 5.5, pode-se observar que os índices *Área* e *PLAND* (Porcentagem de Cenário) apresentaram resultados idênticos para todas as classes, denotando a alta correlação existente entre eles. As classes que mostraram variações mais intensas de área na primeira filtragem (3 x 3 pixels ou 90m) foram cana e solo exposto. Na filtragem de 9 x 9 pixels, aproximadamente equivalente à resolução espacial do sensor

MODIS, a mata apresentou um incremento de 70% em sua área, que cresceu de 18.800 ha (dado original ETM+ com 30m de resolução) para 31.600 ha (resolução espacial simulada de 270m). No sentido inverso, o solo exposto apresentou uma redução de área de 65.700 ha (30m) para 35.900 ha (270m), equivalendo, portanto, a cerca de 45% de decréscimo de área.

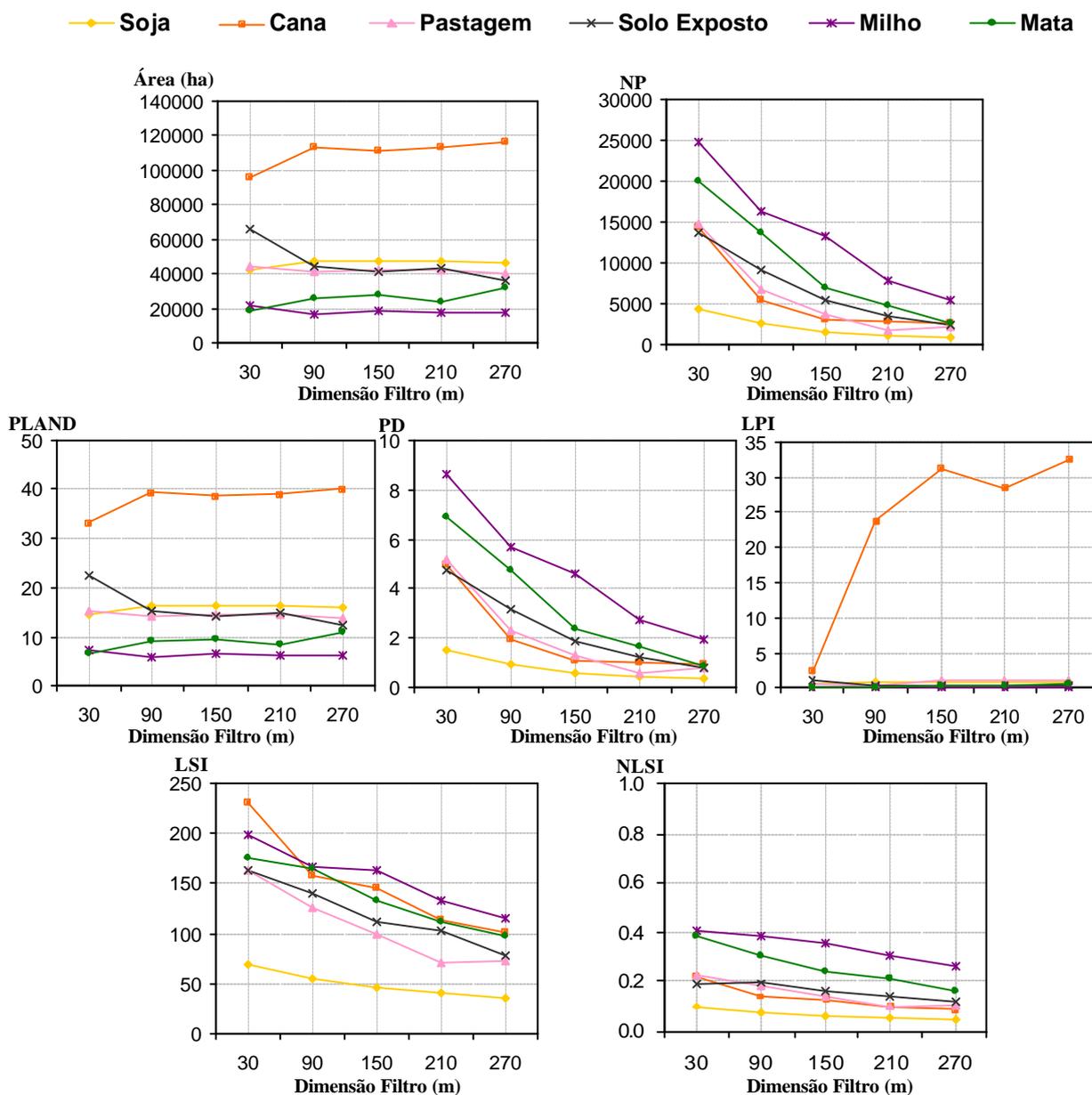


FIGURA 5.5 – Comportamento dos índices de área/densidade/borda à variação da dimensão do filtro de textura (resolução espacial).

Os indicadores *NP* (Número de Polígonos), e *PD* (Densidade de Área) apresentaram também comportamento idêntico entre si na Figura 5.5, o qual pode ser explicado pela análise comparativa das Equações 2.3 e 2.5, que demonstram a combinação linear existente entre eles. Conforme era esperado, com a eliminação de polígonos pouco representativos em função da degradação, houve uma redução muito significativa do número de polígonos (*NP*) para todas as classes que compõem as análises.

Os índices *LSI* (Forma de Classe) e *NLSI* (Forma de Classe Normalizada), correlacionados entre si, também mostraram uma redução na magnitude de seus valores com a degradação da resolução espacial de 30m para 270m, decorrente da queda da complexidade dos polígonos.

Finalmente, a análise do gráfico de *LPI* (Maior Área), demonstrou que a maior área de cana aumentou de 2% para 32% do cenário, evidenciando que desde a primeira variação do tamanho da janela de filtragem (3 x 3 pixels) já houve forte agregação por parte da referida classe. Essa constatação pode ser observada também na Figura 5.1, que mostra a classe cana se transformando em um grande polígono, decorrente do desaparecimento de polígonos muito fragmentados. As demais classes, no entanto, não foram muito afetadas pela variação da resolução espacial da imagem.

#### **5.2.1.2 Índices de Área Nuclear**

A evolução dos índices de área nuclear, em função da degradação da resolução espacial, é apresentada na Figura 5.6. As diversas classes segregadas na classificação digital apresentam-se separadas.

As métricas *TCA* (Área Nuclear Total) e *CPLAND* (Porcentagem Nuclear de Cenário), apresentaram o mesmo comportamento para todas as classes e para todos os tamanhos de janela (Figura 5.6). O desempenho de *TCA* foi exatamente como previsto, pois à medida que se reduz a resolução espacial espera-se que os polígonos que vão se degradando tendam a não apresentar área nuclear. As classes mais fragmentadas tendem a desaparecer, como se observou para a mata e o milho, que, a partir de 150m,

apresentaram a nulidade para *TCA* e *CPLAND*. É notável, ainda, o comportamento muito similar desses índices para as referidas classes.

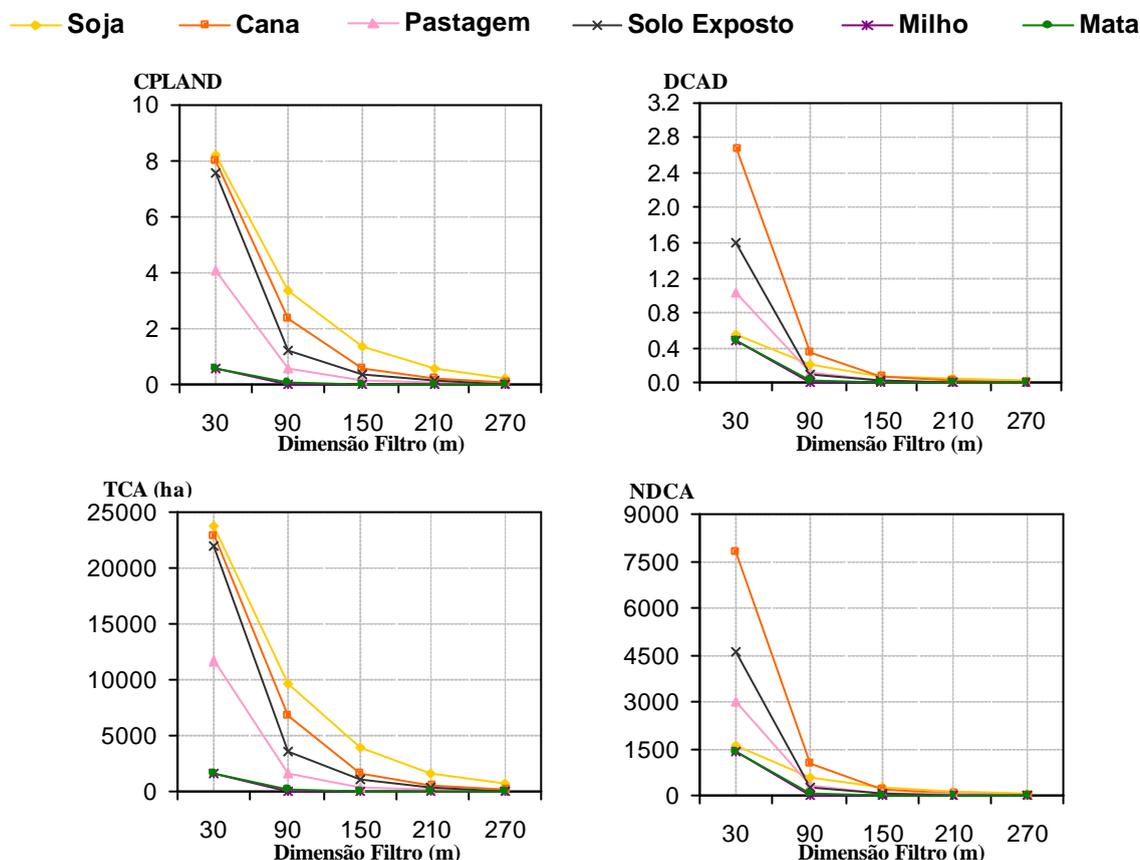


FIGURA 5.6 – Comportamento dos índices de área nuclear à variação da dimensão do filtro de textura (resolução espacial).

Pelo mesmo motivo de *TCA* e *CPLAND*, *DCAD* (Densidade de Área Nuclear Disjunta), e *NDCA* (Número de Áreas Nucleares Disjuntas) trazem a mesma informação para as análises. A observação da Figura 5.6 confirma a combinação linear que as Equações 2.11 e 2.12 denotaram.

A saliente redução de *NDCA* tem explicação semelhante à já apresentada. O desaparecimento de polígonos espacialmente fragmentados ou de dimensão reduzida causa queda excessiva no número de elementos dotados de área nuclear.

### 5.2.1.3 Índices de Contágio/Espalhamento

Novamente, é possível observar na Figura 5.7 que há duas métricas perfeitamente correlacionadas. *AI* (Agregação) e *PLADJ* (Porcentagem de Adjacência) apresentam gráficos, representando o valor do índice através das variações do tamanho da janela de filtragem, idênticos. O acréscimo nos valores destes dois indicadores com a degradação da resolução espacial é decorrente da maior agregação que as classes tendem a assumir para escalas moderadas.

Também medindo a agregação espacial das classes, o índice *CLUMP* (Agrupamento) apresentou resultado bastante semelhante a *AI* e *PLADJ*. Porém, a diferença apresentada para a classe cana indicou a necessidade de maiores estudos sobre a métrica, que se comportou de forma previsível ao denotar aumento da agregação espacial com a degradação da resolução espacial.

O índice *IJI* (Interposição e Justaposição) mostrou-se pouco influenciado pela resolução espacial a partir de 90 m. Quando se compara o incremento do índice para a primeira variação de tamanho da janela de filtragem com as demais mudanças ocorridas nas demais janelas, percebe-se que estas não são consideravelmente significativas.

A Figura 5.7 mostra, ainda, o comportamento da métrica *DIV* (Divisão de Cenário), que é pouco influenciada pela variação da resolução espacial, com exceção da classe cana. Em consonância com esta constatação, a Figura 5.1 ilustra a forte redução da granulometria desta classe.

O gráfico do indicador *MESH* (Tamanho de Trama Efetiva), função da resolução espacial, apresentou perfil inverso ao do *DIV*. Sua definição, dada pela Equação 2.20, corrobora a característica extraída da Figura 5.7. Novamente, apenas a cana se destaca quanto à variação da magnitude da métrica.

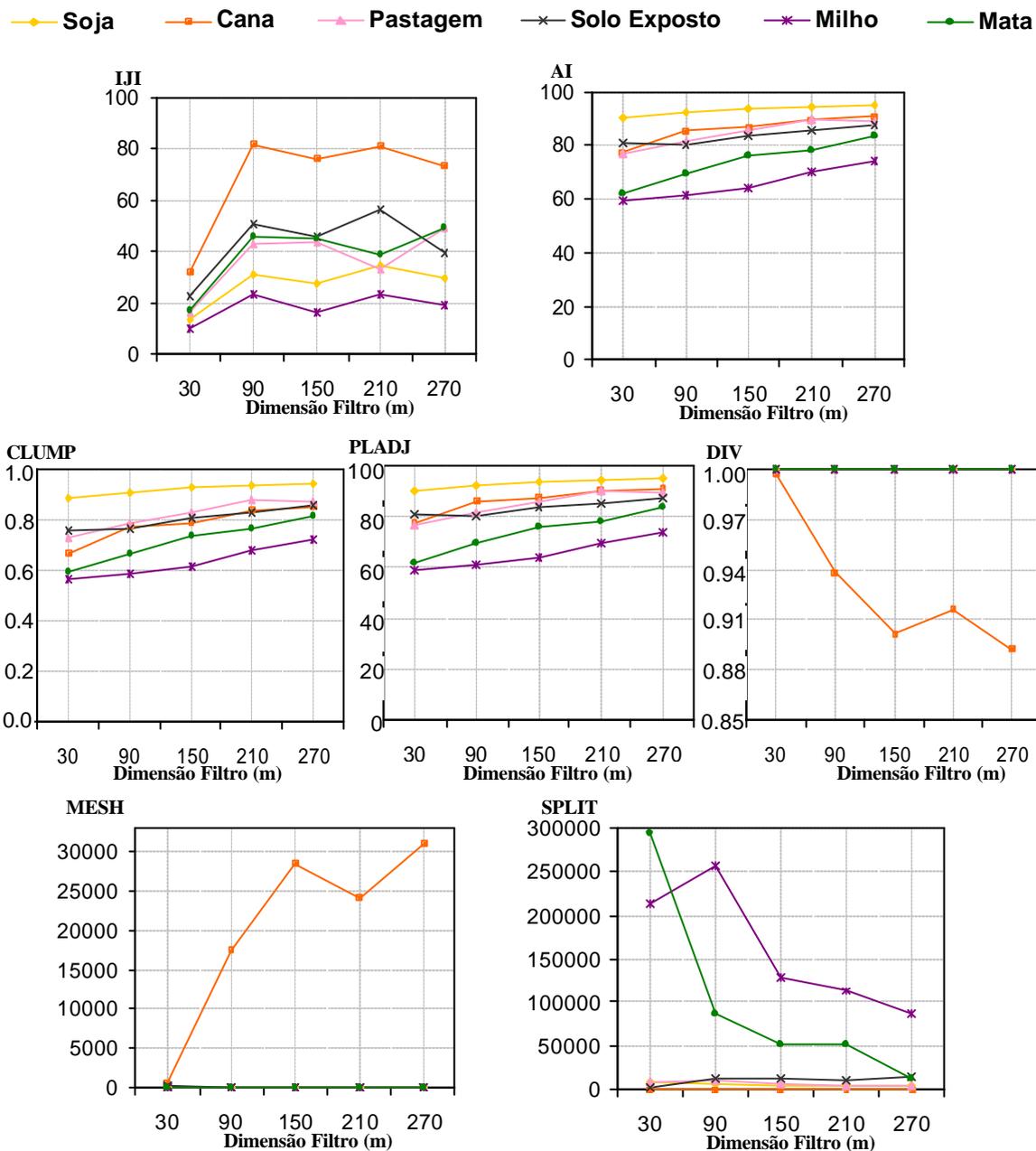


FIGURA 5.7 – Comportamento dos índices de contágio/espalhamento à variação da dimensão do filtro de textura (resolução espacial).

O índice *SPLIT* (Fragmentação), por fim, não mostrou uma tendência específica para todas as classes, apesar de as Equações 2.18 e 2.19 sugerirem uma forte semelhança deste indicador com o *DIV*. É notável, no entanto, o destaque ocorrido para as classes mais fragmentadas, mata e milho. Para elas, é nítido o efeito de célere redução do

número de polígonos que as compõem, típico da diminuição da fragmentação causada pela degradação de resolução espacial.

#### 5.2.1.4 Índices de Forma e de Conectividade

A Figura 5.8 exprime o gráfico de PAFRAC (Dimensão Fractal Perímetro-Área), em função da variação do tamanho da janela de filtragem, demonstrando uma queda significativa para a complexidade das diversas culturas. O milho, no entanto, apresentou um comportamento errático para 90m e 150m, regularizando a partir de então.

O índice COHE (Coesão de Área), por sua vez, apresentou, como esperado, tendência de crescimento com a degradação da resolução, com uma feição inversa à do *SPLIT* abordado na seção anterior. As Equações 2.19 e 2.22 não dão uma indicação imediata desta propriedade. As classes milho e mata corroboram suas constituições dispersas, exibindo valores iniciais de coesão bastante inferiores às demais classes. Ao serem muito afetadas pela agregação espacial, têm seu patamares elevados acentuadamente em função da degradação da resolução espacial.

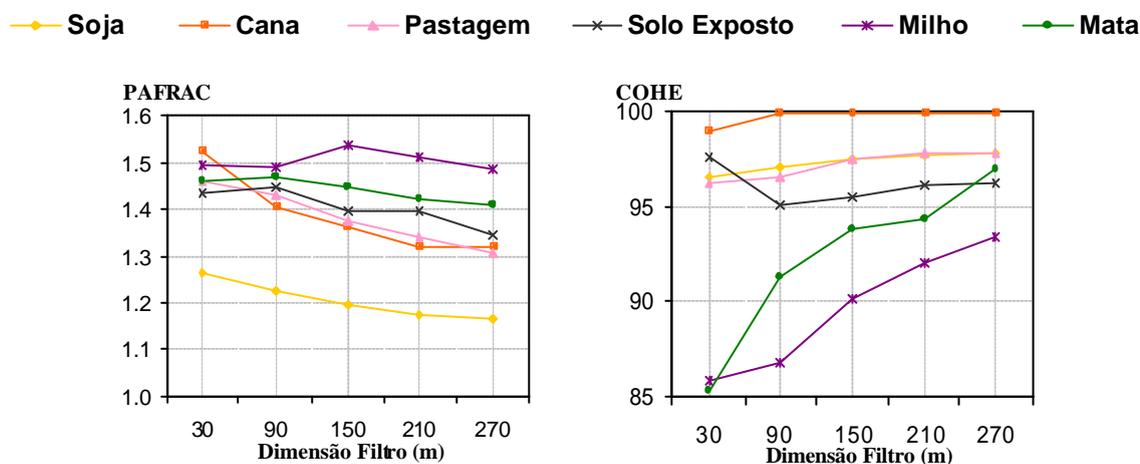


FIGURA 5.8 – Comportamento dos índices de forma (PAFRAC) e de conectividade (COHE) à variação da dimensão do filtro de textura (resolução espacial).

## 5.2.2 Filtragem de Maioria

### 5.2.2.1 Índices de Área/Densidade/Borda

Analogamente ao procedimento adotado para a filtragem de textura, a Figura 5.9 revela a variação dos índices de ADB, em função da resolução espacial simulada pela análise de maioria, para cada uma das classes discriminadas pela classificação digital. Perceberam-se intensas semelhanças na comparação das abordagens, apesar da nítida diferença de contornos mais suaves e controlados proporcionados pela maioria.

Confirmando o comportamento já reportado, os índices *Área* e *PLAND* convergiram perfeitamente para a totalidade das classes, corroborando a correlação elevada existente entre eles.

Os indicadores *NP* e *PD*, assim como *Área* e *PLAND*, não exibiram diferença alguma entre si. Quando comparados graficamente aos mesmos índices obtidos pela filtragem de textura, *NP* e *PD* foram mais bem delineados e previsíveis. Deve-se notar, no entanto, que para classes mais fragmentadas, como milho e mata, a redução no número de polígonos foi bem mais acentuada na análise de maioria. A Tabela 5.1 expõe a comparação para essas classes nas diferentes filtrações. Já a Tabela 5.2 traz a comparação para a cana e a soja, com disparidades menos significativas entre as diferentes simulações, uma vez que são classes dominantes no cenário estudado.

Para os índices *LSI* e *NLSI* há resultados bastante semelhantes quando comparados aos obtidos com a filtragem de textura. A análise comparativa das Figuras 5.5 e 5.9 ilustra bem essa ocorrência.

Enfim, o gráfico de *LPI* (Figura 5.9) demonstra que até a janela de 150m o índice sofreu um suave acréscimo em sua magnitude, para a classe cana e, apenas a partir de então, pôde-se observar um incremento mais substancial. Há, portanto, certa diferença quando se compara com a filtragem de textura, que desde a primeira variação do tamanho da janela de filtragem já ocasionou elevado aumento da área de cana de cerca de 2% para 24% do cenário, atingindo seu ápice em 32%. A justificativa para tal discrepância gira

em torno do maior controle observado quando se operam degradações sobre um mapa temático, ao invés da geração de novas imagens a partir da original.

—◆— Soja    —□— Cana    —△— Pastagem    —×— Solo Exposto    —\*— Milho    —●— Mata

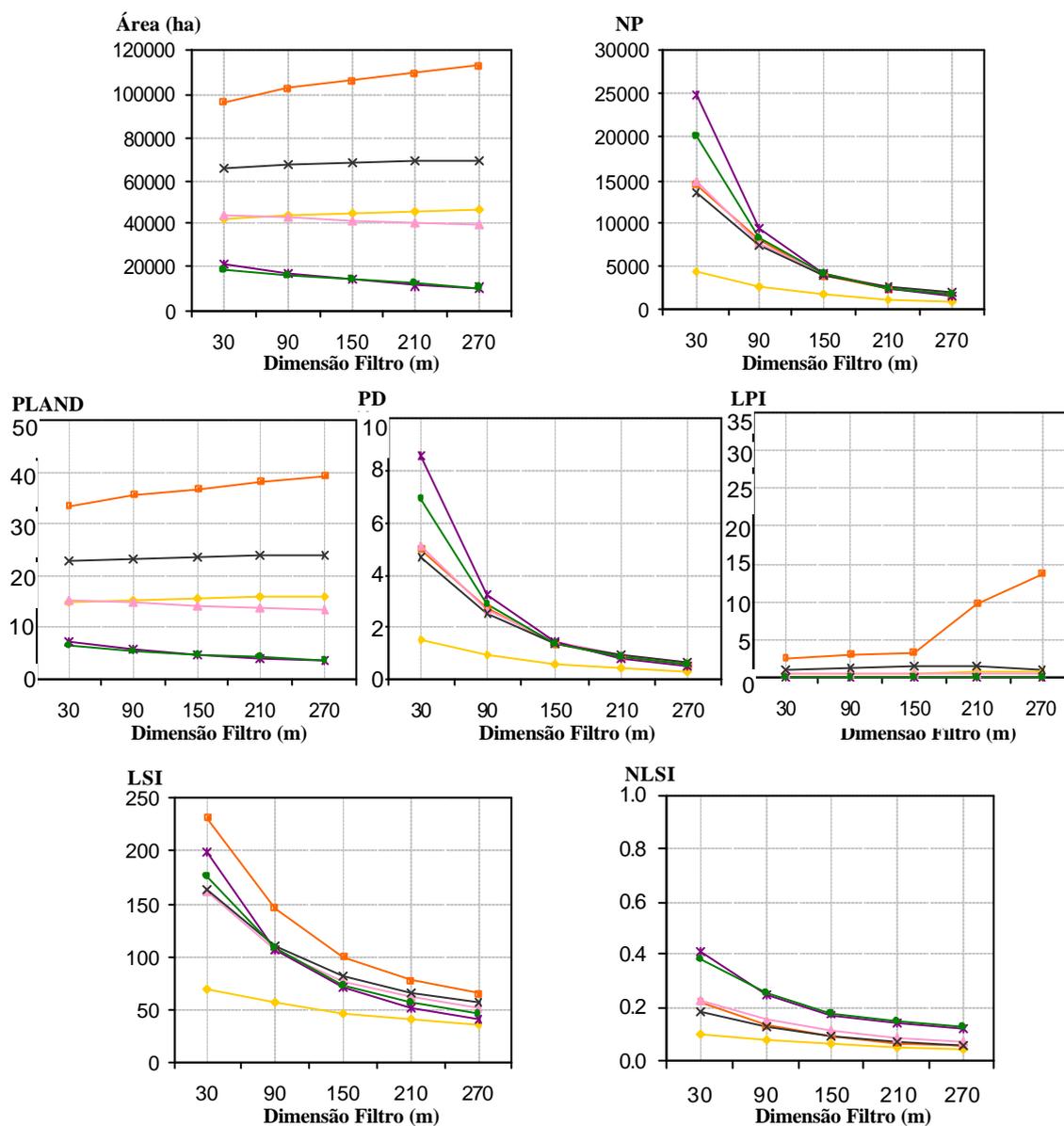


FIGURA 5.9 – Comportamento dos índices de área/densidade/borda à variação da dimensão do filtro de maioria (resolução espacial).

TABELA 5.1 – Comparação do Número de Polígonos (*NP*) para as classes milho e mata nas diferentes filtragens.

<b>Tamanho da Janela (m)</b>	<b>Milho</b>		<b>Mata</b>	
	<b>NP Textura</b>	<b>NP Maioria</b>	<b>NP Textura</b>	<b>NP Maioria</b>
30	24.819	24.819	20.031	20.031
90	16.400	9.311	13.607	8.299
150	13.328	4.066	6.932	3.993
210	7.919	2.303	4.724	2.459
270	5.518	1.533	2.563	1.695

TABELA 5.2 – Comparação do Número de Polígonos (*NP*) para as classes soja e cana nas diferentes filtragens.

<b>Tamanho da Janela (m)</b>	<b>Soja</b>		<b>Cana</b>	
	<b>NP Textura</b>	<b>NP Maioria</b>	<b>NP Textura</b>	<b>NP Maioria</b>
30	4.262	4.262	14.406	14.406
90	2.622	2.653	5.520	7.878
150	1.590	1.654	3.090	3.847
210	1.182	1.156	2.805	2.419
270	946	899	2.630	1.647

### 5.2.2.2 Índices de Área Nuclear

A Figura 5.10 traz a representação gráfica, que delinea a progressão dos índices de área nuclear à medida que degradações sucessivas são aplicadas, utilizando a filtragem de maioria como ferramenta.

Os resultados observados na Figura 5.10 são extremamente semelhantes àqueles obtidos para a filtragem de textura (Figura 5.6). Não há desvio algum que mereça destaque. Portanto, todas as constatações e argumentações são exatamente idênticas, sugerindo-se, assim, a consulta da seção 5.2.1.2.

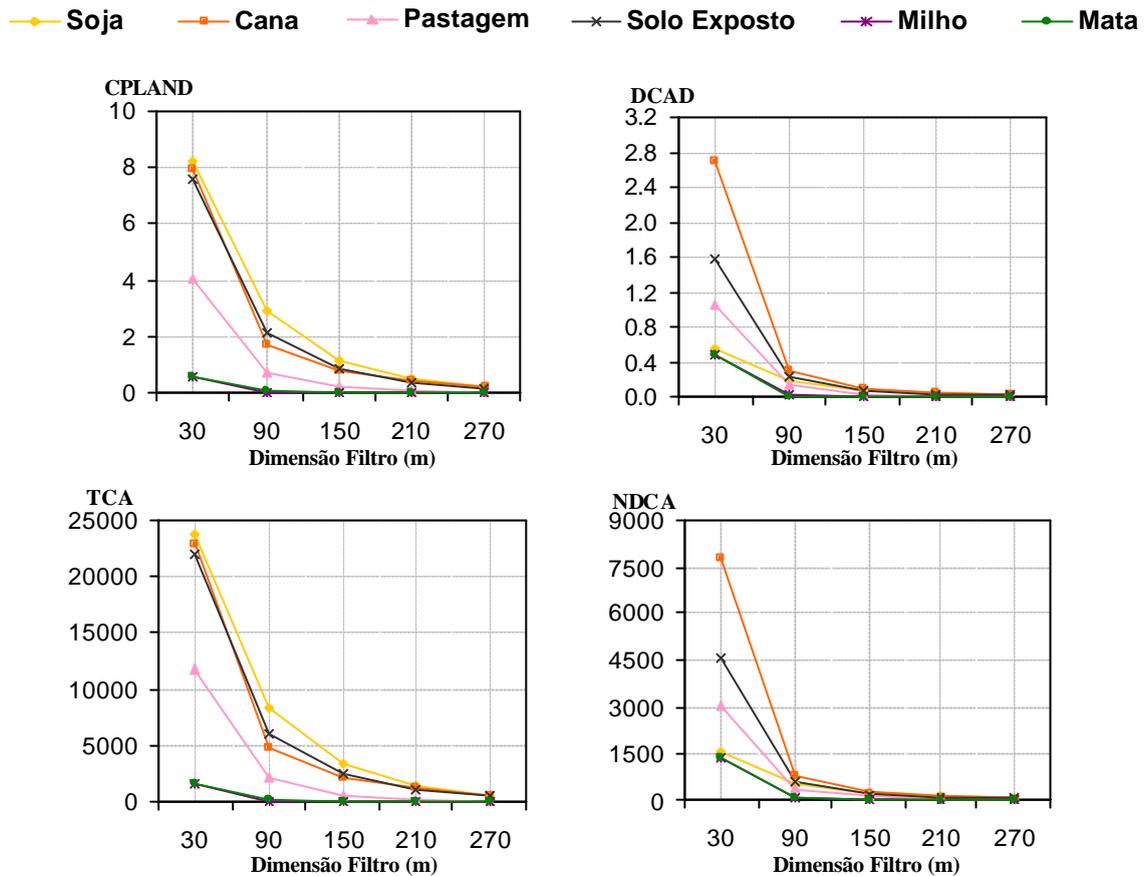


FIGURA 5.10 – Comportamento dos índices de área nuclear à variação da dimensão do filtro de maioria (resolução espacial).

### 5.2.2.3 Índices de Contágio/Espalhamento

As métricas *AI* e *PLADJ*, conforme era previsível, manifestaram resultados idênticos novamente, o que pode ser conferido na Figura 5.11. Não há nada que mereça comentários distintos aos já efetuados para os mesmos índices calculados sobre o resultado da filtragem de textura, exceto a concorrência ligeiramente mais acentuada para a tendência do índice, com perfis cada vez mais próximos à proporção que a janela de filtragem aumenta. O mesmo fenômeno pode ser observado para o indicador *CLUMP*.

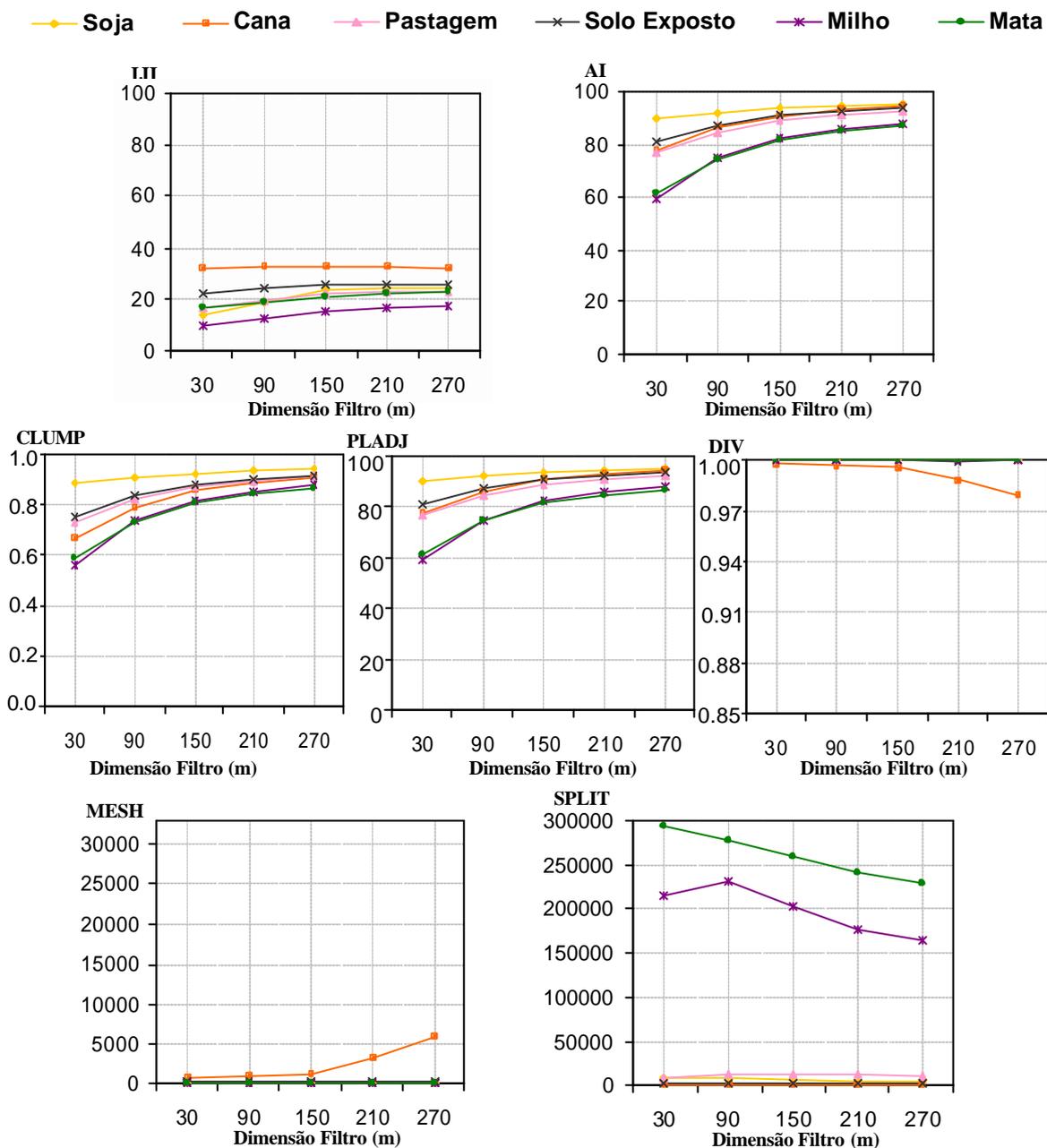


FIGURA 5.11 – Comportamento dos índices de contágio/espalhamento à variação da dimensão do filtro de maioria (resolução espacial).

Confirma-se, por outro lado, a reduzida influência da resolução espacial sobre *III* para qualquer escala. Esta informação, que pode ser observada na Figura 5.11, é muito importante, pois é sinal da forte utilidade do índice para compor um modelo estatístico que refine a estimativa de área obtida por imagens de resolução espacial moderada.

Com relação aos índices *DIV* e *MESH*, eles obtiveram desempenho análogo ao atingido pela análise de textura (Figura 5.7). Na Figura 5.11 nota-se, no entanto, que a extensão dos valores é menor, efeito similar ao observado para a métrica *LPI*.

Concluindo, o índice *SPLIT* continuou com comportamento típico de redução da fragmentação das classes mais dispersas (milho e mata) e constância para as demais. É notável a semelhança dos resultados aqui obtidos com aqueles divulgados para a filtragem de textura. Ocorre, entretanto, a peculiaridade de feições menos abruptas, o que ocasionou menor redução nos patamares da métrica *SPLIT* para a análise de maioria.

#### **5.2.2.4 Índices de Forma e de Conectividade**

A Figura 5.12 exprime o gráfico de *PAFRAC* e *COHE*, em função da variação do tamanho da janela de filtragem, não havendo necessidade de adicionais esclarecimentos aos já mencionados na seção 5.2.1.4. O ambiente mais controlado da filtragem de maioria, todavia, foi suficiente para eliminar a pequena inconsistência observada em *PAFRAC*, para o milho, bem como atenuou o acentuado acréscimo de *COHE*, para o milho e a mata.

### **5.3 Aspectos Comparativos entre as Filtragens de Maioria e de Textura**

Por todos os estudos realizados até este momento, pode-se concluir, de uma forma geral, que há uma forte semelhança entre os resultados obtidos por intermédio das filtragens de textura e de maioria. Vale notar, contudo, que, ao aumentar o tamanho da janela de filtragem em função dos diversos índices, a análise de maioria propiciou perfis muito mais bem delineados, com feições muito mais suaves nos gráficos descritivos.

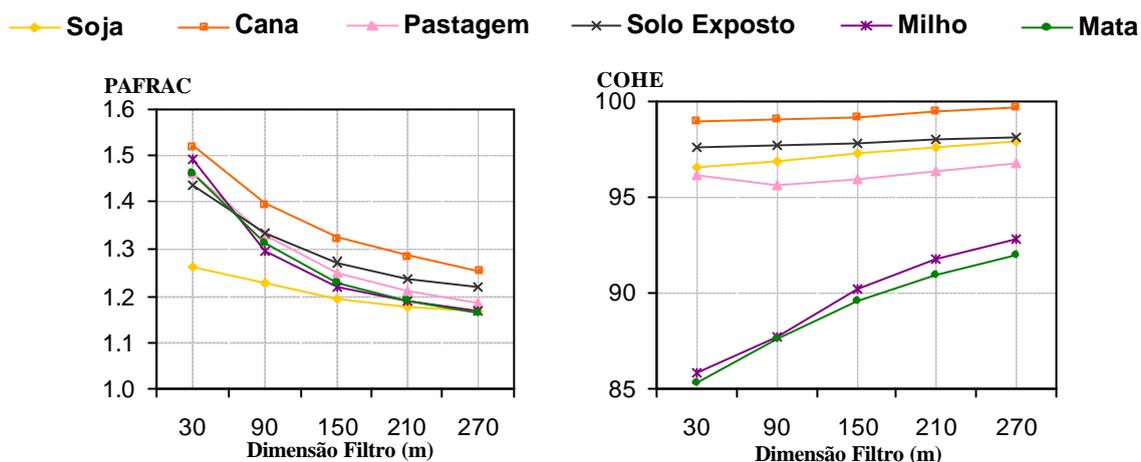


FIGURA 5.12 – Comportamento dos índices de forma (PAFRAC - Dimensão fractal Perímetro-Área) e de conectividade (COHE - Coesão de Área) à variação da dimensão do filtro de maioria (resolução espacial).

Para possibilitar uma análise comparativa mais ampla e abrangente, principalmente sobre as áreas das classes obtidas pelos diversos métodos de filtragem, a Figura 5.13 expõe os histogramas separados por processos e classes, além de conter a linha de tendência observada para a dimensão da variável.

Nota-se que as classes soja, cana e pastagem revelaram um comportamento similar à medida que se constata a degradação da resolução espacial. A soja e a cana são culturas pouco fragmentadas e compostas, via de regra, por grandes polígonos homogêneos, o que justifica essa independência ao método de filtragem utilizado. A pastagem, por sua vez, é uma classe de padrão espacial intermediário, nem tão fragmentada como milho e mata, nem tão compacta como soja e cana, sendo suficiente, entretanto, para garantir a estabilidade apresentada.

Fica claro, portanto, que a troca de elementos entre classes se reduziu à mata, milho e solo, esta com padrão espacial intermediário, enquanto aquelas bastante fragmentadas justificam inteiramente o desempenho observado. Curiosamente, solo foi a classe que exibiu a maior diferença quando comparados os dois métodos de filtragem, sendo esta disparidade repassada para as outras duas classes, principalmente para a mata, cujas linhas de tendência expõem um caráter inverso.

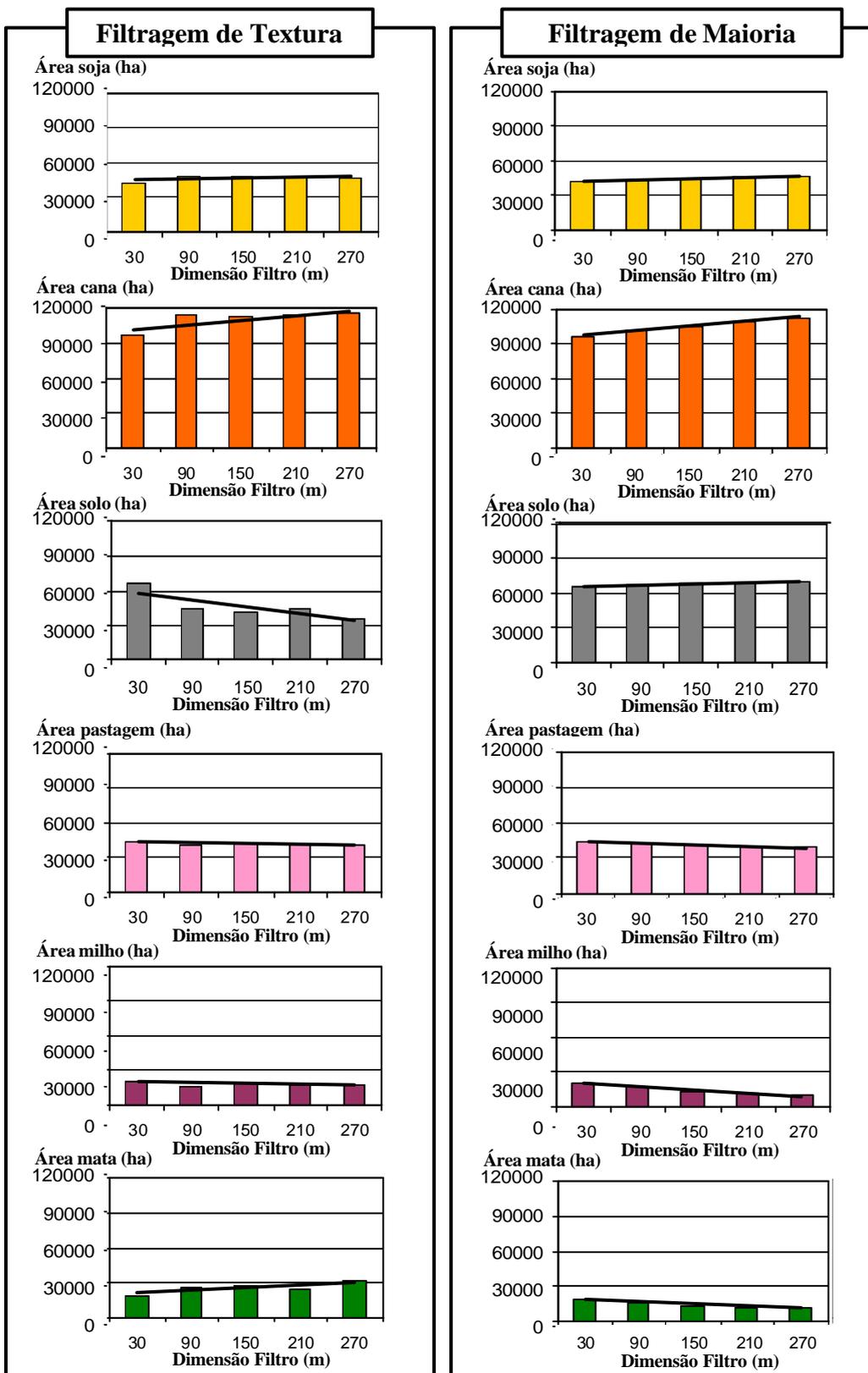


FIGURA 5.13 – Evolução da área das diversas culturas, em função da dimensão do filtro de textura e Maioria aplicado (resolução espacial).

Depois de efetuados os recortes das imagens e gerados, conseqüentemente, os 216 subconjuntos para cada uma das nove imagens temáticas, foi possível efetuar uma análise estatística exploratória da adequação das simulações à imagem ETM+ original. Pelas regressões observadas nas Figuras 5.14, 5.15 e 5.16, foi feita a regressão entre a área de cada subconjunto de imagem ETM+ e a área de cada um dos recortes correspondentes nas filtragens. As linhas de tendência vermelhas e os pontos magentas são relativos à análise de textura, enquanto os traços pretos e pontos azuis aludem à maioria, para as imagens de dimensão  $125 \times 125$  *pixels*. As referidas Figuras trazem, ainda, a equação das regressões, bem como os correspondentes coeficientes de determinação.

Para a cultura de soja, observou-se um  $R^2$  (coeficiente de determinação) muito próximo da unidade para as duas filtragens, com destaque para a maioria, para todos os tamanhos de janela. A Figura 5.14 sugere, ainda, que os coeficientes angulares das retas de regressão reduziram-se ligeiramente, indicando que à medida que se degradou a imagem ETM+ houve uma tendência de superestimação da classe soja. Isso é coerente com o padrão espacial compacto da cultura.

A cana-de-açúcar, igualmente, apresentou elevados valores de coeficiente de determinação, embora menores e menos estáveis do que o exposto pela soja. Conclui-se, também, que sua superestimação através das degradações foi um pouco mais elevada para as duas filtragens.

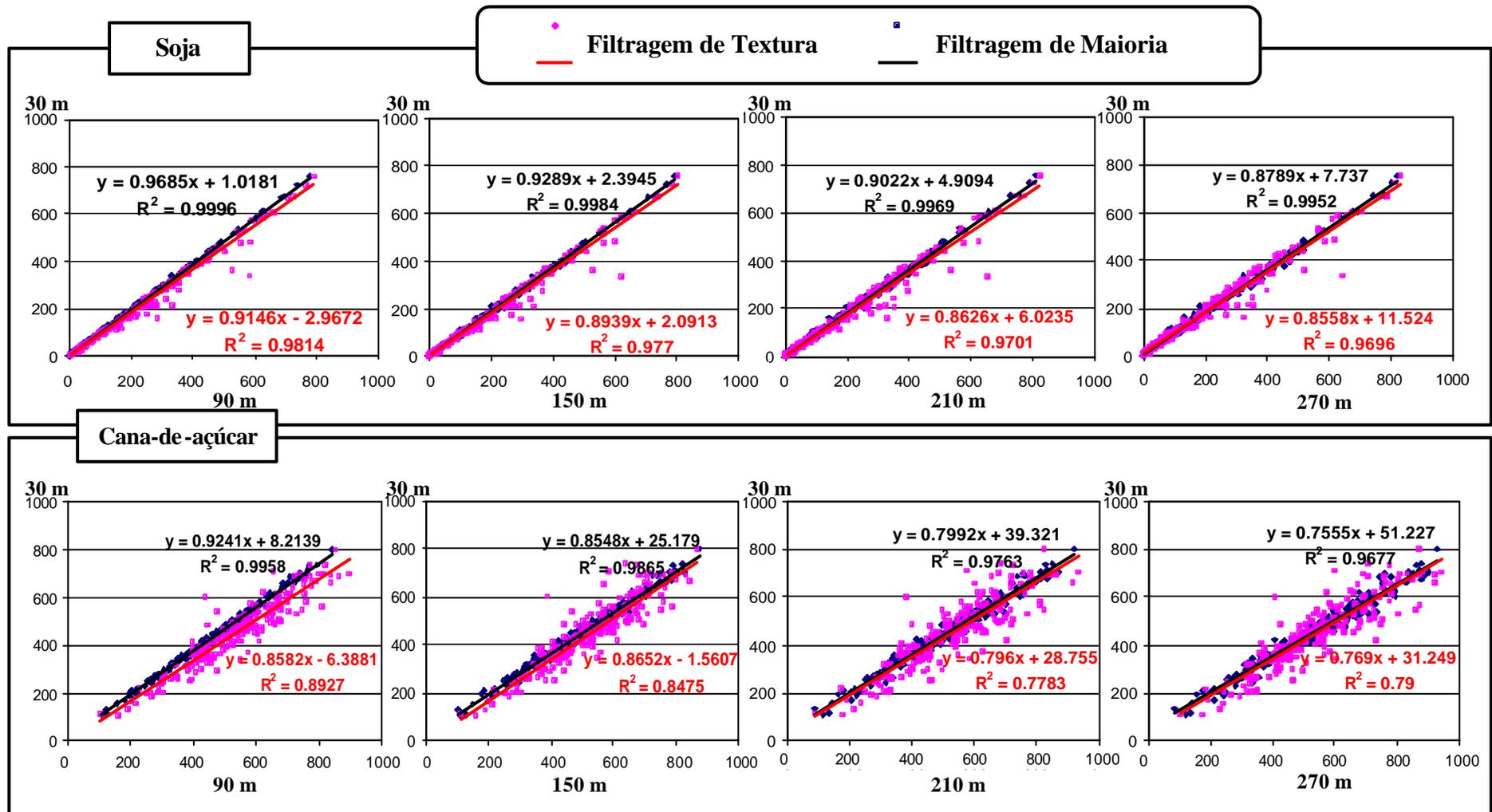


FIGURA 5.14 – Regressões entre as áreas das culturas cana e soja, em hectares, na imagem ETM+ e nas imagens resultantes dos processos de filtragem.

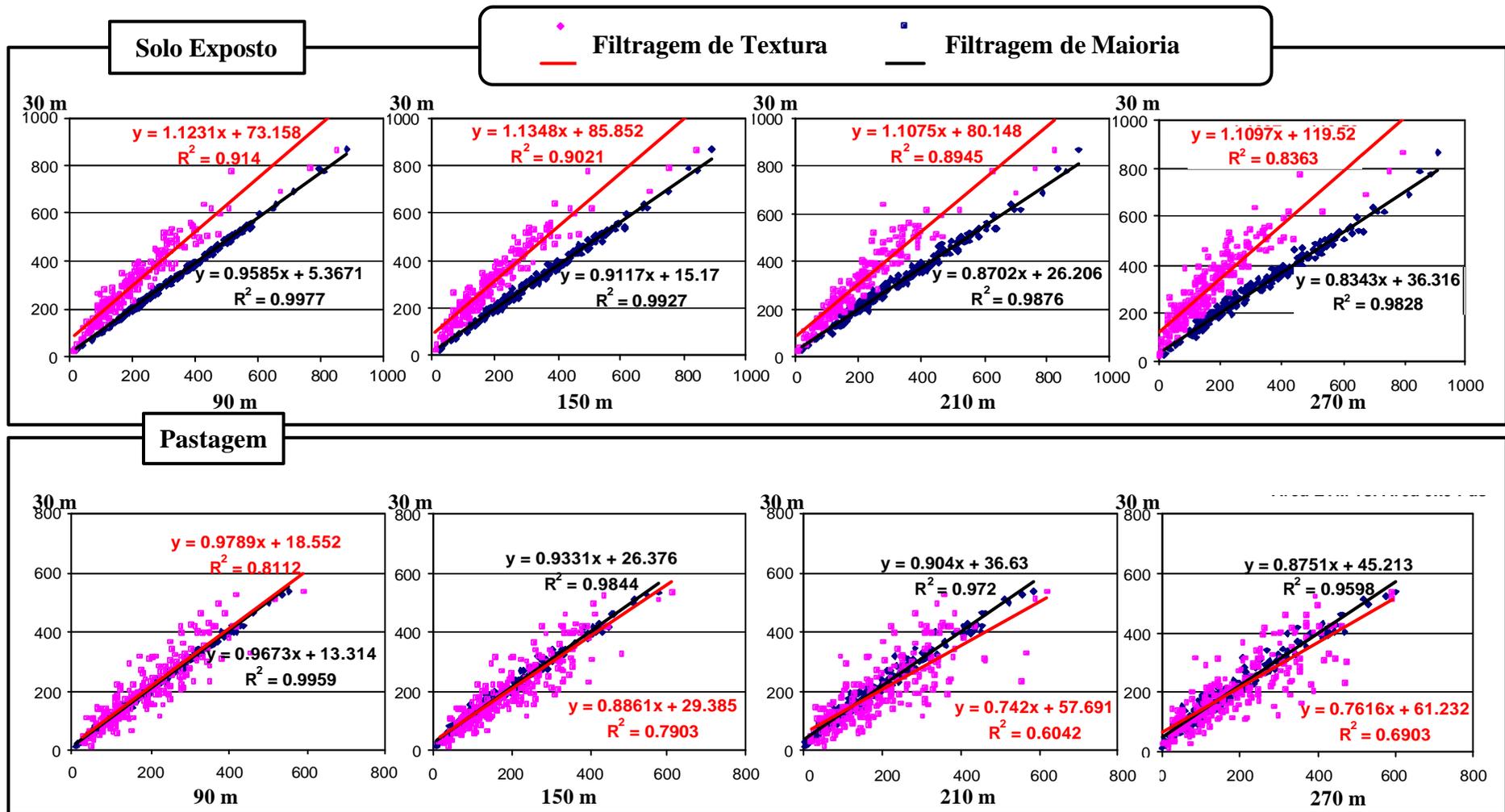


FIGURA 5.15 – Regressões entre as áreas das classes solo e pastagem, em hectares, na imagem ETM+ e nas imagens resultantes dos processos de filtragem.

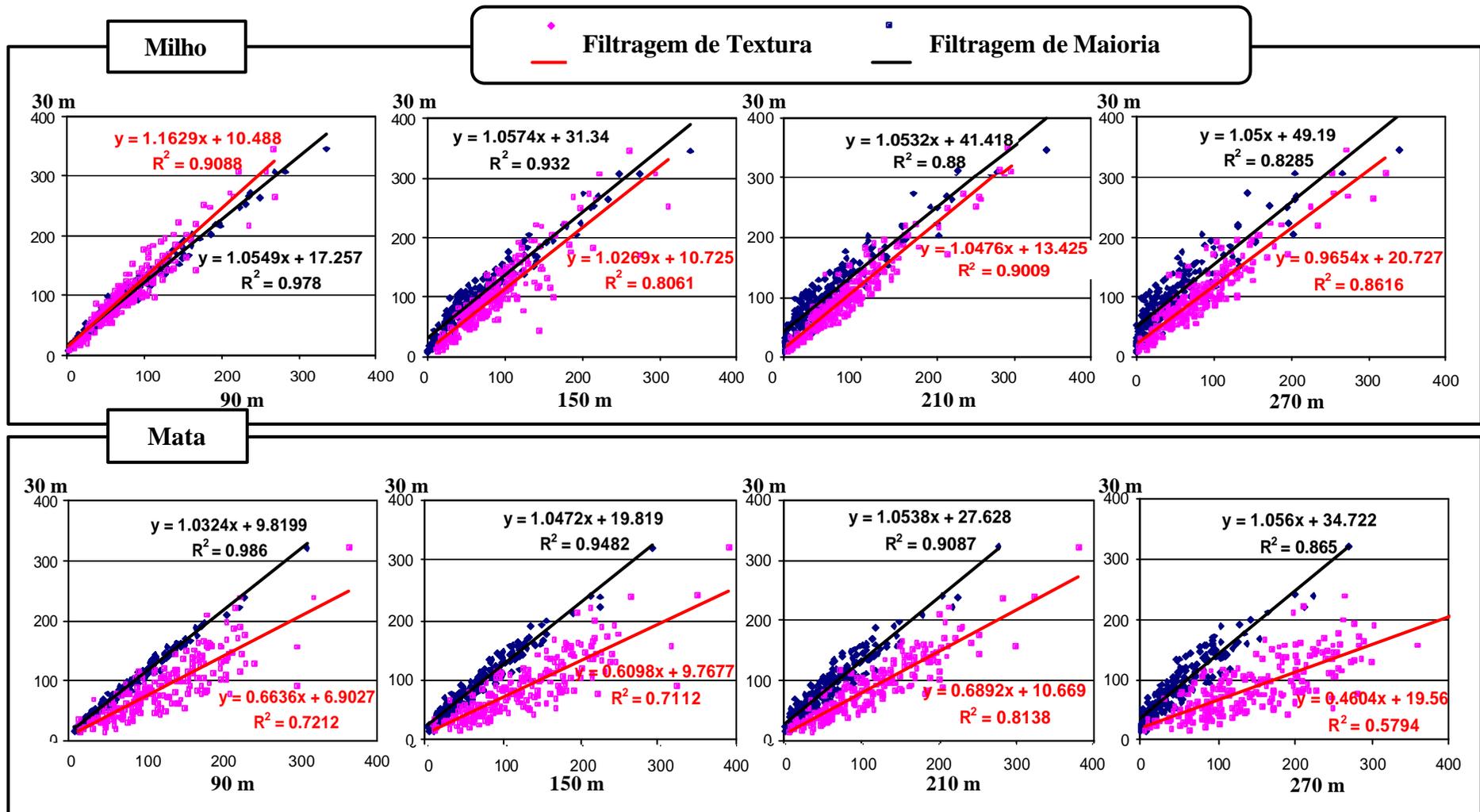


FIGURA 5.16 – Regressões entre as áreas das classes milho e mata, em hectares, na imagem ETM+ e nas imagens resultantes dos processos de filtragem.

A Figura 5.15, por sua vez, traz a evolução do solo exposto, cuja predição pelas degradações foi bastante satisfatória, sobretudo para a filtragem de maioria, que apresentou elevado  $R^2$ . Algo curioso é observado quando se analisa a inclinação das retas de regressão. Pela filtragem de textura houve uma tendência de subestimativa, enquanto a maioria tendeu, gradativamente, a superestimar a área de solo exposto presente no cenário.

A pastagem já apresentou de maneira mais notável os efeitos da degradação da resolução espacial. Os coeficientes de determinação para a filtragem de maioria continuaram altos. Porém, para a análise por textura foi nítido um declínio mais acentuado do  $R^2$ . A superestimação causada pela degradação foi observada e crescente.

A cultura de milho, exibida pela Figura 5.16, foi a primeira classe cujo coeficiente de determinação da regressão entre a área calculada pela imagem ETM+ e a obtida pelas filtragens de maioria foi inferior 0.9. Nela observou-se, também pela primeira vez, um desempenho melhor da filtragem de textura para algum tamanho de janela de filtragem. A alta fragmentação desta classe mostrou-se, portanto, suficiente para impor sua assinatura nos resultados.

A mata também alcançou uma maior redução para o  $R^2$  da regressão envolvendo a filtragem de maioria. Para a textura constatou-se o maior declínio observado, atingindo 0,58 para a janela de 270m. Enquanto que para filtragem de maioria houve tendência de manutenção da área, para a análise de textura ocorreu uma acentuada superestimação da área, com o aumento do fator de degradação da resolução espacial. Não há dúvida que o padrão espacial desta classe impactou diretamente no desempenho das imagens simuladas.

Para sintetizar o que se obteve nas Figuras 5.14, 5.15 e 5.16, a Figura 5.17 congrega os coeficientes de determinação para cada um dos processos de filtragem e separados por classes temáticas. Nela é mais nítido o comportamento do  $R^2$  para a análise de maioria, que tendeu a se reduzir, sobretudo para as classes milho e mata, as mais fragmentadas. Para a filtragem de textura, todavia, algumas discontinuidades foram reveladas. A

classe milho revelou uma queda excessiva para a janela de 150m, o que forçou uma elevação do  $R^2$  (coeficiente de determinação) para o filtro de dimensão de 210m, normalizando a partir de então. Já a mata apresentou um pico para o coeficiente de determinação em 210m, o que impeliu a ocorrência de acentuada redução em 270m, de forma a compor um perfil de redução do  $R^2$  ao longo das escalas. Finalmente, o perfil do gráfico da pastagem também sofreu uma descontinuidade pontual (210m), mas, assim como aconteceu para o milho e a mata, houve sua recomposição para a janela de filtragem subsequente (270m).

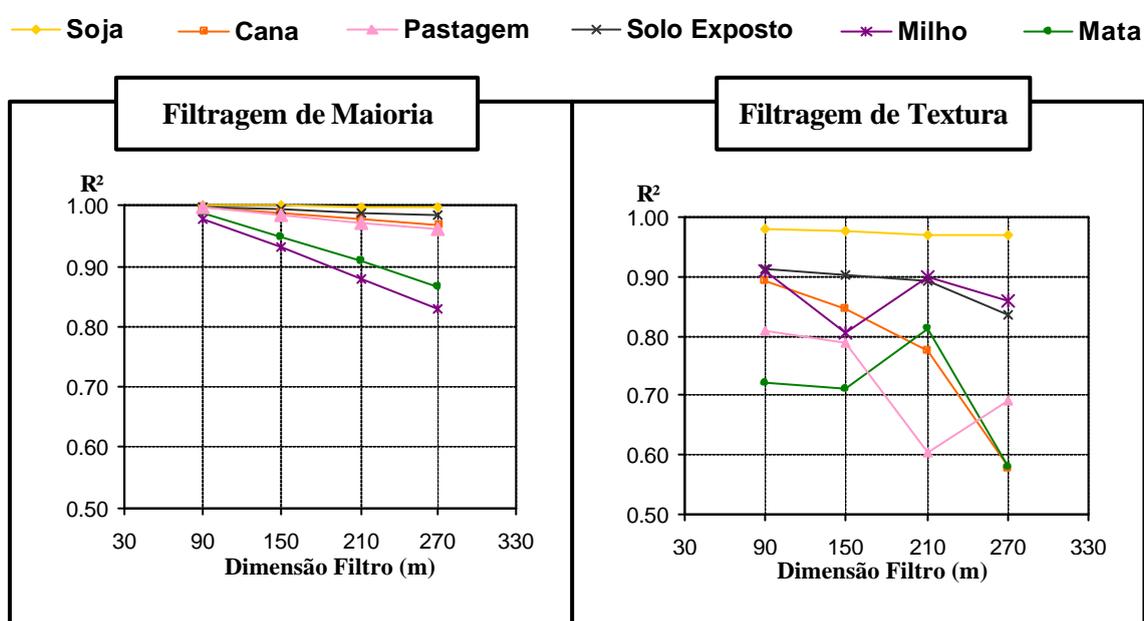


FIGURA 5.17 – Evolução do  $R^2$  das regressões entre as áreas obtidas dos dados simulados e aquelas extraídas da imagem do sensor ETM+/Landsat-7, para as diversas culturas.

A informação de descontinuidade é muito relevante, uma vez que não se está abordando a ocorrência de uma única observação, mas a tendência de um conjunto de 216 observações. O evento corrobora, novamente, a menor estabilidade da filtragem de textura, bem como expõe a fragilidade da estimativa de áreas mais fragmentadas com sensores de resolução espacial moderada.

A soja, que se manteve praticamente estável, bem como o solo e a cana exibiram um desempenho regular para a filtragem de textura, essa última com uma acentuada diminuição do coeficiente de determinação (Figura 5.17).

Finalmente, a Figura 5.17 traz a informação de que, a partir da resolução espacial de 150m, torna-se bastante limitada a reprodução do dado 30m. Tal efeito é destacado para as classes temáticas mais fragmentadas (milho e mata).

#### **5.4 Determinação da Técnica de Filtragem mais Realística para a Simulação da Resolução Espacial Moderada (MODIS/Terra)**

A Figura 5.18 ilustra de forma comparativa as relações entre as áreas das culturas obtidas pela simulação da resolução espacial de 270m (filtragens de textura e maioria) e aquela obtida da classificação digital da imagem original MODIS/Terra (250m). Tendo em vista que um claro objetivo deste trabalho é a avaliação da fidelidade do processo de simulação de imagens de resolução espacial moderada, as constatações advindas da Figura 5.18 foram muito pertinentes no intuito de selecionar o método mais adequado para representação de sensores hipotéticos.

A filtragem de textura apresentou coeficientes de determinação sempre maiores do que os alcançados pela filtragem de maioria, independentemente do padrão espacial da cultura. Para classes mais fragmentadas, entretanto, as diferenças de  $R^2$ , entre as abordagens, foi mais acentuada, com destaque para a mata.

A soja e a cana foram as culturas cuja simulação gerou resultados mais similares com a imagem MODIS. Para a filtragem de textura a pior estimativa foi obtida para a classe pastagem ( $R^2=0,48$ ), enquanto que a mata foi muito mal avaliada pela análise de maioria, atingindo um reduzido  $R^2$  de 0,17.

Finalmente, a Figura 5.19 expõe uma regressão exploratória global da estimativa de área para todas as classes analisadas, na qual as simulações, cujo tamanho de janela é 270m, explicam a área calculada pela imagem MODIS. Dessa forma o número de observações disponíveis para o procedimento foi de 1296, ou seja, 216 para cada uma das 6 classes presentes no cenário. Conforme era previsível, a filtragem de textura conseguiu predizer com muito mais eficiência a área estimada pela classificação digital da imagem MODIS. Assim sendo, a análise de maioria foi menos eficiente, atingindo um

coeficiente de determinação da ordem de 0,48, inferior, portanto, ao de 0,79, apresentado pela filtragem de textura.

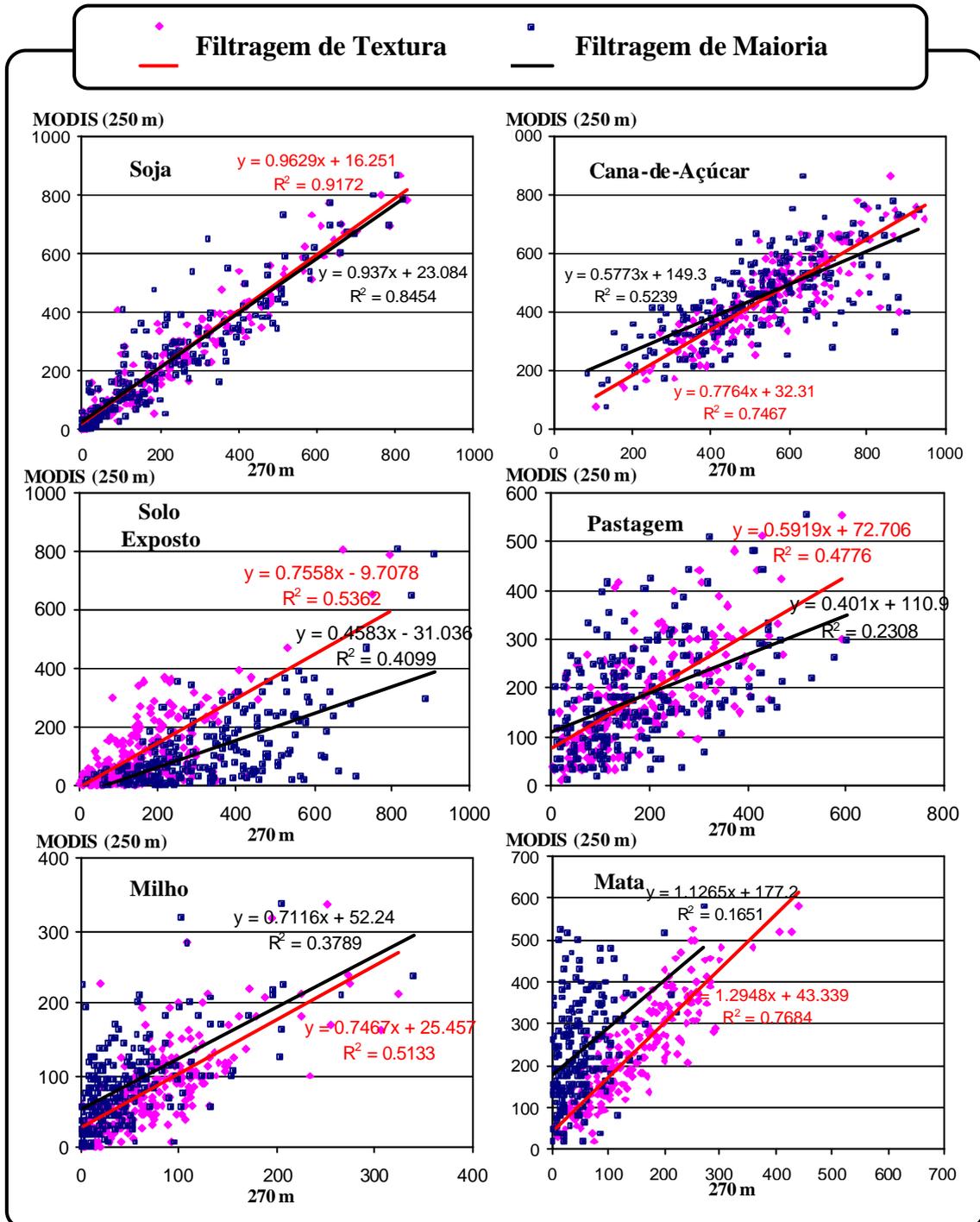


FIGURA 5.18 – Comparação entre a área (em hectares), apresentada pela imagem MODIS/Terra (250m) e aquela extraída das simulações da resolução espacial efetuadas por filtragem de textura e maioria (270m), por classe temática.

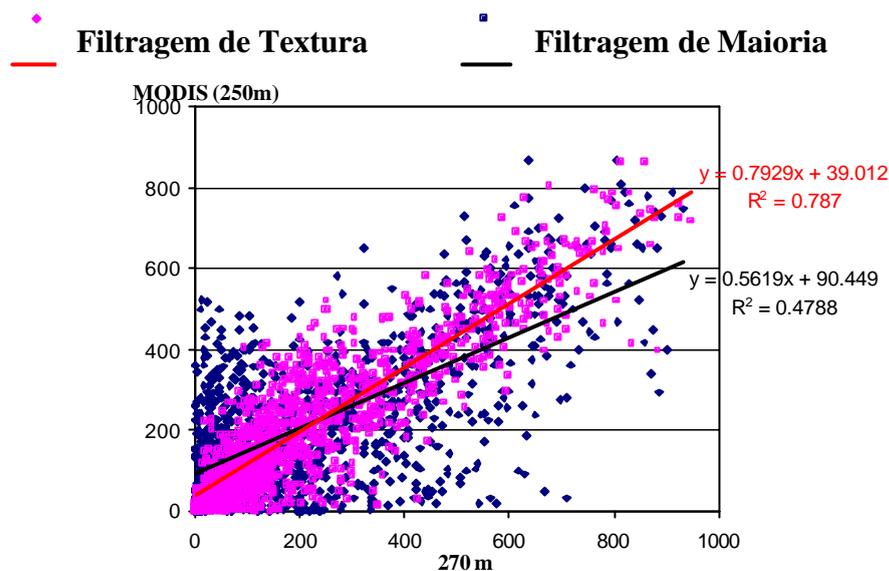


FIGURA 5.19 – Comparação entre a área (em hectares) apresentada pela imagem MODIS/Terra (250m) e os valores de áreas resultantes das simulações de resolução espacial efetuadas por filtragem de textura e maioria (270m), para todas as classes de cobertura analisadas.

A filtragem de maioria mostrou-se sempre mais estável e previsível do que a análise de textura. Este fenômeno se deve ao fato de a maioria ser um processo de redução do ruído aplicado como uma pós-classificação, sobre um mapa temático. A textura, por outro lado, é uma característica física que sugere a resolução espacial. Sua filtragem é implementada no espaço da imagem e gera um novo produto com mesmo número de bandas e com igual domínio ao do dado original, configurando-se, assim, procedimento mais complexo e responsável por uma resposta mais próxima de que uma imagem de um sensor poderia apresentar.

## 5.5 Confeção do Modelo Estatístico para Predição de Áreas Agrícolas do Sensor de Resolução Espacial Fina a partir da Moderada e de Métricas da Paisagem

### 5.5.1 Pré-Seleção das Métricas para a Modelagem Estatística com Base na Filtragem Espacial

Os resultados das simulações da resolução espacial moderada utilizando a técnica de filtragem de textura foram bastante positivos, pois contribuíram para um melhor

entendimento do comportamento dos polígonos ao longo de uma gama de valores de resolução espacial e da evolução das métricas com a degradação da resolução espacial.

Estes resultados também serviram de subsídio para a pré-seleção de métricas potencialmente úteis na composição de modelos de regressão múltipla, visando ao estudo da predição de áreas agrícolas de sensores de resolução espacial fina por dados de resolução espacial moderada. Esta pré-seleção foi feita com base nos seguintes critérios: a) Para índices que apresentaram equivalência entre gráficos de sua magnitude, em função da variação da janela de filtragem espacial, foi mantido apenas um representante dessa informação; b) De acordo com a proposição de Frohn (1998), uma métrica de paisagem eficiente para predição de área deve ser relativamente insensível ou previsivelmente sensível à variação da resolução espacial, bem como influenciada pelo padrão espacial associado às classes estudadas.

Com base nestes critérios, das 20 métricas consideradas neste estudo, foram pré-selecionadas 15, quais sejam: *Área*, *NP* (Número de Polígonos), *LPI* (Maior Área), *LSI* (Forma de Classe), *PAFRAC* (Dimensão Fractal Perímetro-Área), *TCA* (Área Nuclear Total), *NDCA* (Número de Áreas Nucleares Disjuntas), *CLUMP* (Agrupamento), *PLADJ* (Porcentagem de Adjacência), *IJI* (Interposição e Justaposição), *COHE* (Coesão de Área), *DIV* (Divisão de Cenário), *MESH* (Tamanho de Trama Efetiva), *SPLIT* (Fragmentação), *NLSI* (Forma de Classe Normalizada).

## **5.5.2 Análise Exploratória dos Dados**

### **5.5.2.1 Análise de normalidade da variável dependente**

A primeira análise dos dados consistiu da verificação de como era a distribuição da variável *Área* (ETM+), que será explicada pelo modelo. A Figura 5.20 expõe graficamente esse processo comparando os dados disponíveis com uma distribuição normal esperada.

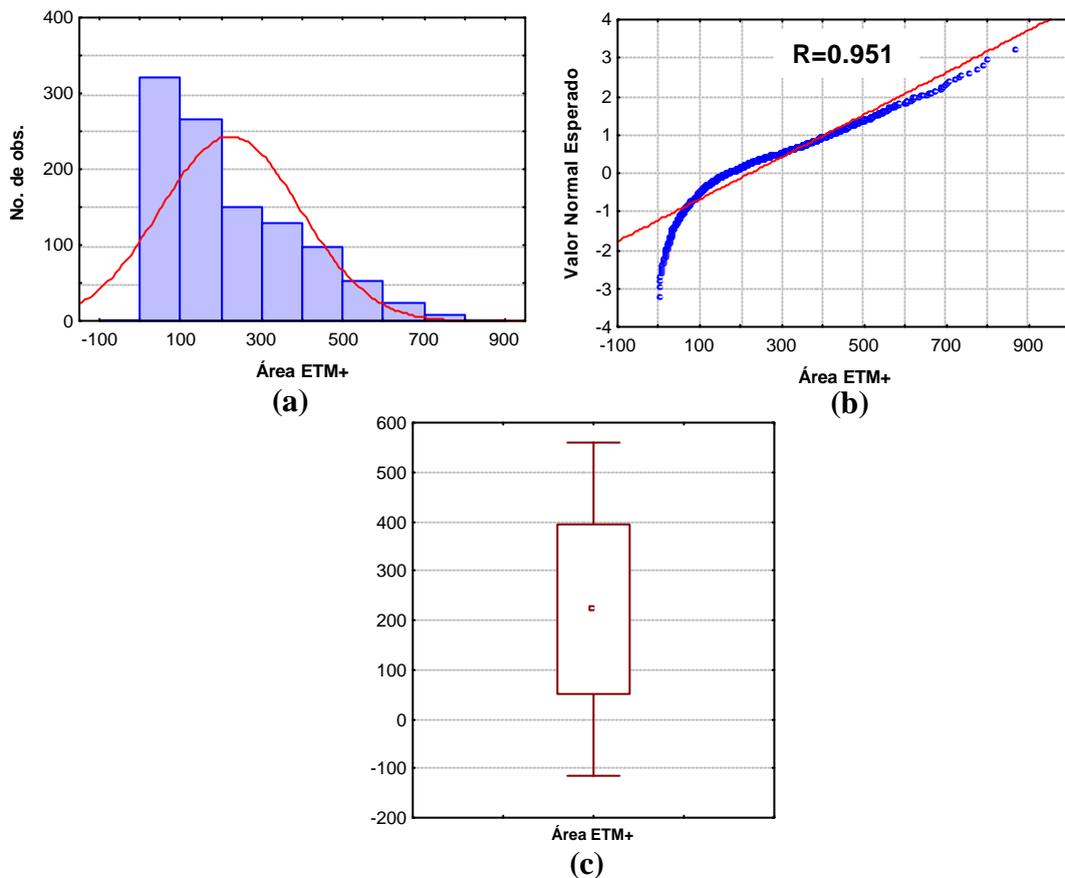


FIGURA 5.20 – Estudo da normalidade da variável dependente (Área ETM+): (a) Histograma; (b) Probabilidade normal; (c) Box-plot.

Da inspeção da Figura 5.20, pode-se notar a completa falta de normalidade da variável. O *box-plot* (Figura 5.20c) mostra uma boa distribuição dos dados tendo em vista o numeroso conjunto de observações. No entanto, o histograma de frequências (Figura 5.20a) constata a existência de um número muito maior de ocorrências de áreas pequenas. O diagrama de folhas (*stem-and-leaf*) gerado (não mostrado neste trabalho) corrobora a afirmação, bem como o gráfico de probabilidade normal, exposto na Figura 5.20b.

O teste de normalidade Shapiro-Wilk foi elaborado, resultando um  $W$  de 0,90 e um  $p$  praticamente nulo. O coeficiente de correlação ( $R$ ) do gráfico de probabilidade normal (Figura 5.20) foi bastante inferior ao valor crítico extraído da tabela organizada por Looney e Gullledge (1985), que é superior 0,987, para 1056 observações, a 5% de

significância. Portanto, definitivamente, não há normalidade para a variável Área ETM+.

Para normalização da variável, foram feitas diversas transformações, sendo a raiz quadrada a que forneceu resultados mais satisfatórios. A Figura 5.21 apresenta o histograma (Figura 5.21a), o gráfico de probabilidade normal (Figura 5.21b) e o *box-plot* (Figura 5.21c) da nova variável dependente transformada matematicamente pelo cálculo da raiz quadrada.

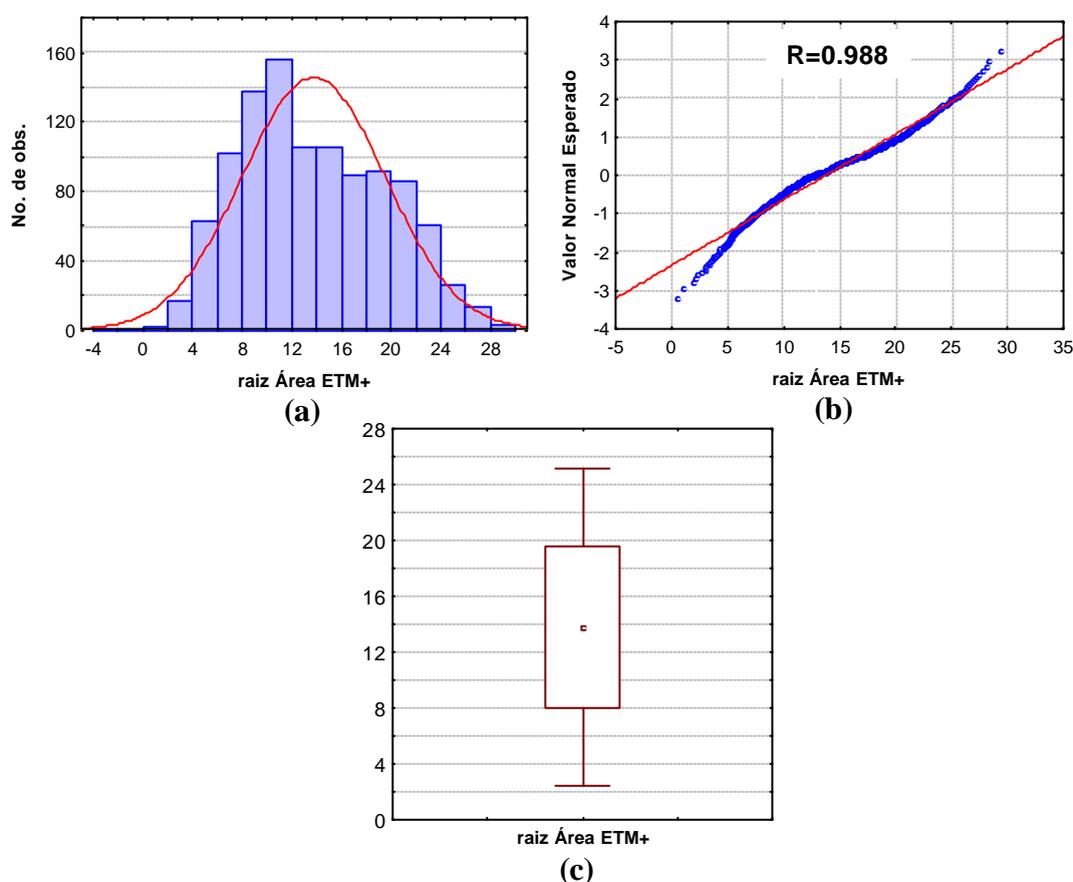


FIGURA 5.21 – Estudo da normalidade da variável dependente transformada (raiz de Área ETM+): (a) Histograma; (b) Probabilidade normal; (c) Box-plot.

Percebe-se, nitidamente, que os dados passaram a ter uma distribuição muito mais próxima da normal. O *box-plot* se manteve simétrico e o histograma se aproximou da curva da normal. No gráfico de probabilidade normal, os pontos se aproximaram da reta de tendência, o que é evidenciado pelo  $R$  maior (0,988). O teste Shapiro-Wilk realizado, entretanto, resultou um  $W$  de 0,98 e  $p$  praticamente nulo, o que refutaria a normalidade.

O elevado coeficiente de correlação do gráfico de probabilidade normal, bem como o elevadíssimo número de observações, todavia, permitiram aceitar a normalidade a níveis de significância inferiores a 6%.

### **5.5.2.2 Transformações das Variáveis Preditivas**

Para alcançar uma melhor relação dos índices com a variável dependente, foram realizadas transformações matemáticas nas variáveis preditivas. A Tabela 5.3 apresenta a matriz de correlação entre as variáveis, sendo possível observar, ainda, as operações realizadas. Note-se que as variáveis PLAND, PD, CPLAND, DCAD e AI não constam na Tabela por já terem sido descartadas pelas análises de filtragem.

Avaliando a matriz de correlação, pôde-se identificar algumas variáveis com forte correlação, o que sinalizou a existência de muita informação redundante. As variáveis *SPLIT* e *MESH*, por exemplo, exibiram correlação máxima, sendo esta a única que seguiu para estudo da composição do modelo final. De qualquer forma, é possível prever que algumas variáveis serão excluídas do modelo por não contribuírem com informação nova.

Uma observação importante foi que a transformação realizada para normalizar a variável dependente contribuiu para normalizar os resíduos das regressões simples. A Figura 5.22 ilustra essa ocorrência para a variável *Área* (MODIS), sendo importante ressaltar que tal fenômeno foi observado para todas as variáveis.

### **5.5.3 Constituição do Modelo de Regressão Simples**

Conforme fora dito anteriormente, a regressão simples entre as áreas obtidas pelas imagens MODIS e ETM+ foi considerada como ponto de partida, a partir da qual se deseja construir um modelo mais completo (regressão múltipla), que possa melhor caracterizar a variável dependente, através da incorporação de métricas da paisagem.

TABELA 5.3 – Matriz de correlação dos índices de padrão espacial.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<b>1. Área</b>	1.00	0.23	0.93	0.64	-0.10	0.44	0.51	0.16	0.72	0.28	0.86	-0.79	0.96	0.96	-0.44
<b>2. NP</b>		1.00	0.00	0.78	-0.72	-0.13	-0.15	-0.30	-0.24	0.26	-0.16	0.13	0.08	0.08	0.38
<b>3. raiz LPI</b>			1.00	0.47	0.06	0.45	0.54	0.31	0.82	0.24	0.97	-0.83	0.99	0.99	-0.58
<b>4. LSI</b>				1.00	-0.47	-0.03	0.01	-0.20	0.08	0.36	0.35	-0.25	0.53	0.53	0.15
<b>5. PAFRAC</b>					1.00	0.08	0.11	0.19	0.19	-0.11	0.16	-0.11	0.01	0.01	-0.23
<b>6. TCA</b>						1.00	0.76	0.16	0.47	-0.03	0.40	-0.63	0.44	0.44	-0.28
<b>7. NDCA</b>							1.00	0.18	0.54	0.02	0.50	-0.69	0.53	0.53	-0.34
<b>8. CLUMP</b>								1.00	0.57	0.02	0.35	-0.15	0.29	0.29	-0.85
<b>9. PLADJ^2</b>									1.00	0.11	0.86	-0.64	0.82	0.82	-0.83
<b>10. IJI</b>										1.00	0.20	-0.14	0.26	0.26	-0.02
<b>11. COHE^4</b>											1.00	-0.78	0.96	0.96	-0.63
<b>12. DIV</b>												1.00	-0.82	-0.82	0.39
<b>13. raiz 4 MESH</b>													1.00	1.00	-0.56
<b>14. 1/raiz4 SPLIT</b>														1.00	-0.56
<b>15. NLSI</b>															1.00

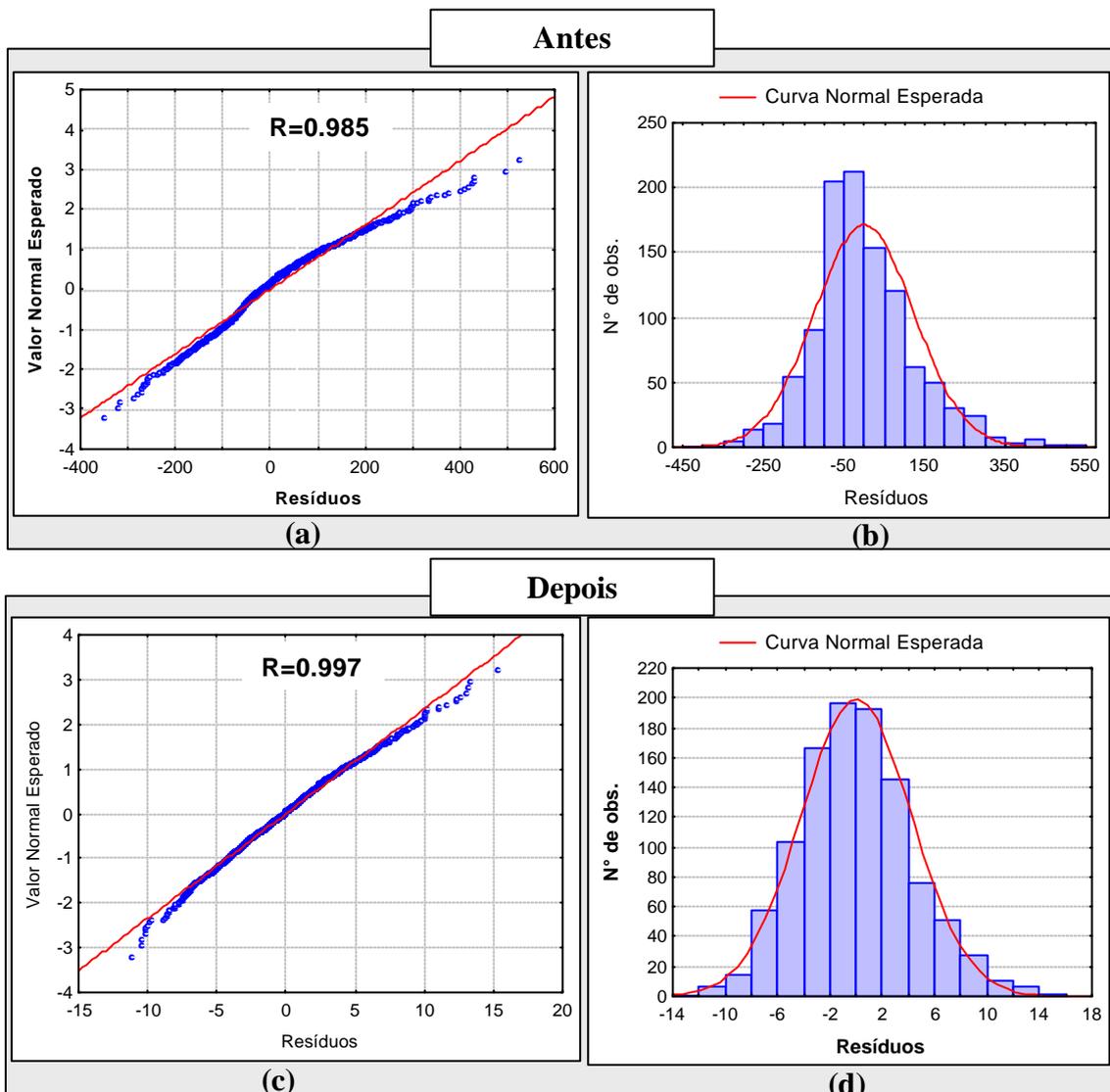


FIGURA 5.22 – Normalização dos resíduos brutos após transformação da variável dependente.

### 5.5.3.1 Sumário da Regressão Simples

A Tabela 5.4 apresenta os parâmetros estimados para a regressão simples, com coluna *p-valor* denotando a significância da relação. A análise da variância (ANOVA) é exibida pela Tabela 5.5 e indica que a variável preditiva está contribuindo para explicação da dependente (Área ETM+), já que o valor de *F* é estatisticamente significativo ao nível de confiança de 95%.

TABELA 5.4 – Parâmetros da regressão simples.

	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	8,911180	0,207870	0,00
<b>Área</b>	0,022611	0,000754	0,00

TABELA 5.5 – Análise da variância (ANOVA).

	<b>Soma de Quadrados</b>	<b>df</b>	<b>Médios</b>	<b>F</b>
<b>Regressão</b>	16269,99	1	16269,99	899,9327
<b>Resíduo</b>	19055,39	1054	18,08	
<b>Total</b>	35325,38			

Portanto a Relação entre as variáveis fica:

$$\hat{Y} = 8,91 + 0,02 \cdot X_1 \quad (5.1)$$

Onde:

$\hat{Y}$  é o valor estimado da raiz de Área ETM+;

$X_1$  é Área MODIS.

Essa regressão apresentou um coeficiente de determinação ( $R^2$ ) de 0,46.

### 5.5.3.2 Análise de Variância

Pela análise da Figura 5.23, é possível observar uma tendência em relação aos resíduos da regressão simples. As linhas verdes indicam que ocorreram maiores variâncias para menores valores da variável dependente. O teste modificado de *Levene* acabou por confirmar, estatisticamente, a constatação, com um *t crítico* de 1,97 e um  $|t^*|$  de 2,56. Como o critério para aceitar a hipótese de heterocedasticidade é  $|t^*| > t \text{ crítico}$ , concluiu-se que a variância dos resíduos não é constante.

Para correção da tendenciosidade da variância, o método dos mínimos quadrados (MMQ) foi substituído por sua ponderação, dando origem aos parâmetros contidos na Tabela 5.6.

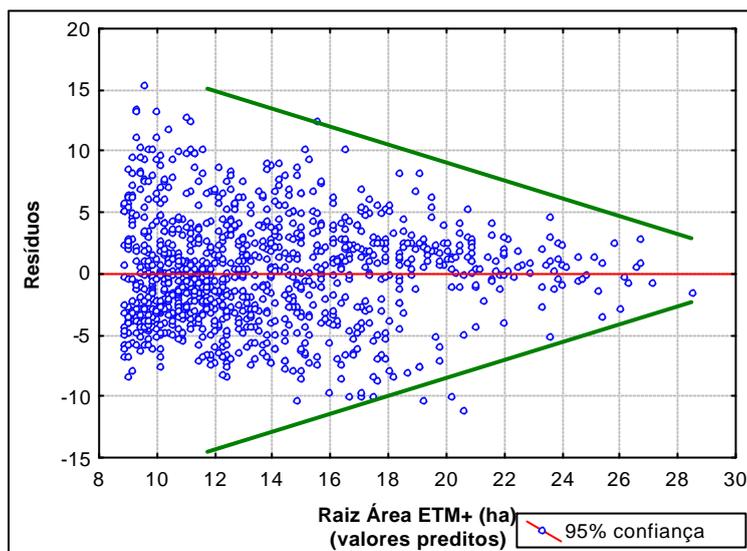


FIGURA 5.23 – Relação dos resíduos brutos com variável dependente.

TABELA 5.6 – Parâmetros da regressão simples ponderada.

	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	8,864927	0,069789	0,00
<b>Área</b>	0,022791	0,000213	0,00

Conforme se pode observar ao comparar as Tabelas 5.4 e 5.6, os parâmetros da regressão sofreram alterações insignificantes. Contudo, os intervalos de confiança de suas estimativas se tornaram bastante mais estreitos, conforme visto na Tabela 5.7. Estes resultados se alinharam aos apresentados por Neter (1996).

TABELA 5.7 – Intervalos de confiança da estimativa de parâmetros.

	<b>MMQ</b>		<b>MMQ ponderado</b>	
	<b>-95,00%</b>	<b>+95,00%</b>	<b>-95,00%</b>	<b>+95,00%</b>
<b>Intercepto</b>	8,503293	9,319068	8,728126	9,001728
<b>Área</b>	0,021132	0,024090	0,022373	0,023209

### 5.5.3.3 Análise de Normalidade

O teste de Shapiro-Wilk realizado rejeitou a hipótese de normalidade, uma vez que resultou um  $p=0,00012$  para um  $W=0,993$ . O gráfico de probabilidade normal, porém, revela uma forte tendência à normalidade (Figura 5.24), com um coeficiente de correlação muito próximo da unidade. A Figura 5.25 traz, finalmente, o histograma da distribuição dos resíduos brutos, cuja proximidade da curva normal esperada é nítida.

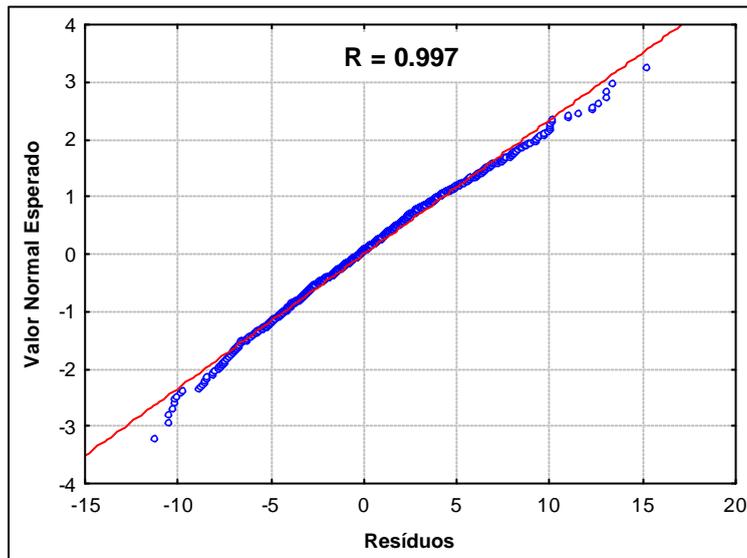


FIGURA 5.24 – Gráfico de probabilidade normal dos resíduos brutos.

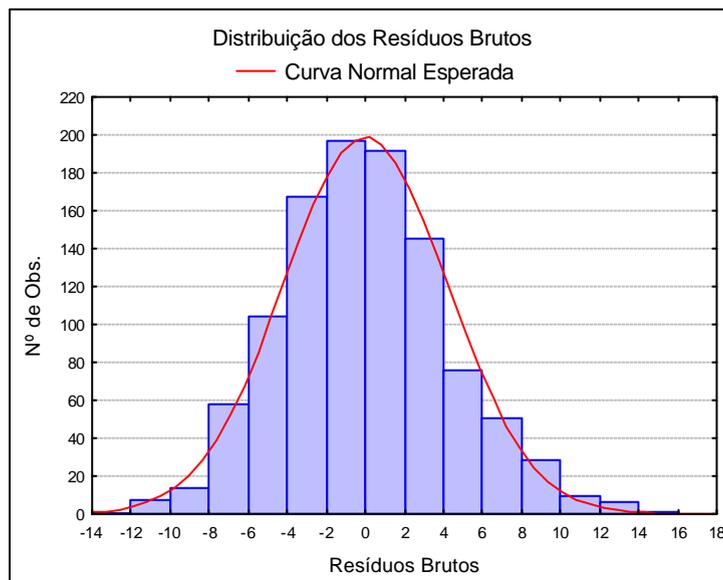


FIGURA 5.25 – Histograma de frequência dos resíduos brutos.

As análises apresentadas permitiram desconsiderar a não normalidade expressa pelo teste de *Shapiro-Wilk*, já que o tamanho da amostra é bastante elevado, além de os valores de *student residuals*, que são resíduos ponderados por seus desvios padrões, serem muito baixos (Neter, 1996).

#### 5.5.3.4 Identificação de *Outliers* Influentes

Para identificação de *outliers* influentes, foi calculado o *DFFITs*, não tendo sido identificado nenhum caso, uma vez que o maior valor absoluto de *DFFITs* encontrado ficou na casa dos 4 centésimos, enquanto que o crítico é 0,09.

#### 5.5.3.5 Validação da Regressão Simples

O primeiro passo para estudar a validade do modelo de regressão simples foi o cálculo do MSPR, cujo valor atingiu 18,40. A observação de Tabela 5.5 permitiu a conclusão de que o Quadrado Médio dos Resíduos (*MSE*), que é de 18,08, é muito semelhante ao *MSPR*. Portanto, concluiu-se ser válido o modelo estatístico gerado.

Para trazer mais robustez ao processo de validação foi procedido o Ajuste à Reta Diagonal (ARD), cujos testes de hipóteses são apresentados a seguir.

A Tabela 5.8 apresenta os parâmetros estimados da regressão para o conjunto de validação, bem como o *p-valor*, através do qual já é possível afirmar que o intercepto é estatisticamente nulo, uma vez que ele é maior que o nível de significância (5%).

TABELA 5.8 – Parâmetros da regressão da amostra de validação.

	<b>B</b>	<b>p-valor</b>
<b>Intercepto</b>	1,691249	0,079559
<b>raiz Área ETM+ (<math>\hat{Y}</math>)</b>	0,872881	0,000000

Foi procedido, então, um teste de hipótese para avaliar se  $b_1=1$ , resultando um valor absoluto para  $t^*$  de 1.92. Como  $t(0,95; 238) = 1.97$ , a hipótese nula é aceita, garantindo, definitivamente, a validade do modelo de regressão simples.

#### 5.5.4 Constituição do Modelo de Regressão Múltipla

Nesta seção é feita uma análise de regressão múltipla, com o objetivo de explicar a variável raiz Área ETM+ pelas 14 variáveis remanescentes. A regressão com todas elas é exposta na Tabela 5.9, sendo apresentadas em destaque vermelho as variáveis estatisticamente significativas.

TABELA 5.9 – Parâmetros da regressão múltipla.

	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	7,83887	1,058564	0,000000
<b>Área</b>	0,02992	0,005305	0,000000
<b>NP</b>	-0,12425	0,144349	0,389577
<b>raiz LPI</b>	-0,55370	1,228785	0,652370
<b>LSI</b>	-1,33288	0,642670	0,038333
<b>PAFRAC</b>	-0,00819	0,005539	0,139309
<b>TCA</b>	-0,04657	0,023070	0,043771
<b>NDCA</b>	-1,03413	0,607512	0,089016
<b>CLUMPY</b>	0,56421	1,114136	0,612678
<b>PLADJ^2</b>	0,00077	0,000472	0,105168
<b>IJI</b>	0,02385	0,009876	0,015926
<b>COHE^4</b>	-0,00000	0,000000	0,413203
<b>DIV</b>	1,67216	2,108063	0,427835
<b>raiz 4 MESH</b>	0,85505	3,585782	0,811576
<b>NLSI</b>	1,89785	2,428501	0,434696

O teste  $p$  para todas as variáveis do modelo (Tabela 5.9) indicou que apenas as variáveis *Área*, *LSI*, *TCA* e *IJI* foram significativas. Logo, foi necessário realizar um teste de hipótese para constatar a razoabilidade estatística da eliminação de todas as outras variáveis, concomitantemente.

A Tabela 5.10 deu suporte à avaliação, tendo sido organizada pelo cálculo da Soma dos Quadrados dos Resíduos (SSE) para o modelo com as 14 variáveis (completo) e do modelo com apenas as quatro variáveis significativas (reduzido).

TABELA 5.10 – Soma do quadrado dos resíduos (SSE).

	<i>SSE</i>	df
Completo	17321,25	1020
Reduzido	18126,89	1051

Com um  $F^* = 1,530376$ , superior, portanto, ao  $F(0,95;31;1020) = 1,462497$ , concluiu-se pela rejeição da hipótese de nulidade de todos os parâmetros. Será procedida, assim, uma busca por todas as regressões possíveis, para analisar quais variáveis devem entrar no modelo final.

A Figura 5.26 exibe o gráfico do critério  $C_p$  de seleção de variáveis. O conjunto selecionado é composto por seis elementos relativos aos índices *Área*, *raiz LPI*, *LSI*, *TCA*, *PLADJ^2* e *IJI*.

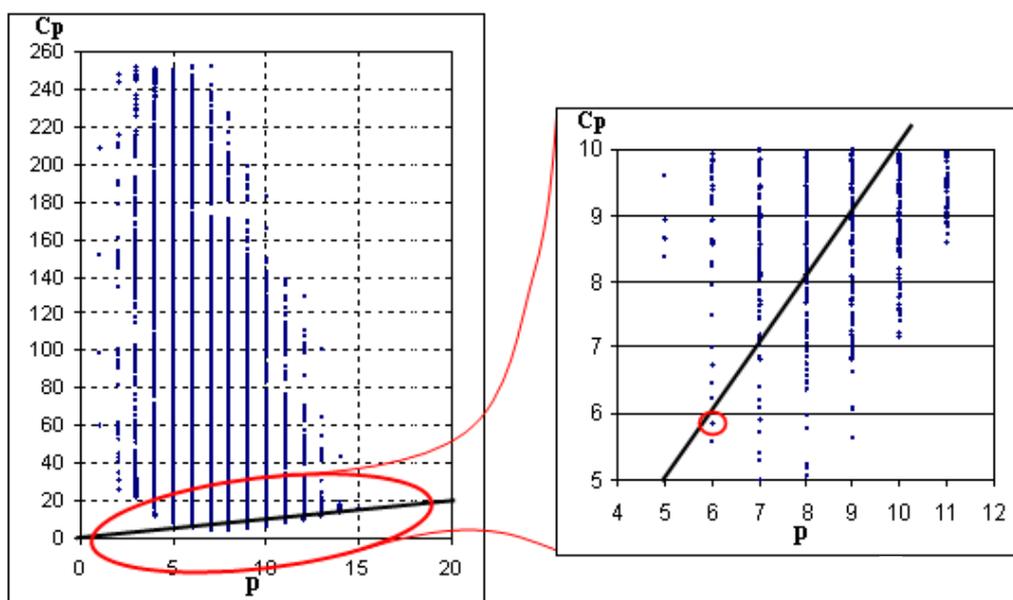


FIGURA 5.26 – Seleção dos índices espaciais: Critério  $C_p$ .

As variáveis do modelo definido pelo critério  $C_p$  encontram-se na Tabela 5.11, cuja coluna do *p-valor* confirma a significância de todos os parâmetros. A Tabela 5.12 exibe, por sua vez, a ANOVA associada.

TABELA 5.11 – Parâmetros do modelo elaborado pelo critério  $C_p$ .

	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	9,159645	0,700634	0,000000
<b>Área</b>	0,027381	0,002560	0,000000
<b>raiz LPI</b>	-0,637525	0,299863	0,033737
<b>LSI</b>	-0,985063	0,263209	0,000192
<b>TCA</b>	-0,056517	0,017221	0,001066
<b>PLADJ^2</b>	0,000731	0,000222	0,001039
<b>IJI</b>	0,029780	0,009104	0,001107

Essa regressão apresentou um coeficiente de determinação ( $R^2$ ) de 0,49.

Os critérios  $R_p^2$ ,  $R_a^2$  também foram calculados e constam, respectivamente, das Figuras 5.27 e 5.28. Ambos os critérios selecionaram nove variáveis idênticas cujos parâmetros

são revelados pela Tabela 5.13, que exibe os parâmetros significativos, a 95% de confiança, destacados na cor vermelha.

TABELA 5.12 – ANOVA do modelo elaborado pelo critério  $C_p$ .

	Soma de Quadrados	df	Médios	F
<b>Regressão</b>	16850,10	6	2808,350	165,3767
<b>Resíduos</b>	17457,02	1028	16,982	
<b>Total</b>	34307,12			

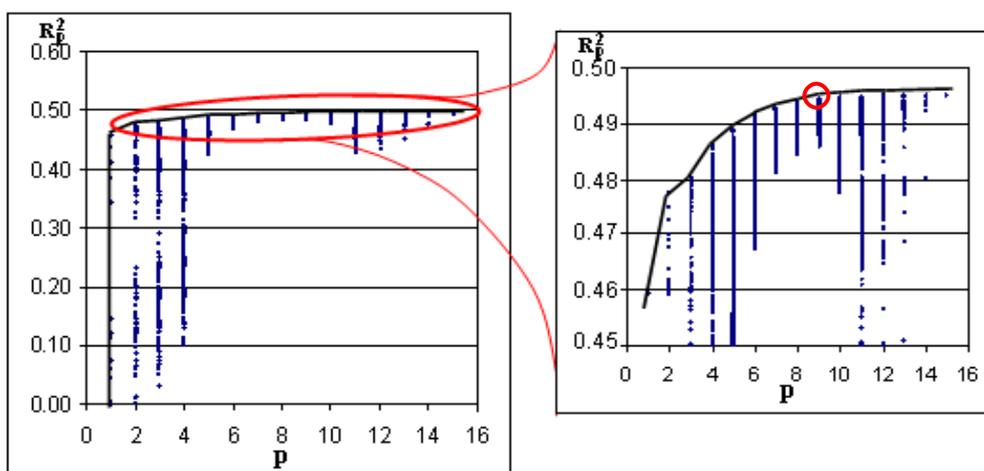


FIGURA 5.27 – Seleção dos índices espaciais: Critério  $R_p^2$ .

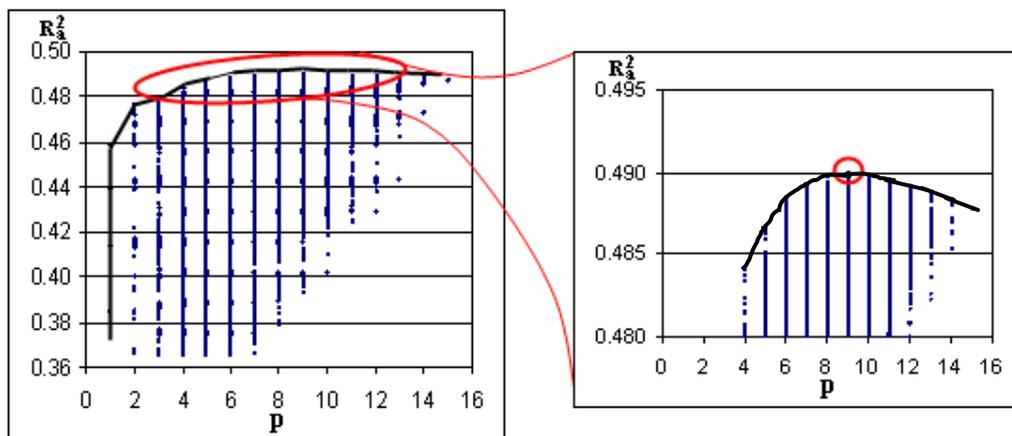


FIGURA 5.28 – Seleção dos índices espaciais: Critério  $R_a^2$ .

TABELA 5.13 – Parâmetros do modelo elaborado pelos critérios  $R_p^2$  e  $R_a^2$ , com dados estatisticamente significantes indicados em vermelho.

	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	7,91217	1,054636	0,000000
<b>Área</b>	0,03070	0,002997	0,000000
<b>raiz LPI</b>	-0,53155	0,314904	0,091720
<b>LSI</b>	-1,52597	0,350330	0,000015
<b>PAFRAC</b>	-0,00527	0,004389	0,229745
<b>TCA</b>	-0,04046	0,022062	0,066962
<b>NDCA</b>	-0,90451	0,589835	0,125462
<b>PLADJ^2</b>	0,00052	0,000244	0,033859
<b>IJI</b>	0,02468	0,009744	0,011453
<b>DIV</b>	3,16776	1,567172	0,043506

Conforme se pode observar pela coluna p-valor, nem todos os parâmetros são estatisticamente significativos. Dessa forma, houve a necessidade de avaliar quais variáveis poderiam ser eliminadas para composição do modelo final.

Tendo em vista que o critério  $C_p$  já selecionara um grupo de variáveis, que está contido no conjunto obtido por  $R_p^2$  e  $R_a^2$ , foi efetuado um teste para verificar se as três variáveis que não estão presentes no modelo de  $C_p$  (*PAFRAC*, *NDCA*, *DIV*) poderiam ser descartadas. A Tabela 5.14 suporta a avaliação.

TABELA 5.14 – Soma do quadrado dos resíduos (*SSE*).

	<i>SSE</i>	df
Completo	17348,68	1025
Reduzido	17457,02	1028

Portanto,  $F^* = 2,133595$ , o que é inferior a  $F(0,95; 3; 1025) = 2,613586$ , permitindo a exclusão das variáveis *PAFRAC*, *NDCA*, *DIV*.

Finalizando a seleção de variáveis, pôde-se concluir que os 3 critérios convergiram para o mesmo conjunto de variáveis, cujos parâmetros encontram-se expostos na Tabela 5.11.

As análises realizadas para a regressão simples, quais sejam análise de normalidade e variância, identificação de *outliers* influentes e, por fim, a validação do modelo foram executadas e serão apresentadas a seguir.

#### 5.5.4.1 Análise de Variância

A Figura 5.29 indica a existência de tendência em relação aos resíduos da regressão. As linhas verdes indicam que ocorreram maiores variâncias para menores valores da variável dependente. O teste modificado de *Levene*, realizado sobre a variável *Área*, corroborou a análise visual, assumindo um *t crítico* de 1,97 e um  $|t^*|$  de 2,51. Portanto, concluiu-se pela heterocedasticidade dos dados.

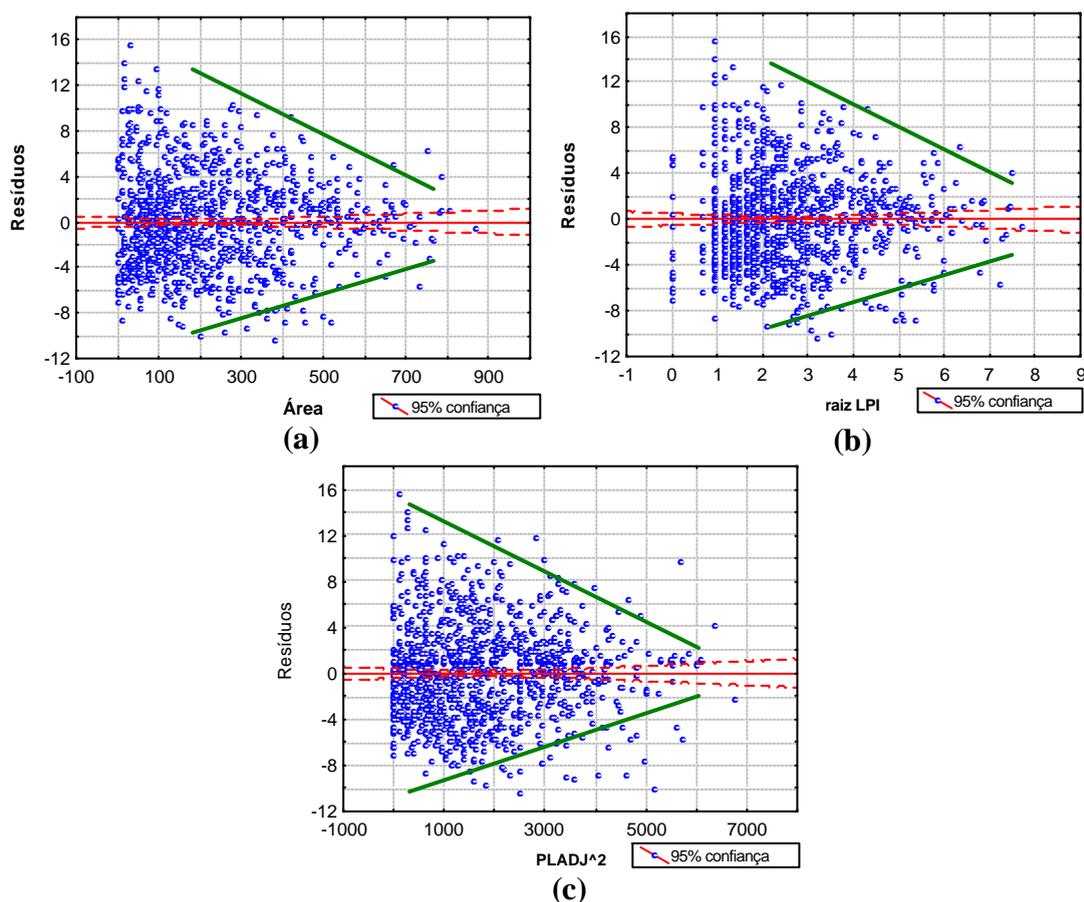


FIGURA 5.29 – Relação dos resíduos brutos com as variáveis independentes *Área*, *raiz LPI*, *PLADJ*<sup>2</sup>.

Analogamente ao procedido para a regressão simples, foi criado um modelo baseado no MMQ ponderado, cujos parâmetros são apresentados pela Tabela 5.15.

TABELA 5.15 – Parâmetros da regressão múltipla ponderada.

	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	8,885800	0,236931	0,000000
<b>Área</b>	0,027465	0,000741	0,000000
<b>raiz LPI</b>	-0,681347	0,084055	0,000000
<b>LSI</b>	-0,970695	0,080811	0,000000
<b>TCA</b>	-0,046723	0,004147	0,000000
<b>PLADJ^2</b>	0,000708	0,000068	0,000000
<b>IJI</b>	0,034778	0,002908	0,000000

As Tabelas 5.11 e 5.15 apresentam parâmetros da regressão muito próximos. Entretanto, conforme já mencionado, a Tabela 5.16 mostra que os intervalos de confiança de suas estimativas tornaram-se muito mais rígidos, o que traz mais segurança à estimativa realizada.

TABELA 5.16 – Intervalos de confiança da estimativa de parâmetros.

	<b>MMQ</b>		<b>MMQ ponderado</b>	
	<b>-95,00%</b>	<b>+95,00%</b>	<b>-95,00%</b>	<b>+95,00%</b>
<b>Intercepto</b>	7,78481	10,53448	8,42137	9,350231
<b>Área</b>	0,02236	0,03240	0,02601	0,028916
<b>raiz LPI</b>	-1,22594	-0,04911	-0,84611	-0,516583
<b>LSI</b>	-1,50155	-0,46858	-1,12910	-0,812290
<b>TCA</b>	-0,09031	-0,02272	-0,05485	-0,038593
<b>PLADJ^2</b>	0,00029	0,00117	0,00058	0,000840
<b>IJI</b>	0,01192	0,04764	0,02908	0,040479

#### 5.5.4.2 Análise de Normalidade

Foi realizado o teste de Shapiro-Wilk para análise da normalidade dos resíduos, o qual resultou um  $p$  praticamente nulo e um  $W$  de 0,990, rejeitando, portanto a hipótese de normalidade. A Figura 5.30 traz, porém, o gráfico de probabilidade normal (Figura 5.30a), o histograma (Figura 5.30b) e o *box-plot* (Figura 5.30c), que denotam a boa proximidade da distribuição estatística dos resíduos com a curva normal esperada.

O enorme tamanho da amostra, novamente, proporcionou um valor crítico de  $W$  exacerbado, o que é evidenciado pelo  $p$ -valor nulo. A Figura 5.30a, todavia, exibe um altíssimo valor de coeficiente de correlação, além do cálculo dos valores de *student*

*residuals* ter mostrado a dimensão reduzida deste resíduo. Deste modo, não foi tomada nenhuma providência para correção de uma eventual não normalidade dos resíduos.

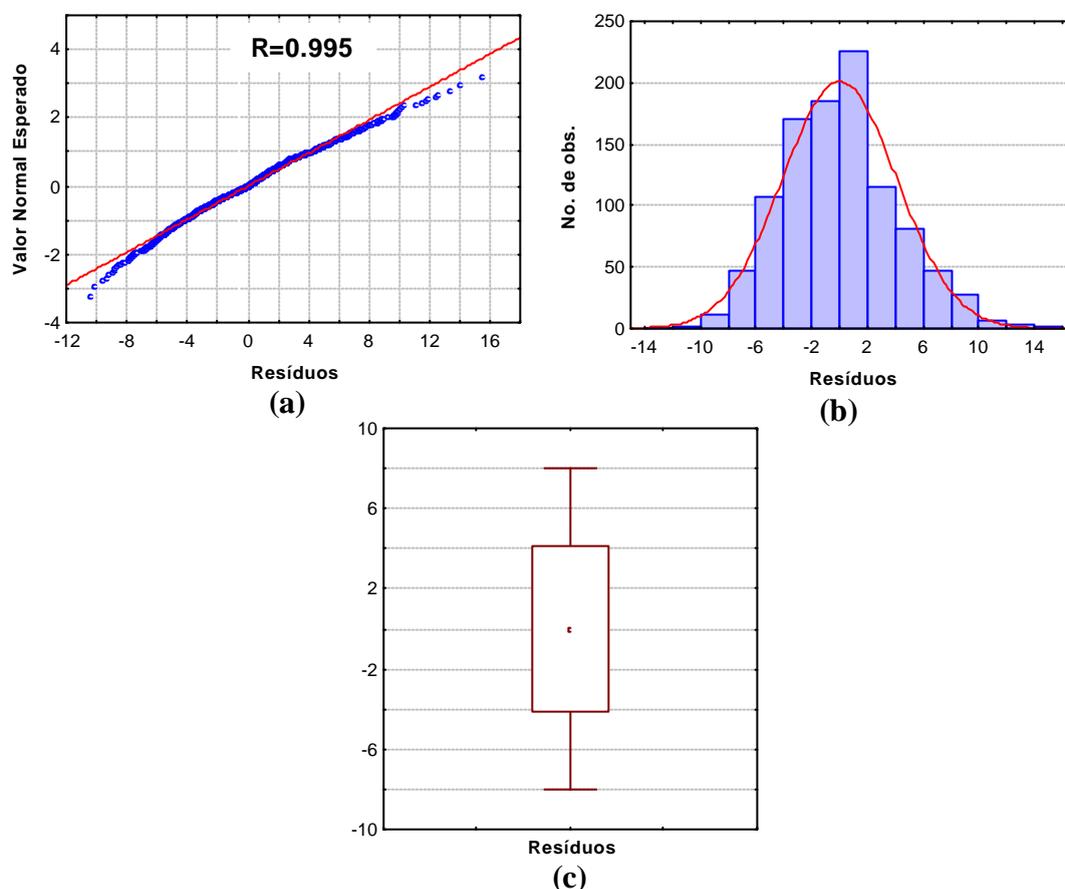


FIGURA 5.30 – Análise da normalidade dos resíduos: a) Gráfico de probabilidade normal; b) Histograma; c) *Box-plot*.

#### 5.5.4.3 Identificação de *Outliers* Influentes

O cálculo de *DFFITs*, realizado para a regressão múltipla, não detectou a existência de *outliers* influentes que pudessem prejudicar a performance do modelo estatístico.

#### 5.5.4.4 Efeito de Multicolinearidade

O cálculo do *VIF* é consolidado na Tabela 5.17 e indica a existência de multicolinearidade no modelo proposto. Note-se, entretanto, que apenas duas variáveis superaram, moderadamente, o valor limítrofe de 10. O *VIF* médio foi além da unidade,

mas não chegou a atingir um valor descontrolado. Dessa forma, nenhuma medida foi utilizada para remediar os efeitos da correlação entre as variáveis.

TABELA 5.17 – VIF do modelo de regressão múltipla.

	CA	raiz LPI	LSI	TCA	PLADJ^2	IJI
$VIF_k$	11,92	10,95	4,05	1,53	5,17	1,39
$\bar{VIF}_k$	7,00					

#### 5.5.4.5 Validação do Modelo de Regressão

O valor de MSPR obtido foi de 14,12, o que é muito próximo do valor do Quadrado Médio dos Resíduos (MSE), que é de 16,98 (Tabela 5.12). Como o critério para que o modelo seja adequado passa por um MSPR próximo ao MSE do modelo correspondente, conclui-se que há um forte indicativo da adequação do modelo.

Para o teste de Ajuste à Reta Diagonal (ARD), é apresentada a Tabela 5.18 com os parâmetros estimados da regressão para o conjunto de validação, bem como o *p-valor*, pelo qual já é possível afirmar que o intercepto é estatisticamente nulo, uma vez que ele é maior que o nível de significância (5%).

TABELA 5.18 – Parâmetros da regressão da amostra de validação.

	B	p-valor
<b>Intercepto</b>	0,963051	0,312153
<b>raiz ETM+ (<math>\hat{Y}</math>)</b>	0,912509	0,000000

O teste de hipótese para avaliar se  $b_1=1$  resultou um valor absoluto para  $t^*$  de 1,36. Como  $t(0,95; 238) = 1,97$ , aceitou-se a hipótese nula, avaliando, categoricamente, a validade do modelo de regressão múltipla.

Num processo lógico de composição final de um modelo de regressão, o próximo passo seria a incorporação das observações do conjunto de validação à base de dados e, assim, reestimar os parâmetros definitivos. Entretanto, tendo em vista a enorme quantidade de observações disponíveis para composição do modelo, esta fase foi negligenciada, sendo estabelecidos os parâmetros constituintes da Tabela 5.15 como finais.

## 5.6 Análise Comparativa entre a Abordagem Geral versus a Específica Utilizando Modelos de Regressão Simples e Múltipla

Com a abordagem geral (todas as classes consideradas em conjunto), a regressão simples apresentou um coeficiente de determinação ( $R^2$ ) de 0,46, enquanto que a regressão múltipla mostrou um valor de  $R^2$  de 0,49. Dessa forma, a inclusão dos índices que denotam o padrão espacial das classes temáticas aumentou o potencial de explicação, pelo sensor MODIS, da área obtida pela imagem ETM+.

Este incremento de cerca de 7% não chegou a impressionar, mas deu indício de que, pela inclusão da informação de padrão espacial das classes temáticas, há possibilidade de melhoria no processo de estimativa de área com sensores de resolução espacial moderada.

Em relação à abordagem específica (modelos por classe), as Tabelas 5.19, 5.20 e 5.21 apresentam os parâmetros das regressões simples, respectivamente, para as classes agrícolas milho, soja e cana-de-açúcar. As Tabelas 5.22, 5.23 e 5.24, por sua vez, trazem, na mesma ordem, os parâmetros das regressões múltiplas, para as mesmas classes agrícolas.

TABELA 5.19 – Parâmetros da regressão simples específica para o milho.

<b>Milho</b>	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	6,629388	0,280199	0,000000
<b>Área</b>	0,033757	0,002609	0,000000

TABELA 5.20 – Parâmetros da regressão simples específica para a soja.

<b>Soja</b>	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	6,142465	0,300568	0,000000
<b>Área</b>	0,028905	0,001042	0,000000

TABELA 5.21 – Parâmetros da regressão simples específica para a cana-de-açúcar.

<b>Cana</b>	<b>B</b>	<b>Desv. Pad.</b>	<b>p-valor</b>
<b>Intercepto</b>	12,58117	0,567008	0,000000
<b>Área</b>	0,01834	0,001199	0,000000

TABELA 5.22 – Parâmetros da regressão múltipla específica para o milho.

Milho	B	Desv. Pad.	p-valor
<b>Intercepto</b>	4,4179	0,7084	0,0000
<b>NP</b>	0,2555	0,0754	0,0009
<b>raiz LPI</b>	-3,1221	1,5268	0,0424
<b>CLUMPY</b>	0,0255	0,0092	0,0062
<b>IJI</b>	-0,0254	0,0097	0,0096
<b>raiz 4 MESH</b>	9,4358	2,6824	0,0006

TABELA 5.23 – Parâmetros da regressão múltipla específica para a soja.

Soja	B	Desv. Pad.	p-valor
<b>Intercepto</b>	3,3741E+00	4,3846E-01	0,0000
<b>Área</b>	1,9741E-02	2,8739E-03	0,0000
<b>NP</b>	7,3158E-01	1,0536E-01	0,0000
<b>NDCA</b>	-1,8665E+03	9,0471E+02	0,0406
<b>DCAD</b>	2,6122E+04	1,2665E+04	0,0407
<b>COHESION^4</b>	7,1136E-08	1,8484E-08	0,0002

TABELA 5.24 – Parâmetros da regressão múltipla específica para a cana-de-açúcar.

Cana	B	Desv. Pad.	p-valor
<b>Intercepto</b>	-38,8831	15,83294	0,015062
<b>Área</b>	0,0337	0,00619	0,000000
<b>LSI</b>	3,9220	0,87243	0,000013
<b>CLUMPY</b>	43,4461	12,96585	0,000993
<b>IJI</b>	-0,0748	0,01637	0,000009
<b>NLSI</b>	44,5468	15,27085	0,004010

De forma a permitir um estudo comparativo entre os resultados obtidos com as abordagens geral (todas as classes temáticas compondo o modelo) e específicas (um modelo para cada cultura agrícola), a Tabela 5.25 exhibe os coeficientes de determinação ( $R^2$ ) para as regressões simples e múltiplas de ambas as abordagens.

TABELA 5.25 – Coeficientes de determinação ( $R^2$ ) para as regressões simples e múltiplas com abordagem geral e específica.

	Regressão Simples		Regressão Múltipla	
	Geral	Específica	Geral	Específica
<b>Milho</b>	0,46	0,49	0,49	0,52
<b>Soja</b>		0,81		0,87
<b>Cana</b>		0,57		0,67

A Tabela 5.25 mostra que sempre a modelagem específica é mais eficiente, inclusive para a cultura agrícola mais fragmentada, que é o milho. Para a classe mais compacta, que é a soja, o incremento chegou a cerca de 80%, para regressão múltipla, utilizando a abordagem específica, quando comparada à regressão múltipla com abordagem geral. Quando se confronta a regressão múltipla específica com a regressão simples geral obtém-se aumento no  $R^2$  de 13%, 90%, 46%, respectivamente, para o milho, a soja e a cana.



## CAPÍTULO 6

### CONCLUSÕES E RECOMENDAÇÕES

#### 6.1 Aspectos Conclusivos

Os resultados obtidos neste trabalho demonstram que sensores de resolução espacial moderada (p. ex., MODIS/Terra; 250m) podem ser utilizados para prever áreas vistas por sensores de resolução mais fina (ETM+/Landsat-7; 30m), para classes temáticas que não apresentem fragmentação acentuada, como soja e cana-de-açúcar na área de estudo. Essa habilidade potencializa-se com a adoção de modelos estatísticos específicos para as diferentes culturas agrícolas e que incorporem métricas de paisagem.

Sinteticamente, as seguintes conclusões foram obtidas no trabalho:

- 1) As filtragens espaciais de maioria e textura apresentaram, de modo geral, comportamento compatível entre si. Entretanto, a filtragem de textura mostrou-se mais realística. Sua estimativa de área com janela  $9 \times 9$  *pixels* (270m), comparada aos dados originais do sensor MODIS (250m), produziu uma regressão com coeficiente de determinação ( $R^2$ ) de 0,79, enquanto que a filtragem de maioria mostrou um valor de 0,48;
- 2) Das 20 métricas de paisagem consideradas na análise, foram selecionadas 14 índices para modelagem estatística da estimativa de área, com base em sua evolução com a degradação da resolução espacial por filtragem e em critérios estatísticos.
- 3) Com a abordagem geral (todas as classes temáticas em conjunto), para o modelo de regressão simples dos dados originais, verificou-se a existência de uma relação de regressão limitada, com coeficiente de determinação de 0,46, entre as estimativas de áreas obtidas com as imagem ETM+ (30m) e MODIS (250m). A adoção de um modelo de regressão múltipla, com a incorporação das seis métricas da paisagem denotando o padrão espacial das culturas, aumentou o potencial de explicação, pelo sensor MODIS,

da área obtida pela imagem ETM+. Tal conclusão vem da comparação do coeficiente de determinação da regressão simples (0,46) com o da múltipla (0,49). Este incremento de cerca de 7% não chega a impressionar, mas dá indício de que há possibilidade de melhoria no processo de estimativa de área agrícola com sensores de resolução espacial moderada e a inclusão de métricas da paisagem;

4) Os melhores resultados para estimativa de área foram obtidos com a adoção de modelos de regressão múltipla e de abordagens específicas (por cultura), inclusive para a cultura agrícola mais fragmentada (milho). Os coeficientes de determinação ( $R^2$ ) obtidos de 0,52, 0,67 e 0,87 para o milho, a cana e a soja, respectivamente, com a utilização de regressão múltipla e de índices espaciais, foram superiores ao valor de 0,49 atingido pela abordagem geral equivalente. Quando comparados com o valor de  $R^2$  da regressão simples geral (0,46), estes coeficientes representam um incremento de 13%, 90%, 46% para o milho, a soja e a cana, respectivamente. Deve-se notar, todavia, a limitação do sensor MODIS para representar a cultura agrícola mais fragmentada (milho). Portanto, para padrões espaciais semelhantes ao do milho na área de estudo, a estimativa de área com um sensor de resolução espacial moderada será limitada com ou sem a adoção de modelos de regressão.

A metodologia desenvolvida neste trabalho possibilitou a avaliação das diferenças obtidas nas estimativas de áreas de culturas agrícolas quando são utilizados sensores de resolução espacial moderada, ao invés de fina. O estudo da evolução dos valores das métricas de paisagem, em função de degradações sucessivas sobre a imagem ETM+, permitiu que se entendesse o comportamento dos índices espaciais à medida que a agregação espacial se torna mais severa. Todos os gráficos gerados foram coerentes com a definição conceitual de cada índice.

Pôde-se observar, também, que há uma forte influência do algoritmo de simulação que é aplicado para obter as imagens degradadas. Dessa forma, na impossibilidade de utilização de sensores que estejam em operação, é fundamental uma escolha criteriosa da metodologia a ser empregada para a simulação das imagens desejadas. O procedimento de filtragem espacial possibilitou, ainda, a identificação de índices

redundantes, sendo útil, portanto, para apoiar a criação de modelos estatísticos eficientes. A filtragem de textura foi mais eficiente na tarefa de simular o sensor MODIS (250m). Todavia tal desempenho indicou que se fazem necessários estudos adicionais para identificar, ou desenvolver, algoritmos que permitam resultados mais fidedignos, que tragam maior confiabilidade às análises realizadas sobre as imagens.

## **6.2 Reflexões Finais e Recomendações**

A metodologia aqui desenvolvida é coerente à aplicação para a estimativa de desflorestamento. É possível antecipar que o padrão espacial característico, bem como a simplicidade da composição das classes (floresta e não-floresta, apenas), tendem a facilitar a análise e interpretação dos resultados. Deste modo, projetos como o de estimativa de desmatamento da Amazônia (PRODES), desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE), podem ser aperfeiçoados com o emprego da técnica desenvolvida nesta pesquisa, para o cálculo de área agrícola.

A existência de diversas métricas de paisagem perfeitamente correlacionadas é um fator crítico para o sucesso de um modelo de regressão múltipla que as utilize como variáveis independentes. Portanto, recomenda-se, também, a pesquisa de novas métricas que denotem outros aspectos das classes, ainda que, para se lograr êxito considerável, seja necessária a criação de índices verdadeiramente novos.

Nessa linha de raciocínio, como alternativa ao método desenvolvido neste trabalho, sugere-se, por fim, a confecção do cálculo do padrão espacial das culturas utilizando o conceito de cenário, em substituição à abordagem por classe. Tal procedimento permite a aplicação de diversas outras métricas de paisagem, que podem atuar de maneira diferente na explicação da área das culturas agrícolas.



## REFERÊNCIAS BIBLIOGRÁFICAS

- Agrobyte. **Cana de açúcar**. Disponível em:  
<[http://www.agrobyte.com.br/cana\\_n.htm](http://www.agrobyte.com.br/cana_n.htm)>. Acesso em: 17 fev. 2005.
- Agência Espacial Brasileira (AEB). **Programa Nacional de Atividades Espaciais (PNAE)**. Brasília: MCT/AEB, 2005. 114 p.
- Bailey, T. C.; Gatrell, A. C. **Interactive spatial data analysis**. Essex: Longman, 1995. 413 p.
- Banon, G. J. F. Simulação de imagens de baixa resolução. **SBA: Controle e Automação**, v. 2, n. 3, p. 180-192, 1990.
- Belward, A. S.; Lambin, E. Limitations to the identification of spatial structures from AVHRR data. **International Journal of Remote Sensing**, v. 11, n. 5, p. 921-927, May 1990.
- Centro Nacional de Referência de Biomassa (CENBIO). **Cana-de-açúcar no Brasil**. São Paulo. Disponível em: <[http://infoener.iee.usp.br/scripts/biomassa/br\\_cana.asp](http://infoener.iee.usp.br/scripts/biomassa/br_cana.asp)>. Acesso em: 17 fev. 2005.
- Chavez Jr., P. S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. **Remote Sensing of Environment**, v. 24, n. 9, p. 459-479, Sep. 1988.
- Confederação da Agricultura e Pecuária do Brasil (CNA). **Agropecuária brasileira: balanço 2004 – perspectivas 2005**. Disponível em  
<<http://www.cna.org.br/cna/publicacao/noticia.wsp?tmp.noticia=4256>>. Acesso em 10 fev. 2005.
- Coordenadoria de Assistência Técnica Integral (CATI). **Projeto Lupa**. Disponível em:  
<<http://www.cati.sp.gov.br>>. Acesso em: 01 mar. 2006.

- Coppedge, B. R.; Engle, D. M.; Fuhlendorf, S. D.; Masters, R. E.; Gregory, M. S. Landscape cover type and pattern dynamics in fragmented southern Great Plains grasslands, USA. **Landscape Ecology**, v. 16, n. 8, p. 677–690, Dec. 2001.
- Cullinan, V. I.; Thomas, J. M. A comparison of quantitative methods for examining landscape pattern and scale. **Landscape Ecology**, v. 7, n. 3, p. 211-227, Sep.1992.
- D'Agostino, R. B.; Stephens, M. A. **Goodness-of-fit techniques**. New York: Marcel Dekker, Inc. 1986, 560 p.
- Duarte, J. O. **Importância econômica do milho**. Sete Lagoas: Embrapa Milho e Sorgo. Disponível em: <<http://www.cnpms.embrapa.br/publicacoes/milho/importancia.htm>>. Acesso em: 10 fev. 2005.
- Empresa Brasileira de Pesquisa Agropecuária (Embrapa-Soja). **A soja**. Londrina. Disponível em: <[http://www.cnpso.embrapa.br/index.php?op\\_page=22&cod\\_pai=16](http://www.cnpso.embrapa.br/index.php?op_page=22&cod_pai=16)>. Acesso em:17 fev. 2005a.
- Empresa Brasileira de Pesquisa Agropecuária (Embrapa-Soja). **A soja**. Londrina. Disponível em: <[http://www.cnpso.embrapa.br/index.php?op\\_page=24&cod\\_pai=22](http://www.cnpso.embrapa.br/index.php?op_page=24&cod_pai=22)>. Acesso em: 17 fev. 2005b.
- Epiphany, J. C. N.; Neto, O. O. B.; Luiz, A. J. B.; Formaggio, A. R. Sistema de amostragem em imagem como base para estimativa de áreas de culturas agrícolas no município de Ipuã-SP. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., 2001, Foz do Iguaçu. **Anais...** São José dos campos: INPE, 2001. p. 59-66. Disponível na biblioteca digital URLib: <<http://iris.sid.inpe.br:1908/>>. Acesso em: 12 fev. 2005.
- Fonseca, J. S.; Martins, G. A.; Toledo, G. L. **Estatística aplicada**. São Paulo: Atlas, 1985. 267 p.
- Frohn, R. **Remote sensing for landscape ecology: new metric indicators for monitoring, modeling, and assessment of ecosystems**. Boca Raton, FL: Lewis, 1998. 99 p.

Frohn, R. C.; Hao, Y. Landscape metric performance in analyzing two decades of deforestation in the Amazon Basin of Rondonia, Brazil. **Remote Sensing of Environment**, v. 100, n. 2, p. 237-251, Jan.2006.

Gürtler, S. **Estimativa de área agrícola a partir de sensoriamento remoto e banco de pixels amostrais**. 2003. 179 p. (INPE-9774-TDI/858). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003. Disponível na biblioteca digital URLib: <<http://iris.sid.inpe.br:1912/rep-/sid.inpe.br/jeferson/2003/06.02.07.29>>. Acesso em: 17 fev. 2005.

Hale, R. C.; Hanuschak, G. A.; Craig, M. E. Appropriate role of remote sensing in U.S. agricultural statistics. In: Seminar on Remote Sensing for Agricultural Statistics. (FAO), 9 – 11 June 1999, Bangkok, Thailand. **Proceedings...**Bangkok: FAO, 1999.

Haralick, R. M. Statistical and structural approaches to texture. **Proceedings of the IEEE**. v. 67, n. 5, p. 786-804, May 1979.

Haralick, R. M.; Shanmugan, K. ; Dinstein, I. Textural features for image classification. **IEEE Transactions, Man, and Cybernetics**, v. 3, n. 6, 610–621 Nov.1973.

Henderson-Sellers, A.; Pitman, A. J. Land-surface schemes for future climate models: specification, aggregation and heterogeneity. **Journal of Geophysical Research**, v. 97, n. D3, p. 2687-2696, Feb.1992.

Henderson-Sellers, A.; Wilson, M. F.; Thomas, G. The effect of spatial resolution on archives of land cover type. **Climate Change**, v.7, n. 4, 391-402, Dec. 1985

Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo agropecuário de 1995-1996**. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 15 fev. 2005a.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Cidades**. Disponível em: <<http://200.255.94.66/cidadesat/default.php>>. Acesso em: 18 fev. 2005b.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Pesquisas agropecuárias**. Rio de Janeiro, 2002. Disponível em:

<<http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/pesquisasagropecuarias2002.pdf>>. Acesso em: 10 fev. 2005.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Sistema IBGE de Recuperação Automática (SIDRA)**. Disponível em: <<http://www.sidra.ibge.gov.br>>.

Acesso em: 01 mar. 2006.

Jeanjean, H.; Malingreau, J. P.; Achard, F. Forest fragmentation: typology and characterization. In: European Symposium on Satellite Sensing, 2. (ESSS), 26-30 Sep. 1994, Rome, Italy. **Proceedings...** Rome: SPIE, 1994.

Jensen, J. R. **Remote Sensing of the Environment**: an earth research perspective. Upper Saddle River: Prectice Hall, 2000. 544 p.

Justice, C. O.; Markham, B. L.; Townshend, J. R. G. e Kennard, R. L. Spatial degradation of satellite data. **International Journal of Remote Sensing**, v. 10, n. 9, p. 1539-1561, Sep. 1989.

Justice, C.; Vermote, E.; Townshend, J. R. G.; Defries, R.; Roy, D. P.; Hall, D. K.; Salomonson, V. V.; Privette, J.; Riggs, G.; Strahler, A.; Lucht, W.; Myneni, R.; Knjazihhin, Y.; Running, S.; Nemani, R.; Wan, Z.; Huete, A.; van Leeuwen, W.; Wolfe, R. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 4, p. 1228-1249, July 1998.

Landis, J.; Koch, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 3, p. 159-174, Mar. 1977.

Legendre, P.; Fortrin, M.-J. Spatial pattern and ecological analysis. **Vegetatio**, v. 80, n. 2, p. 107-138, Feb. 1989.

- Looney, S. W.; Gullledge, T.R. Use of the correlation coefficient with normal probability plots. **The American Statistician**, v. 39, n. 1, p. 75-79, Feb. 1985.
- Luiz, A. B. Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto. dez. 2003. (INPE--TDI/). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.
- Luiz, A. J. B.; Oliveira, J. C.; Epiphanyo, J. C. N.; Formaggio, A. R. Auxílio das imagens de satélite aos levantamentos por amostragem em agricultura. **Agricultura em São Paulo**, v. 49, n.1, p. 41-54, 2002. Disponível em: <<http://www.iea.sp.gov.br/out/publicacoes/asp-1-02.htm>>. Acesso em: 12 fev. 2005.
- Malingreau, J. P.; Belward, A. S. Scale considerations in vegetation monitoring using AVHRR data. **International Journal of Remote Sensing**, v. 13, n.12, p.2289-2307, Aug.1992.
- Masuoka, E.; Fleig, A.; Wolfe, R. E.; Patt, F. Key characteristics of MODIS data products. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 4, p. 1313-1323, July 1998.
- Mather, P. M. **Computer processing of remotely-sensed images: an introduction**. West Sussex: John Wiley & Sons, 2<sup>a</sup> Ed. 1999, 292 p.
- Mayaux, P.; Lambin E. F. Estimation of tropical forest area from coarse spatial resolution data: a two-step correction function for proportional errors due to spatial aggregation. **Remote Sensing of Environment**, v. 53, n. 1, p. 1-15, July 1995.
- Mayaux, P.; Lambin E. F. Tropical forest area measured from global land-cover classifications: inverse calibration models based on spatial textures. **Remote Sensing of Environment**, v. 59, n. 1, p. 29-43, Jan. 1997.
- McGarigal, K.; Marks, B. J. **FRAGSTATS: spatial pattern analysis program for quantifying landscape structure**. Technical Report PNW-GTR-351. Portland: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 1995.122 p.

Millington, A. C.; Velez-Liendo, X. M.; Bradley, A. V. Scale dependence in multitemporal mapping of forest fragmentation in Bolivia: implications for explaining temporal trends in landscape ecology and applications to biodiversity conservation. **ISPRS International Journal of Photogrammetry and Remote Sensing**, v. 57, n. 4, p. 289-299, Feb. 2003.

Ministério de Ciência e Tecnologia (MCT). **A cana-de-açúcar no Brasil**. Disponível em: <[http://www.mct.gov.br/clima/comunic\\_old/cana3.htm](http://www.mct.gov.br/clima/comunic_old/cana3.htm)>. Acesso em: 17 fev. 2005.

Moody, A. Using landscape spatial relationships to improve estimates of land-cover area from coarse resolution remote sensing. **Remote Sensing of Environment**, v. 64, n. 2, p. 202-220, May 1998.

Moody, A.; Woodcock, C. E. Scale-dependent errors in the estimation of land-cover proportions – implications for global land-cover datasets. **Photogrammetric Engineering and Remote Sensing**, v. 60, n. 5, p. 585-594, May 1994.

Moody, A.; Woodcock, C. E. The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. **Landscape Ecology**, v.10, n. 6, p. 363-379, Dec. 1995.

National Aeronautics and Space Administration (NASA). **About MODIS**, 2005a. Disponível em: <<http://modis.gsfc.nasa.gov/about/index.html>>. Acesso em: 13 fev. 2005.

National Aeronautics and Space Administration (NASA). **Landsat 7 science data users handbook**. Disponível em: <[http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook\\_toc.html](http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html)>. Acesso em: 16 fev. 2006.

National Aeronautics and Space Administration (NASA). **MODIS data**, 2005c. Disponível em: <<http://modis.gsfc.nasa.gov/data/dataproducts.html>>. Acesso em: 21 fev. 2005.

National Aeronautics and Space Administration (NASA). **MODIS technical specifications**, 2005b. Disponível em: <<http://modis.gsfc.nasa.gov/about/specs.html>>. Acesso em: 13 fev. 2005.

Nelson, R.; Holben, B. Identifying deforestation in Brazil using multiresolution satellite data. **International Journal of Remote Sensing**, v. 7, n. 1, p. 429-448, Jan. 1986.

Neter, J.; Kutner, M. N.; Nachtsheim, C. J.; Wasserman, W. **Applied linear statistical models**. Boston: WCB/McGraw-Hill, 4ª Ed. 1996, 791 p.

Pax-Lenney, M.; Woodcock, C. E. The effect of spatial resolution on the ability to monitor the status of agricultural lands. **Remote Sensing of Environment**, v. 61, n. 2, p. 210-220, Aug. 1997.

Pino, F. A. Estimação subjetiva de safras agrícolas. **Informações Econômicas**, v. 31, n. 6, p. 55-58, jun. 2001.

Ponzoni, F. J.; Galvão, L. S.; Epiphanyo, J. C. N. Spatial resolution influence on the identification of land cover classes in the Amazon environment. **Anais da Academia Brasileira de Ciências**, v. 74, n. 4, p. 717-725, Dec. 2002.

Ponzoni, F. J.; Rezende, A. C. P. Influência da resolução espacial de imagens orbitais na identificação de elementos da paisagem em Altamira – PA. **Árvore**, v. 26, n. 4, p. 403-410, jul./ago. 2002.

Richards, J. A. **Remote sensing digital image analysis: an introduction**. New York: Springer-Verlag, 1986. 281 p.

Rizzi, R. **Geotecnologias em um sistema de estimativa da produção de soja: estudo de caso no Rio Grande do Sul**. 2004. 214 p. (INPE-12271-TDI/983). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004. Disponível na biblioteca digital URLib: <<http://mtc-m12.sid.inpe.br:80/rep-/sid.inpe.br/jeferson/2004/07.02.14.17>>. Acesso em: 19 abr. 2006.

Salomonson, V. V.; Barnes, W. L.; Maymon, P. W.; Montgomery, H. E.; Ostrow, H. MODIS: Advanced facility instrument for studies of the earth as a system. **IEEE Transactions on Geoscience and Remote Sensing**, v. 27, n. 2, p. 145-153, Mar. 1989.

Sanches, I. D. **Sensoriamento remoto para o levantamento espectro-temporal e estimativa de área de culturas agrícolas**. 2004. 172 p. (INPE-10290-TDI/909). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004. Disponível na biblioteca digital URLib: <<http://iris.sid.inpe.br:1913/rep-/sid.inpe.br/jeferson/2004/05.14.10.59>>. Acesso em: 13 fev. 2005.

Saura, S. Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. **Landscape Ecology**, v.19, n. 2, p. 197-209, Feb. 2004.

Schowengerdt, R. A. **Remote sensing, models and methods for image processing**. San Diego: Academic Press, 2 Ed. 1997, 522 p.

Shapiro, S. S.; Wilk, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591-611, Dec. 1965.

Tukey, J. W. **Exploratory data analysis**. Reading, MA: Addison Wesley , 1977, 506 p.

Turner, M. G. ; O'Neill, R. V.; Gardner, R. H.; ; Milne, B. T. Effects of changing spatial scale on the analysis of landscape pattern. **Landscape Ecology**, v. 3, n. 3-4, p. 153-162, 1989.

United States Geological Survey (USGS). **MODIS Reprojection Tool user's manual**. South Dakota, 2004.

United States Geological Survey (USGS). **Thematic Mapper**. Sioux Falls. Disponível em: < <http://eros.usgs.gov/products/satellite/band.html>>. Acesso em: 20 fev. 2006a.

United States Geological Survey (USGS). **Thematic Mapper**. Sioux Falls. Disponível em: <<http://eros.usgs.gov/products/satellite/landsat7.html>>. Acesso em: 20 fev. 2006b.

Vermote, E. F. **MODIS surface reflectance; atmospheric correction algorithm products (also called spectral reflectance) (MOD 09)**. Disponível em: <[http://modis.gsfc.nasa.gov/data/dataproduct/pdf/MOD\\_09.pdf](http://modis.gsfc.nasa.gov/data/dataproduct/pdf/MOD_09.pdf)>. Acesso em: 21 fev. 2005

Wu, J.; Shen, W.; Sun W.; Tueller, P. T. Empirical patterns of the effects of changing scale on landscape metrics. **Landscape Ecology**, v.17, n. 8, p. 761-782, Nov. 2002.