

UNIVERSIDADE DE BRASÍLIA
CENTRO DE DESENVOLVIMENTO SUSTENTÁVEL

**USO DE TÉCNICAS DE ANÁLISE MULTIVARIADA PARA A PREDIÇÃO DE
DESMATAMENTO NA AMAZÔNIA: O MODELO ‘AMAZON-PD’**

Darcton Policarpo Damião

Orientador: Prof. Dr. Newton Moreira de Souza

Tese de Doutorado

Brasília-DF: Julho / 2007

Damião, Darcton Policarpo

Uso de técnicas de análise multivariada para a predição de desmatamento na Amazônia: o modelo 'AMAZON-PD'./
Darcton Policarpo Damião

Brasília, 2007.

128 p.:

Tese de Doutorado. Centro de Desenvolvimento Sustentável,
Universidade de Brasília, Brasília.

1. Desmatamento, Modelos de mudança de uso e de cobertura do solo, Sensoriamento remoto, Geoprocessamento. Universidade de Brasília. CDS.

II. Título.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta tese e emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta tese de doutorado pode ser reproduzida sem a autorização por escrito do autor.

UNIVERSIDADE DE BRASÍLIA
CENTRO DE DESENVOLVIMENTO SUSTENTÁVEL

**USO DE TÉCNICAS DE ANÁLISE MULTIVARIADA PARA A PREDIÇÃO DE
DESMATAMENTO NA AMAZÔNIA: O MODELO 'AMAZON-PD'**

Darcton Policarpo Damião

Tese de Doutorado submetida ao Centro de Desenvolvimento Sustentável da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do Grau de Doutor em Desenvolvimento Sustentável, área de concentração em Política e Gestão Ambiental.

Aprovado por:

Prof. Newton Moreira de Souza, Dr. (UnB)
(Orientador)

Prof. Antônio Brasil, Dr. (UnB)
(Examinador Interno)

Prof. Fernando Scardua, Dr. (UnB)
(Examinador Interno)

Prof. Paulo Roberto Meneses, Dr. (UnB)
(Examinador Externo)

Prof.^a Thelma Krug, Dra. (INPE)
(Examinadora Externa)

Prof.^a Tatiana Mora Kuplich, Dra. (INPE)
(Examinadora Externa)

Brasília-DF, 27 de julho de 2007

Dedico este trabalho aos meus filhos Alexandre, André Luiz e Ana Luíza, na esperança de influenciá-los ainda mais na busca do conhecimento.

AGRADECIMENTOS

Ao Estado-Maior da Aeronáutica, pelas excepcionais condições de trabalho proporcionadas ao longo de toda a minha permanência como oficial-adjunto da Terceira Subchefia, em especial no período de 2003 a 2006. Em particular, agradeço ao apoio prestado pelo Maj Brig Ar Salamone e pelo Brig Ar Robinson.

Ao Comando-Geral de Operações Aéreas, pela pronta-resposta no atendimento à solicitação de apoio aéreo, fundamental à condução do trabalho de campo.

Aos integrantes do 1º ETA e do 1º/8º GAV, nas pessoas de seus comandantes (2005) – Ten Cel Siqueira e Ten Cel Ciarlini. Em particular, agradeço à camaradagem e à dedicação dos tripulantes do FAB 8650 – Ten Gava, Ten Cardote, SO Schoffen e SO Carlos.

À empresa Imagem, na pessoa de Luiz Leonardi, pela generosidade singular na disponibilização de técnicos que auxiliaram sobremaneira na compreensão do ArcGIS 9.2.

A Valdete Duarte, pelo desprendimento e pela atenção dedicados quando da orientação acerca do PRODES.

Aos professores José Ângelo Belloni e Eduardo Freitas, pelo acesso à análise multivariada e aos dados categorizados.

Ao amigo Vitor Lopes Mendes pela disposição irrestrita em ajudar, mesmo numa época em que o desenho do trabalho ainda não passava de rascunho.

À mente brilhante de Antônio Carlos Marins Pedroso, 77 puro-sangue, pelas preciosas considerações feitas em um momento se suma importância – a defesa.

A minha amiga Simone Shiki agradeço o prazer do convívio e as trocas de idéias sempre enriquecedoras.

Ao amigo, quase irmão, Eduardo Henrique Geraldi Araújo, pela ajuda sem a qual a conclusão desta tese não seria possível, pelo menos não dentro do prazo previsto.

Ao meu orientador, Prof. Dr. Newton Moreira de Souza, pela confiança externada e pelos *insights* sempre precisos e esclarecedores sobre o melhor caminho a seguir.

A Cláudia, minha cara-metade, por existir simplesmente. Seu entusiasmo e sua dedicação relativos ao meu trabalho são tão inspiradores que me compelem a combater insistentemente a mediocridade.

“Uma resposta errada para um problema certo pode, via de regra, ser reparada. Uma resposta certa para um problema errado, no entanto, é algo muito difícil de se corrigir, a começar pela dificuldade de seu diagnóstico.”

(Peter F. Drucker, 2004)

RESUMO

Dono da maior floresta tropical do mundo, o Brasil tem envidado grandes esforços, conduzidos em setores diversos da sociedade, dirigidos para mitigar os efeitos do desmatamento na Amazônia. Esta tese propõe um estudo de predição de desmatamento na Amazônia, por meio de técnicas de análise multivariada, mais especificamente um modelo de regressão logística. Dadas as características da região, o uso de dados de sensoriamento remoto de média resolução, organizados e modelados em sistemas de informação geográfica, constitui uma prática cada vez mais consagrada. Neste caso, buscou-se levantar as relações espaço-temporais existentes entre variáveis biofísicas identificadas em uma única cena Landsat, localizada no município de São Félix do Xingu, Sul do Pará, observada em seis datas espaçadas de 1985 a 2004. Assim, o modelo de Análise Multivariada Aplicada a Zoneamento para Predição de Desmatamento (AMAZON-PD) foi formulado, testado e validado. Para tanto, na etapa de pré-processamento, os dados foram divididos em quatro classes – floresta, não-floresta, hidrografia e desmatamento – tomando-se por base os procedimentos metodológicos desenvolvidos pelo INPE para o PRODES DIGITAL. Em seguida, após a escolha e a geração das variáveis empregadas no modelo, ano a ano foram gerados os coeficientes de regressão logística que permitiram compreender, inicialmente de maneira estática e depois dinamicamente, o efeito das variáveis independentes na variável-resposta (impacto ambiental = desmatamento). Os resultados alcançados possibilitaram inferir a localização futura de desmatamentos, cuja precisão foi verificada a partir de tabelas de classificação.

Palavras-chave: desmatamento, sensoriamento remoto, geoprocessamento, sistemas de informação geográfica, análise multivariada, regressão logística, modelos de mudanças na cobertura do solo.

ABSTRACT

Owner of the world largest rainforest, Brazil put forth its best effort in many different segments of the society looking forward to mitigating Amazon deforestation effects. This thesis proposes a prediction study of Amazon deforestation throughout multivariate analysis techniques, more specifically a logistic regression model. Given regional characteristics, usage of remote sensing data of medium resolution, organized and modeled in geographical information systems, constitutes an increasingly adopted *praxis*. In this case, the study was intended to raise space-temporal relationships that exist among biophysical variables identified in a single Landsat scene, located in São Félix do Xingu, Southern of Pará State, Brazil, observed along with six different dates from 1985 to 2004. Therefore, the “Multivariate Analysis Applied to Zoning for Deforestation Prediction” model (AMAZON-PD) was conceived, tested, and validated. In order to achieve such goal, during the pre-processing stage, the data were divided in four categories – forest, non-forest, hydrography, and deforestation – taking into account the methodological procedures developed to PRODES DIGITAL by the Brazilian National Institute for Space Research (INPE). Right after selection and formation of the variables to be used in the model, the following step was to generate the logistic regression coefficients in a year basis. This procedure played a key role in the understanding of the independent variables effects on the response-variable (environmental impact = deforestation), first in a static fashion and later on in a dynamic way. The attained results, which accuracy were verified upon classification tables, allowed to infer the future location of deforestation sites.

Keywords: deforestation, remote sensing, geoprocessing, geographical information systems, multivariate analysis, logistic regression, land cover change models.

RESUMÉ

Propriétaire de la plus grande forêt du monde, le Brésil emploie des grands efforts, conduits en divers secteurs de la société, dirigés pour soulager les effets de la déforestation en Amazonie. Cette thèse propose une étude de prédiction de déforestation en l'Amazonie, par des techniques d'analyse multivariées, plus spécifiquement un modèle de régression logistique. En considérant les caractéristiques de la région, l'utilisation de données de télédétection satellitale de moyenne résolution, organisées et modelées en système d'information géographique, constitue une pratique de plus en plus consacrée. Dans ce cas, les relations espace-temps existantes ont été recherchées parmi les variables biophysiques identifiées dans une seule scène Landsat, placée au village de São Félix do Xingu, au sud de l'État de Pará, observée en six dates distribuées entre 1985 et 2004. De cette façon, le modèle d'Analyse Multivariée Appliquée à la Démarcation de Zones pour Prédiction de Déforestation (AMAZON-PD) a été formulé, testé et validé. Pour cela, à l'étape d'élaboration, les données ont été divisées en quatre classes – forêt, non-forêt, hydrographie et déforestation – en considérant les procédés méthodologiques développés par INPE pour le PRODES DIGITAL. En suite, après le choix et la génération des variables appliquées sur le modèle, à chaque année, les chiffres de régression logistique ont été définis et ont permis de comprendre, premierment de manière statique et après, dynamiquement, l'effet des variables indépendantes sur la variable-réponse (impact environnemental = déforestation). Les résultats atteints ont permis de déduire la localisation de déforestation future, dont la précision a été vérifiée à partir de tableaux de classification.

Mots clés: déforestation, télédétection satellitale, geoprocédures, système d'information géographique, analyse multivariée, régression logistique, modèles de changement de couverture du sol.

SUMÁRIO

LISTA DE FIGURAS	III
LISTA DE TABELAS	V
INTRODUÇÃO	1
CONTEXTUALIZAÇÃO E IMPORTÂNCIA DO TRABALHO.....	1
O PROBLEMA INVESTIGADO	4
<i>O objeto e a hipótese da pesquisa</i>	5
<i>Hipóteses secundárias</i>	6
VARIÁVEIS.....	6
<i>Fatores de ordem biofísica</i>	6
<i>Impacto ambiental</i>	7
UNIVERSO GEOGRÁFICO E TEMPORAL.....	7
PROCEDIMENTOS METODOLÓGICOS.....	7
ESTRUTURA DO TRABALHO	8
1 - ARCABOUÇO TEÓRICO DA MUDANÇA NO USO E NA COBERTURA DO SOLO	9
1.1 - O ESTUDO DAS MUDANÇAS NO USO E NA COBERTURA DO SOLO.....	11
1.2 - RELAÇÃO ENTRE TEORIAS E MODELOS DE ESTUDO DA MUDANÇA NO USO E NA COBERTURA DO SOLO	20
1.3 - TEORIAS DO ESTUDO DA MUDANÇA NO USO E NA COBERTURA DO SOLO.....	21
1.3.1 - <i>A tradição teórica da economia urbana e regional</i>	25
1.3.2 - <i>A tradição teórica sociológica e de política econômica</i>	26
1.3.3 - <i>A tradição teórica natureza-sociedade</i>	26
1.4 - MODELAGENS DO ESTUDO DA MUDANÇA NO USO E NA COBERTURA DO SOLO	27
1.4.1 - <i>Modelos estatísticos e econométricos</i>	28
1.4.2 - <i>Modelos de interação espacial</i>	31
1.4.3 - <i>Modelos de otimização</i>	31
1.4.4 - <i>Modelos integrados</i>	32
1.4.5 - <i>Abordagens recentes</i>	34
1.4.6 - <i>Considerações sobre os modelos apresentados</i>	35
1.5 - EXEMPLOS DE MODELOS DE MUDANÇA NO USO E NA COBERTURA DO SOLO - AMAZÔNIA.....	36

2 -	MODELOS DE REGRESSÃO	39
2.1 -	REGRESSÃO LINEAR	41
2.1.1 -	<i>Regressão linear simples</i>	41
2.1.2 -	<i>Regressão Linear Múltipla</i>	42
2.2 -	REGRESSÃO LOGÍSTICA	44
2.2.1 -	<i>Origem do conceito</i>	44
2.2.2 -	<i>O Modelo Logístico</i>	45
2.2.3 -	<i>A transformação logito</i>	47
2.2.4 -	<i>Derivação da equação da razão de chances</i>	49
2.2.5 -	<i>Interpretação de Coeficientes de Regressão Logística</i>	50
2.3 -	O PROBLEMA MULTIVARIADO DA PESQUISA	54
2.3.1 -	<i>Análise Simples</i>	55
2.3.2 -	<i>Interação Multiplicativa</i>	56
2.3.3 -	<i>Efeitos de confusão e de interação - o Modelo EVW</i>	57
2.3.4 -	<i>Estratégia de modelagem</i>	58
3 -	CONCEPÇÃO DO MODELO – PROCEDIMENTOS METODOLÓGICOS	62
3.1 -	COLETA DE DADOS.....	63
3.1.1 -	<i>Concepção inicial</i>	63
3.1.2 -	<i>Escolha da área-teste e aquisição de imagens</i>	64
3.1.3 -	<i>Realização do trabalho de campo</i>	65
3.2 -	ORGANIZAÇÃO DOS DADOS DE ENTRADA.....	66
3.2.1 -	<i>Georreferenciamento e registro de imagens</i>	68
3.2.2 -	<i>Classificação temática</i>	68
3.2.3 -	<i>Escolha e geração das variáveis</i>	74
4 -	FORMULAÇÃO E IMPLEMENTAÇÃO DO MODELO – RESULTADOS E DISCUSSÃO	87
4.1 -	FORMULAÇÃO DO MODELO.....	87
4.1.1 -	<i>Regressão logística entre diferentes datas – o modelo inicial</i>	87
4.1.2 -	<i>Eliminação de variáveis – o modelo ideal</i>	89
4.2 -	IMPLEMENTAÇÃO DO MODELO.....	101
4.2.1 -	<i>AMAZON – o modelo estático</i>	108
4.2.2 -	<i>AMAZON-PD – o modelo dinâmico</i>	108
	CONCLUSÕES E RECOMENDAÇÕES	113
	REFERÊNCIAS BIBLIOGRÁFICAS	117
	ANEXOS	123

LISTA DE FIGURAS

Fig. 2.1 – Problemas intrínsecos à regressão linear, quando a variável dependente é binária.	43
Fig. 2.2 – Gráfico da função logística.	45
Fig. 2.3 – Eliminação de variáveis no modelo EVW proposto por Kleinbaum e Klein (2002).	61
Fig. 3.1 – Esquema geral adotado na concepção do modelo AMAZON.	64
Fig. 3.2 – Rotas percorridas pela aeronave FAB 8530 do 1º ETA, a partir do aeródromo de São Félix do Xingu.	66
Fig. 3.3 – A área de estudo em relação ao estado do Pará e ao Arco do Desmatamento.	67
Fig. 3.4 – Área – teste, formada pelas coordenadas (1) 052°40’W/05°49’S, (2) 052°06’W/05°53’S, (3) 052°46’W/06°22’S, e (4) 052°12’W/06°27’S.	67
Fig. 3.5 – Esquema de georrefenciamento adotado.	69
Fig. 3.6 – Imagens-fração de solo (a), sombra (b) e vegetação (c), geradas a partir do modelo linear de mistura espectral, para o ano de 1996.	71
Fig. 3.7 – Imagem-fração de 1996 (sombra) segmentada pelo algoritmo de crescimento de região.	72
Fig. 3.8 – Imagem classificada (1996) com base em algoritmo de agrupamento.	73
Fig. 3.9 – Mapeamento final para o ano de 1996.	74
Fig. 3.10 – Esquema geral para obtenção da variável PAD.	76
Fig. 3.11 – Esquema geral para obtenção da variável ROD.	77
Fig. 3.12 – Esquema geral para obtenção da variável HID.	77
Fig. 3.13 – Esquema geral para obtenção da variável EPA.	78
Fig. 3.14 – Esquema geral para obtenção das variáveis DEC e ALT.	79
Fig. 3.15 – Espacialização da variável DEC no entorno da área de estudo.	79
Fig. 3.16 – Espacialização da variável ALT no entorno da área de estudo.	80
Fig. 3.17 – Espacialização da variável IFR para todos os anos do período estudado.	82
Fig. 3.18 – Esquema geral para obtenção das variáveis IFR e APE.	82
Fig. 3.19 – Espacialização da variável APE para todos os anos do período estudado.	83

Fig. 3.20 – Exemplos de espacialização das variáveis W_1 (esquerda) e W_2 (direita) para 2000.	84
Fig. 3.21 – Esquema geral para obtenção das variáveis W_1 (a) e W_2 (b).	85
Fig. 4.1 – Evolução das classes ‘floresta’ e ‘desmatamento’, de 1985 a 2004.	89
Fig. 4.2 – Evolução da Proximidade da Área Desmatada (PAD) de 1985 a 2004. No intuito de se manter um padrão de cores constante, a escala de distâncias para PAD assume um caráter relativo nessas imagens. Em consequência, as distâncias classificadas com a cor verde em 1985, por exemplo, são bem maiores do que aquelas de 2004.	92
Fig. 4.3 – Datas de catalogação pelo Imazon de novas rodovias na área de estudo.	94
Fig. 4.4 – Exemplo de descompasso entre novas rodovias (brancas) que ainda não se encontram catalogadas e rodovias mais antigas (pretas), estas já catalogadas.	95
Fig. 4.5 – Evolução da variável EPA de 1996 a 2004.	96
Fig. 4.6 – Exemplo de queimada “morro acima”, diversas vezes observada no trabalho de campo.	98
Fig. 4.7 – Exemplo de células (ano-base de 2000) com baixo índice de fragmentação (verde).	100
Fig. 4.8 – Ilustração dos componentes envolvidos na geração das tabelas de classificação.	102
Fig. 4.9 – Melhor desempenho do AMAZON no ano-base de 1985.	103
Fig. 4.10 – Melhor desempenho do AMAZON no ano-base de 1988.	104
Fig. 4.11 – Melhor desempenho do AMAZON no ano-base de 1992.	105
Fig. 4.12 – Melhor desempenho do AMAZON no ano-base de 1996.	106
Fig. 4.13 – Melhor desempenho do AMAZON no ano-base de 2000.	107
Fig. 4.14 – Modelo para o ano-base de 2000, que utiliza somente a variável PAD (P00), mas com a incorporação da dimensão temporal (AMAZON-PD).	110
Fig. 4.15 – AMAZON-PD para o ano-base de 2004 (predição de IPT para 2008) com três diferentes configurações: (a) somente com a variável PAD, consideradas todas as datas anteriores – 1985 a 2000; (b) com as variáveis PAD e ALT, considerados os anos-base de melhor desempenho para essas variáveis – 1988, 1996 e 2000; e (c) com as variáveis PAD, ROD, EPA, DEC e ALT, considerado o “aprendizado” proporcionado pelos anos-base de 1996 e 2000.	112

LISTA DE TABELAS

TABELA 4.1 – Síntese das tabelas de classificação utilizadas para avaliar o AMAZON.....	90
TABELA 4.2 – Comportamento isolado da variável de exposição PAD (1985 a 2000).....	91
TABELA 4.3 – Comportamento de ROD frente à variável de exposição PAD (1985 a 2000).....	93
TABELA 4.4 – Comportamento de EPA frente à variável de exposição PAD (1996 a 2000).....	97
TABELA 4.5 – Comportamento de DEC frente à variável de exposição PAD (1985 a 2000).....	98
TABELA 4.6 – Comportamento de ALT frente à variável de exposição PAD (1985 a 2000).....	99
TABELA 4.7 – Tabela de classificação do AMAZON _{P00}	109

INTRODUÇÃO

Esta tese propõe um estudo de predição de desmatamento na Amazônia, por meio de técnicas de análise multivariada¹, mais especificamente um modelo de regressão logística. Realizado a partir de análises feitas com o auxílio de pesquisa de campo, dados de sensoriamento remoto e de sistemas de informação geográfica, o estudo em questão tem a finalidade de gerar subsídios úteis e consistentes para a formulação de políticas públicas ambientais mais efetivas nessa área.

Para tanto, foi conduzida uma pesquisa baseada em técnicas de sensoriamento remoto e de geoprocessamento, na qual um modelo estatístico de regressão logística é aplicado às relações espaciais entre feições extraídas de imagens do satélite de observação terrestre Landsat 5 e o fenômeno do desmatamento². Uma área localizada no município de São Félix do Xingu, Sul do Pará, foi escolhida para a implementação desse modelo.

CONTEXTUALIZAÇÃO E IMPORTÂNCIA DO TRABALHO

A Amazônia constitui apenas um dos seis biomas brasileiros³, de acordo com a classificação adotada pelo Brasil (MMA), no contexto da Convenção sobre Diversidade Biológica (CAMARGO *et al.*, 2002). Apesar disso, em função das diferentes estratificações a que pode ser submetido o espaço amazônico, são diversas as “Amazônias” contidas nesse único grande bioma.

A Amazônia Oriental, que engloba a área de estudo deste trabalho, constitui uma das quatro partes existentes do novo macrozoneamento proposto por Becker (1999). Foi esta a primeira porção do espaço amazônico a ser articulada na expansão recente da fronteira, tendo como eixo de penetração a rodovia Belém-Brasília.

¹ - Análises estatísticas multivariadas constituem um termo genérico que inclui um conjunto de técnicas estatísticas voltadas à análise das relações entre muitas variáveis. A análise de regressão logística é uma dessas técnicas.

² - Para os fins desta tese, considera-se desmatamento somente aquele referente à primeira ação de retirada da cobertura florestal nativa.

³ - Agrupamento de regiões fitoecológicas, elaborado pelo Instituto Socioambiental a partir do Mapa de Vegetação, IBGE, 1993, escala 1:5.000.000, exceto Pantanal (baseado nas propostas da WWF e Ramsar), que inclui: Amazônia, Mata Atlântica, Pantanal, Cerrado, Caatinga, Campos Sulinos e Costeiro (CAMARGO *et al.*, 2002).

O Estado, por meio de subsídios e créditos, privilegiou a acumulação de fazendeiros individuais e empresas de grupos econômicos. Para tanto, intensa mobilidade da força de trabalho foi estimulada (BECKER *et al.*, 1990). Tal movimento visava a integração nacional idealizada pelo Estado, na medida em que possibilitava a incorporação da última fronteira nacional ao restante do País.

Nas últimas três décadas, os programas e projetos implementados pelo Governo Federal aceleraram o processo de ocupação territorial na Amazônia. Dirigidos para a transformação de sua estrutura produtiva e com vistas à sua integração com o restante do País, esses programas e projetos alteraram sensivelmente os cenários social, econômico e ambiental amazônicos.

Alguns resultados desse processo foram de grande importância para a região, sobretudo os que redundaram em melhorias nas redes de transporte e de comunicação, com a abertura de estradas e a expansão das redes de distribuição de energia elétrica, que atenuaram significativamente o seu isolamento.

Por outro lado, foram geradas sérias contradições sociais e econômicas, caracterizadas por conflitos com as populações tradicionais da região, além da acelerada e inadequada exploração dos seus recursos naturais. Em decorrência do fluxo crescente de grandes contingentes migratórios, atraídos pelas riquezas minerais e a aparente fertilidade natural dos solos florestais, houve um crescimento populacional acelerado.

As possibilidades de emprego nas obras de infra-estrutura de comunicação e abastecimento urbano foram outro fator de atração. Essas obras, no entanto, não acompanharam o ritmo de crescimento das populações urbanas.

Dois fatores atuaram para acelerar o processo de ocupação e o uso dos recursos naturais na Amazônia: (1) o esgotamento dos recursos florestais na fronteira antiga – Norte do Mato Grosso, Sudeste e Nordeste do Pará e Rondônia; e (2) o investimento governamental em novas obras de infra-estrutura na Amazônia.

Esse processo culminou com o surgimento do chamado “Arco do Desmatamento”, um corredor de cerca de 3000 km de extensão, por até 600 km de largura, que abrange sete estados: Acre, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia e Tocantins.

De acordo com Pasquis *et al.* (2001), em termos gerais, o Arco está constituído de várias faixas ou etapas de valorização, que se deslocam e se substituem com base no modelo dominante de colonização, envolvendo sucessivos atores e sistemas de produção diferentes. Essas diferentes faixas são objeto de profundos processos de mudanças nas relações entre as

sociedades rurais e seus territórios. Em cada uma delas, os atores sociais presentes, tendo interesses e percepções diferentes da mesma realidade, lutam para impor seu modelo de desenvolvimento e seu formato de ocupação do espaço, provocando conflitos variados.

O desmatamento, na Amazônia, normalmente obedece ao ciclo “exploração madeireira predatória/assentamento/pecuária extensiva”, que vem sendo implantado na região desde o final da década de 1970. Em função de sua rapidez, esse ciclo acaba por gerar conflitos sociais e ambientais. Via de regra, o resultado é a busca, pelos pequenos produtores, por áreas florestais distantes para desmatar e sobreviver. Em termos de dinâmica, a “mineração da madeira” - exploração discriminada e predatória, precede a ocupação da terra. A um só tempo, essa atividade propicia o benefício econômico advindo de certos recursos madeireiros disponíveis e a “limpeza” do terreno, o que permite, por sua vez, a sua valorização nas etapas seguintes. Realizada *a priori* da chegada dos posseiros, apresenta um ritmo mais acelerado que o deles. (BECKER, 1999)

Uma vez esgotado o recurso, a exploração se desloca para mais adiante, deixando um rastro de terras degradadas e abandonadas. Quando a base econômica (madeira) de um município se perde, há desemprego, migração para outros locais e diminuição dos serviços públicos. (ARIMA, 2001)

De acordo com Sawyer (2001), o ciclo de forte imigração inter-regional para a Amazônia, que se estendeu até meados de 1986, está encerrado. Atualmente, a mobilidade predominante é crescentemente intra-regional e temporária. O crescimento populacional encontra-se em estado vegetativo (fecundidade elevada e mortalidade média).

Seja como for, é fato notório que as diferenças internas de ocupação exercem pressões sobre o meio ambiente. No caso da Amazônia, observa-se um crescimento da pecuária, principalmente nos entornos do Arco do Desmatamento. A pecuária, aliada ao desmatamento, constitui um instrumento de legitimação jurídica da propriedade para os latifundiários.

Para os pequenos produtores, a pecuária significa o remédio para a degradação causada pelas primeiras culturas anuais. É com ela que eles vislumbram a possibilidade imediata de valorização da terra degradada. Não bastassem esses aspectos endógenos, um outro fator que exerce pressão sobre o meio ambiente é a demanda por madeira, que não cessa de crescer, sobretudo no mercado interno.

O fato é que o manejo sustentável de florestas, em especial na Amazônia, constitui um dos maiores e mais importantes desafios para o Brasil enfrentar nos anos por vir. Não se trata

de um desafio unicamente ligado a aspectos ambientais. Mais do que isso, ele envolve também questões econômicas, sociais, geopolíticas e tecnológicas. Qualquer que seja a abordagem para esse desafio, ela deverá contemplar todas essas dimensões.

Uma das maneiras possíveis de se propor soluções para problemas de ordem múltipla é a análise integrada de suas partes, com vistas ao planejamento de ações de caráter abrangente. Em uma região com as características da Amazônia brasileira, tanta complexidade não pode prescindir do aporte da tecnologia. Esta tese aborda essa dimensão tecnológica da questão apresentada.

O PROBLEMA INVESTIGADO

O desmatamento da região amazônica se dá a uma taxa média de desmatamento de aproximadamente 19000 km²/ano - considerado o período de 1988 a 2003 - e possui diferentes dinâmicas, em função de fatores sociais e econômicos que o impulsionam.

Se posto de forma simples, ele ocorre como consequência de queimadas e de derrubadas de árvores. Dada a sua natureza, tais eventos podem ser facilmente monitorados por intermédio de boa parte dos satélites imageadores atualmente disponíveis, dada a combinação das resoluções⁴ espacial, espectral e temporal desses sistemas orbitais.

Ainda que úteis para inventariar os danos ambientais provocados pelo desmatamento, as imagens orbitais produzem uma visão bastante estática do processo de mudança na cobertura do solo⁵. Detectar uma queimada ou uma grande área desmatada simplesmente traduz o achado de um desastre ambiental, com pouco ou nada a fazer a respeito, a não ser as necessárias ações punitivas.

Identificar tendências de desmatamento, no entanto, traz consigo o mérito de possibilitar posturas preventivas, conquanto sejam tomadas providências adequadas e oportunas. Buscar uma forma de qualificar, quantificar e localizar as mudanças na cobertura do solo provocadas por mudanças no uso do solo, criando algo como um mapa de tendências, é ainda mais interessante sob o ponto de vista da gestão ambiental.

⁴ - Aqui, a menção do conceito de resolução, quando aplicada a sistemas sensores, engloba os conceitos de resolução espacial, espectral e temporal.

⁵ - O termo 'solo', para os fins deste trabalho, possui o mesmo significado que 'terra'. Para uma distinção dos conceitos de cobertura e de uso do solo, ver Cap. 1 - 'Arcabouço Teórico da Mudança no Uso e na Cobertura do Solo'.

Por questão de escala, a dinâmica do desmatamento é mais bem compreendida por meio de inspeções feitas a partir da superfície. Naturalmente, a Amazônia brasileira, dadas as suas dimensões, torna esse processo algo impraticável. Assim, o uso de satélites imageadores constitui uma atraente solução para essa tarefa, em função do baixo custo/benefício decorrente de sua visão sinótica. Já o uso de aeronaves, como plataforma de sensores imageadores, pode prover uma solução intermediária, em termos de precisão, mas de alcance relativamente curto, em se tratando de Amazônia, dado o elevado custo envolvido em eventuais acionamentos sistemáticos dessas aeronaves.

O objeto e a hipótese da pesquisa

Com base nas considerações feitas, buscou-se uma forma de estabelecer uma relação de dependência, em uma área de estudo situada no Sul do Estado do Pará, município de São Félix do Xingu, entre o aspecto mais aparente da mudança no uso do solo – o avanço do vetor de desmatamento (objeto de pesquisa) – e a combinação de fatores biofísicos.

A partir dos aspectos relativos ao objeto de pesquisa, busca-se um melhor entendimento da complexidade dos processos de desmatamento na região de estudo. Para tanto, um modelo espacial multivariado de tendências associadas a mudanças na cobertura do solo é testado. Esse modelo, composto de variáveis independentes que caracterizam o desmatamento de maneira espacialmente explícita, visa testar a seguinte hipótese:

O desmatamento, em especial aquele estimulado pela pecuária extensiva no Sul do Pará, possui uma relação de dependência direta com fatores de ordem biofísica.

Tal relação pode ser representada pela função logito de regressão logística, que indica a probabilidade de uma célula do terreno estar desmatada, dadas algumas variáveis indutoras (explicativas), conforme se segue:

$$\text{Logito}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

onde P é a probabilidade de ocorrência de desmatamento para uma dada célula e X_i são as variáveis indutoras desse desmatamento. Os valores de β_i representam os parâmetros desconhecidos de cada variável.

Hipóteses secundárias

As hipóteses listadas a seguir dão suporte à hipótese principal e tratam mormente das variáveis que são objeto do modelo apresentado nesta tese.

- As práticas adotadas na atividade de desmatamento são desenvolvidas segundo um padrão espacial que, por sua vez, é ditado pelas características geográficas da região explorada;
- A intensidade da atividade de desmatamento é inversamente proporcional à distância das células florestais para as áreas desmatadas, com as rodovias e com as hidrovias, bem como é inversamente proporcional à declividade e à altura do terreno;
- Existe uma tendência de redução da atividade de desmatamento quando a célula florestal está localizada em uma unidade de conservação (UC) ou área indígena; e
- A intensidade da atividade de desmatamento é diretamente proporcional ao índice de fragmentação da célula florestal e inversamente proporcional à área de seu polígono envolvente.

VARIÁVEIS

As hipóteses formuladas são testadas por meio das variáveis a seguir:

Fatores de ordem biofísica

Variável **teórica** composta das seguintes variáveis **empíricas**:

- Proximidade da área desmatada – distância da célula até a próxima área desmatada;
- Distância de rodovias – distância da célula até a rodovia mais próxima;
- Distância de hidrovias – distância da célula até o rio mais próximo;
- Existência de área de proteção ambiental – *status* da área em torno da célula;
- Declividade – declividade do terreno na posição da célula;
- Altitude – altitude do terreno na posição da célula;
- Índice de fragmentação florestal – relação entre perímetro e área da célula; e
- Área do polígono envolvente – área do polígono em torno da célula.

Impacto ambiental

Variável **teórica** formada pela seguinte variável **empírica**:

- Cobertura do solo – classificação dicotômica (floresta/desmatamento) da célula.

UNIVERSO GEOGRÁFICO E TEMPORAL

Como forma de incorporar a dimensão temporal ao objeto de estudo, são analisadas imagens que permeiam praticamente toda a existência do satélite Landsat 5, em um período de cerca de 20 anos, com amostragens desde 1985 até 2004. A essas imagens, somam-se dados de um trabalho de campo realizado em 2005.

A área de estudo corresponde à “órbita/ponto” 225/64 do satélite Landsat 5 (cena de 185km x 185km), localizada no município de São Félix do Xingu, região Sul do estado do Pará. Trata-se de uma área situada no “Arco do Desmatamento”, onde há uma intensa atividade madeireira.

PROCEDIMENTOS METODOLÓGICOS

O método (desenho de pesquisa) adotado é a pesquisa experimental. À exceção da revisão bibliográfica, a pesquisa se baseia na obtenção de dados primários. Esses dados formam as variáveis empíricas que, por sua vez, são consolidadas em variáveis teóricas, usadas para testar as hipóteses formuladas.

As imagens do satélite Landsat servem de base de dados para a construção do modelo de regressão logística. Além disso, foi agregado um trabalho de campo com a finalidade de prover observações diretas da área de estudo.

Para conduzir operações de classificação e de geoprocessamento dos dados coletados, optou-se por utilizar, respectivamente, os *softwares* SPRING 4.3.2, desenvolvido pelo INPE, e o ArcGIS 9.2, da ESRI.

De posse dos dados de campo e das imagens, ambos foram georreferenciados a uma mesma base cartográfica, com vistas ao início da atividade de processamento de imagens. A partir daí, um modelo preditivo de regressão foi concebido, de maneira a gerar um mapa de probabilidades das áreas candidatas ao desmatamento, em função das características de variáveis explicativas identificadas nas etapas de coleta de dados e imagens. O grau de aderência do modelo é verificado por meio de tabelas de classificação.

ESTRUTURA DO TRABALHO

Este trabalho está organizado em quatro capítulos. No primeiro, é apresentado um apanhado das principais linhas teóricas e de modelagem, naquilo que envolve o estudo do uso e da cobertura do solo, especialmente em ambientes tropicais, como forma de contextualizar o leitor em termos do arcabouço teórico que fundamenta esse tipo de estudo no Brasil e no mundo. Além disso, o contato com esse universo de teorias e de modelos permite um melhor entendimento acerca das opções feitas na construção do modelo concebido.

No segundo capítulo, é feita uma revisão teórica dos modelos de regressão, em geral, e do modelo de regressão logística, em particular, que serve de base a este estudo. Embora não constitua a única maneira de abordar o tema, a regressão logística se mostra uma ferramenta bastante interessante sob vários aspectos. O principal deles está em sua capacidade de fácil implementação de modelagens complexas.

A caracterização dos sistemas empregados no estudo, bem como dos procedimentos metodológicos para a coleta e a análise dos dados, é o tema do terceiro capítulo. Em suma, dedica-se a explicar as técnicas de sensoriamento remoto e de geoprocessamento utilizadas na concepção do estudo.

O quarto capítulo divide-se entre a formulação e a implementação do modelo avaliado no estudo. Aos preceitos tradicionais dos modelos de regressão logística múltipla, é acrescentada uma abordagem espaço-temporal. Nessa abordagem, são utilizadas cenas de diferentes datas de aquisição, permeadas por técnicas de geoprocessamento, ambientadas em um sistema de informação geográfica. Aqui são também analisados os resultados alcançados e discutidos os desempenhos das diversas variáveis utilizadas.

1 - ARCABOUÇO TEÓRICO DA MUDANÇA NO USO E NA COBERTURA DO SOLO

Ainda que similares, os conceitos ‘cobertura do solo’⁶ e ‘uso do solo’ não são equivalentes. De acordo com Turner *et al.* (1994), cobertura do solo compreende a caracterização do estado físico, químico e biológico da superfície terrestre (floresta, gramínea, água, área construída). Uso do solo, por sua vez, se refere aos propósitos humanos associados àquela cobertura (pecuária, recreação, conservação, moradia, indústria etc.).

Da mesma forma que uma classe de cobertura pode suportar usos diversos, um único sistema de uso pode incluir várias coberturas. No entanto, ainda que mudanças no uso do solo acarretem, via de regra, mudanças na cobertura do solo, nem sempre a recíproca é verdadeira. Em outras palavras, podem ocorrer modificações na cobertura, sem que isso signifique alterações no seu uso.

Preocupações com eventuais impactos ambientais e socioeconômicos das mudanças de uso e de cobertura do solo têm ensejado inúmeras iniciativas no sentido de entender os processos aí envolvidos; diagnosticar regiões de maior incidência de mudanças, a fim de antecipar áreas de risco; e analisar os impactos de tais mudanças, com vistas à proposição de ações pertinentes.

Em linhas gerais, há um consenso em torno do entendimento de que o conhecimento integrado dos processos de mudança de uso e dos fatores que afetam tais processos possibilitará a elaboração de modelos preditivos confiáveis de médio e de longo prazo, em diferentes escalas.

No que diz respeito ao desmatamento atualmente em curso na Amazônia, existe um padrão linear de evolução concentrado em cerca de 50 km de cada lado das rodovias. Algumas constatações importantes sobre o padrão espacial de ocupação são apresentadas por Alves (2001, citado por AGUIAR, 2003):

- Uma concentração importante do desmatamento está em áreas próximas aos eixos e pólos de desenvolvimento e às áreas desflorestadas ainda na década de 1970.

6 - As traduções mais fiéis de *land*, *land cover* e *land use* são, respectivamente, terra, cobertura da terra e uso da terra. Entretanto, esta tese adota os termos ‘solo’, ‘cobertura do solo’ e ‘uso do solo’, acompanhando a prática em voga na comunidade científica brasileira.

Isso é, o processo mais intenso parece ocorrer nas vizinhanças de áreas já abertas, alargando as áreas derrubadas, levando à redução contínua e, em alguns casos, ao esgotamento das reservas legais.

- Os eixos que oferecem acesso mais fácil ao Sul e Centro-Oeste têm concentrado a maior parte do desmatamento. Nota-se que menos de 11% do desmatamento ocorreu na Transamazônica a oeste de Marabá, enquanto mais de 85% das taxas foram observadas ao redor dos corredores que ligam Belém ao Sul do Pará, Mato Grosso e Centro-Oeste; Cuiabá ao Norte Matogrossense e Santarém; e de Cuiabá a Porto Velho e Rio Branco.

O autor conclui que o desmatamento teve incentivo inicial importante com a implantação dos eixos e pólos de desenvolvimento, mas permaneceu elevado em regiões que estabeleceram sistemas produtivos encadeados no Centro-Oeste, Sul e Nordeste, como é o caso dos flancos Oriental, Sul e Sudeste da Amazônia, onde se concentra a maior parte do desmatamento.

O uso de recursos naturais pelo homem implica forçosamente o uso do solo, seja qual for sua finalidade (produção de alimentos, mineração, ocupação, recreação etc.). De acordo com Briassoulis (2000), o uso do solo ocorre sob a influência de duas grandes forças: (1) necessidades humanas; e (2) aspectos e processos ambientais.

Ambas as forças são inerentemente dinâmicas, como dinâmica é a natureza da vida. Como consequência, decorre que a externalidade – positiva ou negativa – mais notável dessas forças é constituída pelas mudanças no uso do solo. Não por acaso, os efeitos provocados por tais mudanças, bem como a expectativa deles, têm sido objeto de atenção da sociedade – leiga e científica – desde tempos ancestrais.

As mudanças mais preocupantes no uso do solo estão associadas à conversão de lavouras e à abertura de florestas, em função das implicações socioeconômicas e ambientais.

Dada a multiplicidade de fatores envolvidos, o estudo desse assunto não se resume a alguns poucos corpos de conhecimento científico. Ao contrário, sua compreensão absoluta depende de uma abordagem transdisciplinar. Por conta disso, teorias de mudança no uso e na cobertura do solo podem surgir como derivativas de teorias mais amplas sobre mudanças socioeconômicas e ambientais. Em suma, não há meios de se estudar diretamente as mudanças no uso do solo.

1.1 - O ESTUDO DAS MUDANÇAS NO USO E NA COBERTURA DO SOLO

Os últimos três séculos foram marcados por uma evolução marcante nos impactos provocados por mudanças no uso do solo. Problemas como a desertificação⁷, eutrofização⁸, acidificação⁹ e a perda de biodiversidade estão entre os mais divulgados e são influenciados, em maior ou menor grau, pela ação humana. (BRIASSOULIS, 2000)

Como forma de precisar essa ação, o Programa Internacional da Geosfera-Biosfera (IGBP) e o Programa Internacional de Dimensões Humanas (IHDP) representam as principais iniciativas científicas, em termos globais. Sob os auspícios desses dois programas, em 1993, foi criado o Projeto LUCC¹⁰ (programa de pesquisas voltadas a mudanças no uso e na cobertura do solo).

As abordagens adotadas para a análise de mudança no uso e na cobertura do solo estão criticamente relacionadas aos objetivos do analista. O propósito dessas análises pode ser agrupado, segundo Briassoulis (2000), em seis categorias: descrição, explanação, predição, investigação do impacto, prescrição e avaliação.

Estudos descritivos de mudança no uso e na cobertura do solo têm fornecido o ímpeto para investigações mais profundas acerca do “porquê” dessas mudanças, bem como para a tomada de ações de contraposição aos impactos negativos das mudanças identificadas.

A descrição, entretanto, não é suficiente para prover a base para o entendimento das mudanças observadas, ou ainda para guiar políticas e tomadas de decisão relativas a maneiras efetivas de se lidar com as implicações adversas dessas mudanças. As análises explanatórias buscam preencher esse vazio. O nível de explanação oferecida é função do nível de análise espacial e temporal.

Além de descrever e explicar mudanças no uso e na cobertura do solo, um importante propósito para conduzir essas análises é o de predizer futuras mudanças. Predições podem ser

⁷ - Processo de degradação do solo, que leva à drástica redução de sua produtividade. Ocorre com maior frequência nas áreas de clima árido e semi-árido e decorre tanto de razões naturais, como de antropismos.

⁸ - Processo de enriquecimento da água em lagos, rios, estuários, mares etc. com nutrientes (carbono, enxofre, potássio, cálcio, magnésio, nitrogênio e fósforo) que leva ao crescimento orgânico acelerado, com conseqüências indesejadas. Estas incluem: infestações por algas, mudanças de coloração das águas e mau cheiro.

⁹ - Processo de mudança na acidez de um meio ambiente (água, solos e ecossistemas). Caracteriza-se pela redução do pH natural para valores inferiores aos normalmente encontrados nesses meios, de 5,5 a 7,0. A acidificação está normalmente associada a corpos d'água, solos e ecossistemas florestais.

¹⁰ - O Projeto LUCC (*Land Use and Cover Change*) é um dos elementos do IGBP e do IHDP. (BRIASSOULIS, 2000)

incondicionais ou condicionais. Predições incondicionais, categoria na qual se enquadra a abordagem apresentada nesta tese (voltada especificamente a mudanças na cobertura do solo), também conhecidas por extrapolações de tendência, fornecem imagens dos padrões de mudança no uso e na cobertura do solo em uma dada área, caso tendências passadas perdurem no futuro. As predições condicionais, por sua vez, produzem alternativas futuras de uso do solo, sob condições ou cenários hipotéticos. Nos dois casos, o fator crítico é novamente função do nível de análise espacial e temporal.

Outro importante propósito da análise de mudança no uso e na cobertura do solo é a investigação de impacto. O interesse contemporâneo não se dá tanto na mudança no uso e na cobertura do solo em si, mas nos diversos impactos ambientais e socioeconômicos em todos os níveis espaciais. O interesse recente, especificamente, está no aspecto mais amplo da sustentabilidade¹¹ do desenvolvimento, na medida em que este é afetado pela mudança no uso e na cobertura do solo decorrente da implementação de políticas ambientais.

Em uma perspectiva normativa, a análise de mudança no uso e na cobertura do solo pode buscar esclarecer a questão de ‘dever ser’; em outras palavras, o propósito é prescrever configurações de uso do solo que assegurem o alcance de objetivos particulares. Atualmente, esses objetivos estão agrupados na ampla busca por “soluções sustentáveis de uso do solo”. O propósito desse tipo de análise é indicar padrões de uso que sejam associados à preservação ambiental, à prosperidade econômica e à equidade social.

Finalmente, a análise de mudança no uso e na cobertura do solo pode ser conduzida para avaliar passado, presente ou futuras (politicamente orientadas) mudanças em padrões de uso do solo, em termos de certos critérios, tais como deterioração (ou melhoria) ambiental, declínio econômico, ou empobrecimento social. Os resultados dessas avaliações podem ser usados para sugerir alternativas de uso do solo capazes de contribuir com a obtenção desses objetivos.

Independentemente do seu propósito, uma análise segura e consistente de mudança no uso e na cobertura do solo requer que certos pré-requisitos sejam satisfeitos. Especificamente, que os termos básicos na análise sejam claramente definidos, que sejam utilizados sistemas de classificação compatíveis com o propósito da análise, que teorias válidas enquadrem a análise, e que as técnicas analíticas usadas representem realmente a mudança no uso e na cobertura do solo que estiver em estudo.

¹¹ - Propriedade de um sistema em manter suas funções e sua produtividade de maneira constante no tempo. O termo ‘desenvolvimento sustentável’ advém do conceito de sustentabilidade, mas considera também as condições nas quais a sustentabilidade pode ser alcançada. De maneira sucinta, são elas: eficiência econômica, proteção ambiental e justiça social. (BRIASSOULIS, 2000)

Tanto no caso de cobertura do solo como no de uso do solo, o significado e a conceituação de mudança são bastante amplos. No caso de cobertura do solo, a literatura distingue entre dois tipos de mudança: conversão e modificação. A conversão da cobertura do solo envolve uma mudança de um tipo de cobertura para outra. Já a modificação envolve alterações estruturais ou funcionais, sem uma mudança total de um tipo para outro de cobertura (mudanças em produtividade, biomassa, ou fenologia¹²).

Meyer e Turner (1996) sugerem que “o uso do solo (seja ele deliberado ou não) altera a cobertura do solo de três maneiras: convertendo a cobertura do solo, ou mudando-a para um estado qualitativamente diferente; modificando a cobertura do solo, ou mudando quantitativamente sua condição, sem que haja uma conversão completa; e mantendo a cobertura do solo em sua condição, contrariamente aos agentes naturais de mudança”.

De maneira análoga, a mudança no uso e na cobertura do solo pode envolver a conversão de um tipo para outro ou a modificação de um certo tipo de uso do solo. A modificação pode implicar mudanças na intensidade de seu uso, da mesma forma que alterações nas características de suas qualidades e atributos.

A razão pela qual a ligação entre mudança no uso e na cobertura do solo é enfatizada está no fato de que os impactos ambientais provocados por mudanças no uso do solo geralmente implicam mudanças na cobertura do solo. Assim sendo, o exame das maneiras pelas quais o uso do solo se relaciona com a mudança na cobertura do solo, em vários níveis de detalhamento espacial e temporal, deve preceder a análise de ambas.

A especificação dos níveis de detalhe espacial e temporal é de crucial importância, na medida em que: (a) guia a seleção dos tipos de uso e de cobertura do solo a serem analisados; (b) determina os direcionadores em processos de mudança que podem ser detectados; e (c) afeta a identificação e a explanação das ligações entre uso e cobertura do solo em um contexto espaço-temporal específico. (BRIASSOULIS, 2000)

A análise de mudança no uso e na cobertura do solo gira em torno de duas questões centrais inter-relacionadas: o que direciona/causa a mudança no uso e na cobertura do solo; e quais são os impactos (ambientais e socioeconômicos) da mudança no uso e na cobertura do solo. Este trabalho trata somente da primeira delas, ou seja, dos direcionadores de mudanças na cobertura do solo.

¹² - Estudo dos aspectos temporais de fenômenos naturais recorrentes. (BRIASSOULIS, 2000)

Existe uma quase unanimidade de que há duas categorias de direcionadores: biofísicos e socioeconômicos. Os direcionadores biofísicos incluem características e processos do ambiente natural, enquanto os direcionadores socioeconômicos compreendem demografia, fatores sociais, econômicos, políticos e institucionais, entre outros.

Deve-se ter em mente que os direcionadores biofísicos normalmente não causam mudança no uso e na cobertura do solo diretamente. Na maior parte das vezes, eles induzem a mudança na cobertura do solo, que, por sua vez, pode influenciar as decisões dos proprietários dessas terras.

É comum a assunção de que todos os impactos de mudanças no uso do solo sejam negativos. Isso não é sempre verdade, por duas razões. Em primeiro lugar, o caráter positivo ou negativo de um impacto depende das escalas espacial e temporal envolvidas. Segundo, forças mitigadoras implementadas pelo homem, se devidamente conduzidas, podem impedir as influências negativas dessas mudanças.

No entanto, a questão relacionada aos impactos da mudança no uso e na cobertura do solo de caráter mais abrangente diz respeito à sustentabilidade do desenvolvimento em todos os níveis espaciais. Nesta perspectiva, o planejamento e o gerenciamento do uso do solo tornam-se imperativos. O objetivo maior, que propicia a gestão do uso do solo e suas mudanças, é desenvolver os recursos terrestres de maneira a capitalizar potenciais e conveniências locais, evitar impactos negativos e responder às demandas sociais atuais e futuras, dentro dos limites suportáveis pelo meio ambiente. (FAO, 1995)

A análise de mudança no uso e na cobertura do solo possui uma dependência crítica do sistema de classificação adotado. A magnitude e a qualidade da mudança no uso e na cobertura do solo é expressa em termos de tipos específicos de usos e coberturas do solo. A avaliação dos impactos ambientais e socioeconômicos somente é possível quando as feições ambientais e socioeconômicas das coberturas e usos do solo são especificadas. Se esse requisito não é atendido, a análise passa a ter valor limitado, no sentido de servir de guia de políticas ambientais, especialmente em escalas maiores. (WOLMAN, 1987)

A escala espacial determina o nível de detalhamento ambiental e socioeconômico contido no sistema de classificação, enquanto o propósito do estudo determina os atributos particulares dos tipos de uso do solo considerados. Além disso, a tecnologia de coleta de dados disponível é determinante para a estrutura e para o conteúdo desse sistema.

O desenvolvimento de sistemas de classificação da superfície terrestre está cada vez mais sendo influenciado pela disponibilidade e pelo uso de dados de satélites e de técnicas de sensoriamento remoto. Essas mudanças tecnológicas estimulam a utilização direta dos dados, em conjunção com as técnicas de análise e os modelos disponíveis. As três abordagens genéricas – ou “perspectivas de entendimento” (LUCC, 1999) - usualmente adotadas para o estudo de mudança no uso e na cobertura do solo são:

- Narrativa – busca aprofundar o entendimento da mudança no uso e na cobertura do solo por meio do detalhamento histórico e da interpretação;
- Baseada em agentes – visa compreender a natureza genérica e as regras do comportamento individual dos agentes em suas tomadas de decisão diárias;
- Sistêmica – busca o entendimento das instituições e organizações da sociedade que estabelecem as oportunidades e restrições à tomada de decisão.

As duas últimas abordagens – baseadas em agentes e sistêmicas – dependem do desenvolvimento de modelos explícitos e de testes empíricos.

LUCC (1999) enfatiza, ainda, que o estudo de mudanças no uso e na cobertura do solo deve primar pela integração de três tradições epistemológicas, a saber:

- “Observar e descrever para compreender” – abordagem indutiva;
- “Modelar para compreender” – abordagem dedutiva;
- “Integrar para compreender” – abordagem dialética.

George Perkins Marsh, nos Estados Unidos, e J. H. von Thunen, na Alemanha, estão entre os pioneiros mais conhecidos no estudo da mudança no uso e na cobertura do solo. O assunto ‘uso do solo’ possui um caráter central no trabalho de Marsh, dado que todas as atividades humanas ocorrem e modificam o espaço para fins particulares. Marsh enfatizou a amplitude e a gravidade dos impactos humanos não intencionais, e assim, a necessidade de entender as complexas interações dos processos naturais anteriores à intervenção humana. Ele também via a transformação do meio ambiente, se devidamente conduzida, como algo desejável. (KATES *et al.* 1990, 3).

Em 1826, J. H. von Thunen desenvolveu uma teoria (ou modelo) mais genérica de como o uso de terras rurais deveria ser arranjado em torno de um mercado urbano. O princípio básico era que cada porção do solo deveria ser dedicada ao uso no qual ele poderia prover a maior renda. Von Thunen via o solo como um recurso econômico, cujo principal atributo a considerar era a produtividade. Deve-se ter em mente que a paisagem, para a qual a atividade

agrícola era considerada, era plana e uniforme em todas as direções. Seu propósito era utilitarista e a análise do padrão “ótimo” de uso do solo, estática. O utilitarismo é uma abordagem filosófica, cujo critério moral para a ação humana é o interesse pessoal; escolhas éticas são feitas com base em benefícios pessoais. O lema do utilitarismo é: “o maior benefício para o maior número de pessoas”. (HOOVER e GIARRATANI 1984, 142-143)

Esses estudos pioneiros representam abordagens pioneiras e diametralmente opostas para o estudo da mudança no uso e na cobertura do solo.

A visão abrangente de Marsh em relação ao solo, ao meio ambiente e ao papel do homem como causador de mudanças ambientais está no centro das teorias sobre a relação natureza-sociedade e dos modelos integrados propostos nos anos que se sucederam e que continuam em voga ainda hoje.

Já o homem econômico e racional de Von Thunen, que produz bens em uma paisagem ordenadamente uniforme e estática, cuja mudança não é o assunto central, fundou as teorias e os modelos das principais correntes econômicas - urbanas e regionais - no Século XX.

As primeiras décadas do Século XX viram mudanças significativas nos usos do solo, provocadas pela industrialização e pela urbanização no mundo ocidental, pelas duas guerras mundiais, bem como por outros eventos socioeconômicos importantes e pelo progresso tecnológico.

De acordo com Briassoulis (2000), o estabelecimento da análise sistemática e “científica” de mudança no uso e na cobertura do solo baseou-se em teorias e modelos fundamentados em uma variedade de áreas científicas, em especial a economia, a sociologia e a geografia. A análise científica e os estudos de mudança no uso e na cobertura do solo prosperaram depois da Segunda Guerra Mundial, com base nas linhas das várias abordagens que tinham sido formuladas anteriormente.

Os números e a diversidade dos estudos existentes tornam uma avaliação completa e abrangente algo virtualmente impossível. Entretanto, a proliferação de estudos e as linhas adotadas para a análise das mudanças no uso e na cobertura do solo não ficaram desconectadas das mudanças teóricas e metodológicas mais amplas nas disciplinas que contribuíram para esses estudos, tampouco das tecnologias de suporte necessárias.

A mais importante dessas mudanças talvez seja a revolução da geografia denominada “quantitativa”, ocorrida nas décadas de 1950 e de 1960. Naquele período, foram propostos teorias e modelos formais de uso do solo e de mudança no uso e na cobertura do solo, que

mais tarde viriam a ser rejeitados - mas não abandonados - quando suas limitações ficaram evidentes e suas fundamentações epistemológicas foram seriamente questionadas.

O recente interesse político nas implicações (negativas) da mudança global – que tem na mudança no uso e na cobertura do solo um de seus componentes - pode constituir uma influência na orientação dos estudos de mudança no uso e na cobertura do solo. Tal influência pode ocorrer na forma de abordagens práticas e instrumentos de apoio à decisão, buscados para conduzir a realização de políticas voltadas ao uso sustentável do solo.

Johnston *et al.* (1994) observam que as análises orientadas economicamente para a mudança no uso e na cobertura do solo possuem características comuns. A mais importante delas é a ênfase no mecanismo de preço (custos imobiliários e de transporte) como o fator determinante para a localização das atividades humanas no espaço. Elas são funcionalistas¹³ e buscam tanto descrever (direta ou indiretamente) padrões de uso do solo e suas mudanças, como prescrever configurações ótimas de uso do solo que satisfaçam objetivos estabelecidos.

As linhas de estudo orientadas sociologicamente continuaram a tradição da ecologia humana, desenvolvida na primeira metade do Século XX, produzindo estudos quantitativos e empíricos da estrutura espacial e social urbana, especialmente nas décadas de 1960 e 1970.

Uma característica notável de todas essas abordagens sociológicas é que elas não fazem nenhuma referência explícita ao real uso do solo e suas mudanças, dentro do contexto estudado das relações sociais causais. Em termos práticos, em sua forma e orientação atuais, essas abordagens são capazes de prover uma análise muito limitada de mudança no uso e na cobertura do solo.

Além das abordagens orientadas social e economicamente, foram desenvolvidas, na segunda metade do Século XX, várias outras voltadas ao estudo de uso do solo e de suas mudanças. Elas combinam elementos das ciências naturais e sociais e são baseadas na noção de equilíbrio ecológico, que atribui mudanças em uma região a mudanças na interação dinâmica de quatro fatores: população, recursos, tecnologia e instituições (COCCOSSIS, 1991; MEYER e TURNER, 1996).

Abordagens teóricas baseadas em ecossistemas e modelos integrados de ambiente-economia-sociedade tornaram-se bastante difundidos na segunda metade do Século XX,

¹³ - Perspectiva analítica na qual o mundo é visto como um jogo de sistemas interdependentes. Suas ações e relações coletivas refletem regularidades repetíveis e previsíveis, nas quais forma e função podem ser relacionadas. (BRIASSOULIS, 2000)

especialmente depois dos anos setenta. O uso do solo e suas mudanças passaram a ser reconhecidos como elementos importantes do contexto mais amplo do sistema natureza-sociedade e como contribuintes não triviais para a mudança ambiental global (SLOCOMBE, 1993; LUTZ, 1994; MANNING, 1991).

O que distingue estas abordagens dos dois grupos anteriores é o tratamento dado aos conceitos de ‘solo’ e ‘uso do solo’ como dotados de propriedades ambientais (e não só econômicas e socioculturais), capacidades e atributos intrínsecos e variáveis, que influenciam e são influenciadas por atividades e ações humanas.

O progresso tecnológico no domínio da gestão de dados (espaciais) e na área de sensoriamento remoto motivou a criação de grandes projetos voltados à observação e ao registro de mudanças no uso e na cobertura do solo. Satélites de observação da Terra, aliados a sistemas avançados de gestão de dados espaciais (principalmente os Sistemas de Informação Geográfica - GIS¹⁴) oferecem a possibilidade de monitorar e mapear a cobertura do solo, ainda que não necessariamente o uso do solo. (LIVERMAN *et. al.*, 1998)

Mudanças no uso e na cobertura do solo não constituíram exceção nos impactos provocados pelo relatório Brundtland¹⁵ ou pela Convenção Rio’92. O tema foi reconhecido como um componente significativo do ambiente global e iniciativas de pesquisa específicas, como o Lucc, foram formadas. Ao mesmo tempo, as áreas científicas que contribuem para a análise de mudanças no uso e na cobertura do solo amadureceram um pouco mais em termos de teorias, modelos e tecnologia. (KATES *et. al.*, 1990)

Como resultado desse amadurecimento, as abordagens são mais holísticas do que eram no passado. Apesar da persistência e da inércia de limites disciplinares fortes, novas formas de cooperação científica são promovidas sob o signo da “transdisciplinaridade”.

Do ponto de vista epistemológico, a maior parte das abordagens de análise de mudança no uso e na cobertura do solo se move, em maior ou menor grau, ao longo de linhas empíricas e positivistas (aplicações de sensoriamento remoto e de GIS, modelos integrados e abordagem neoclássica de maximização do bem-estar econômico).

¹⁴ - **GIS** – *Geographical Information Systems* – Sistemas de Informação Geográfica. Dado o uso da sigla ‘SIG’ para caracterizar o termo ‘Sistemas de Informação Gerencial’ no campo da Administração, optou-se por manter a sigla originalmente cunhada na língua inglesa, já consagrada mundialmente.

¹⁵ - O relatório, elaborado em 1987 sob a coordenação de Gro Harlem Brundtland, ministra da Suécia, tornou-se célebre por ter criado o conceito de desenvolvimento sustentável - aquele que atende às necessidades do presente sem comprometer a possibilidade de as gerações futuras atenderem a suas próprias necessidades.

Estudos de mudança no uso e na cobertura do solo cobrem o espectro que se estende do panorama global ao local. Estudos de mudança no uso e na cobertura do solo em nível global receberam maior publicidade, comparados aos outros níveis, dado o interesse geralmente mais forte na mudança ambiental global. Porém, é reconhecido que muitos dos impactos globais provocados por mudanças no uso do solo resultem principalmente das decisões e das ações locais dos verdadeiros usuários do solo. Como consequência, cresce o interesse na integração entre escalas espaciais, nas análises que envolvem os níveis local, regional e global, bem como nas análises que levem em conta questões rurais e urbanas.

Uma nova área de conhecimento oriunda da ecologia é a ecologia de paisagens, caracterizada por duas abordagens distintas e, por vezes conflitantes. A primeira delas é a geográfica, que enfatiza a influência do homem sobre a paisagem e a gestão do território. A segunda abordagem é a ecológica, focada na importância do contexto espacial sobre os processos ecológicos e as implicações disso em termos de conservação biológica.

Metzger (2001) propõe que a paisagem seja definida como “um mosaico heterogêneo formado por unidades interativas, sendo esta heterogeneidade existente para pelo menos um fator, segundo um observador e numa determinada escala de observação”.

O autor argumenta que, dentro da abordagem geográfica da ecologia de paisagens, o mosaico heterogêneo é visto pelos olhos do homem, a partir de suas necessidades, anseios e planos de ocupação territorial. Como o homem age em amplas extensões de seu território, neste caso a ecologia de paisagens lida obrigatoriamente com escalas espaciais amplas e, por conseguinte, temporais. O conjunto interativo da paisagem é composto por ‘ecossistemas’ ou por unidades de ‘cobertura’ ou de ‘uso e ocupação do território’, cabendo ao observador a escolha de uma dessas três formas de representação das unidades da paisagem.

Na abordagem ecológica, Metzger (2001) considera o mosaico como um conjunto de habitats que apresentam condições mais ou menos favoráveis para a espécie ou a comunidade estudada. Assim, o olhar sobre a paisagem leva em conta essas espécies e suas características biológicas, em particular seus requisitos em termos de área de vida, alimentação, abrigo e reprodução. Dentro desta perspectiva, a escala espaço-temporal de análise não é necessariamente ampla, como na abordagem anterior, pois ela depende da escala de ação e de percepção da espécie ou comunidade em estudo.

Em suma, a ecologia de paisagens lida com mosaicos submetidos a alguma forma de interferência humana, observada a escala na qual o homem modifica seu ambiente.

1.2 - RELAÇÃO ENTRE TEORIAS E MODELOS DE ESTUDO DA MUDANÇA NO USO E NA COBERTURA DO SOLO

É ampla a busca por teorias e modelos de mudança no uso e na cobertura do solo que possam oferecer orientações essenciais e significativas para o entendimento da mudança no uso e na cobertura do solo e para a tomada de decisão voltada a usos sustentáveis futuros, em vários contextos diferentes.

Cada tradição teórica é especializada, mais ou menos, em um dado nível espacial, que determina a natureza e a ênfase colocada nos comportamentos do sistema estudado.

Em níveis mais detalhados, a teorização é normalmente mais concreta e considera de maneira mais realística os agentes, o contexto e os mecanismos de mudança. Em níveis mais abrangentes, a teorização é mais abstrata e a passagem da teoria para o mundo real não é simples e direta.

O papel da tradição teórica é crítico no que diz respeito à identificação dos indutores da mudança no uso e na cobertura do solo. Algumas teorias enfatizam questões econômicas e sócio-políticas, enquanto outras, os determinantes ambientais da mudança no uso e na cobertura do solo.

A possibilidade de uma teoria geral de mudança no uso e na cobertura do solo é uma questão aberta. Aparentemente, uma abordagem sensata para a formulação de um arcabouço teórico geral para a mudança no uso e na cobertura do solo é buscar uma síntese das teorias existentes, utilizando-se cada uma delas no nível espaço-temporal de detalhes mais adequado. (BRIASSOULIS, 2000)

Os modelos de mudança no uso e na cobertura do solo constituem um universo diferente, em termos de propósito e objeto de estudo, nível de agregação, dinâmica, teoria subjacente, especificação funcional, requisitos de dados e aplicações práticas. Embora a mudança no uso e na cobertura do solo implique automaticamente o conceito de tempo, modelos dinâmicos são raros para resoluções espaciais e funcionais referentes às situações mais práticas e relevantes. As dificuldades em se construir modelos verdadeiramente dinâmicos não são somente técnicas, mas também teóricas.

A especificação e a interpretação dos resultados de um modelo dinâmico requerem uma teoria de mudança correspondente. Apesar da diversidade de formas funcionais dos

modelos, a maioria deles adota formas funcionais simples – modelos de programação estatística ou linear – ou baseia-se em técnicas heurísticas¹⁶ (simulação).

A ligação entre teorias e modelos de mudança no uso e na cobertura do solo não se tem mostrado forte ao longo do tempo. Em vários casos, teorias e modelos foram desenvolvidos de maneira independente. Assim sendo, nem as teorias conduziram aos modelos, nem os modelos conduziram às teorias. Quando comparados, pode-se dizer que os modelos se encontram mais desenvolvidos. As razões para isso podem ser divididas em substantivas e práticas.

Razões substantivas pertencem às dificuldades associadas com a formulação de teorias para um fenômeno tão complexo como a mudança no uso e na cobertura do solo e com a tentativa de destrinchar as interações entre seus diversos determinantes. Assim, é comparativamente mais fácil evitar essas dificuldades, adotando-se uma abordagem instrumental para a teoria (instrumentalismo¹⁷), o que simplifica as relações observadas e as reduz a expressões matemáticas quantitativas que podem ser compreendidas e manipuladas.

As razões práticas incluem tanto a disponibilidade de toda sorte de recursos, como as demandas dos “clientes” da tomada de decisão. Os primeiros são críticos e geralmente direcionados a atividades que trazem resultados operacionais “visíveis” razoavelmente rápidos. Os clientes da tomada de decisão, por outro lado, independentemente de sua atitude relativa às teorias, podem colocar maior prioridade a ferramentas operacionais na tomada de decisão, onde esta pode requerer uma agilidade relativa. Nesse caso, assuntos controversos devem ser evitados, e os resultados têm que ser facilmente traduzidos em ações concretas.

1.3 - TEORIAS DO ESTUDO DA MUDANÇA NO USO E NA COBERTURA DO SOLO

Uma teoria de mudança no uso e na cobertura do solo explica porque as mudanças ocorrem, quais são as causas dessas mudanças, e quais são os mecanismos de mudança. Normalmente, elas se referem ao “o quê” ou ao “porquê” e não a ambos.

¹⁶ - Técnicas usadas para modelar um sistema de interesse que não utilizam expressões matemáticas formais e técnicas. Ao contrário, baseiam-se em regras que são usadas para guiar a representação das relações investigadas. (BRIASSOULIS, 2000)

¹⁷ - Filosofia de ciência e uma abordagem ao desenvolvimento da teoria, voltada a dispositivos computacionais concebidos para descrever relações observadas, mas que não questiona a verdade ou falsidade das declarações teóricas produzidas. Modelos matemáticos são a expressão mais direta de instrumentalismo, especialmente quando eles são dirigidos a avaliar a qualidade do ajuste de um conjunto de dados às relações matemáticas especificadas, sem a preocupação com os processos que produzem esses padrões. (BRIASSOULIS, 2000)

Para os fins deste trabalho, uma grande distinção é estabelecida entre as teorias que tratam do solo e, mais importante, da mudança no uso e na cobertura do solo de forma explícita e aquelas nas quais a referência à mudança no uso e na cobertura do solo é mais ou menos indireta e está implícita no contexto de uma discussão mais ampla.

Tomando-se por base esse critério, análises que envolvem a teoria da localização (incluindo os estudos sobre a teoria do lugar central¹⁸) possuem um escopo limitado. Elas não são consideradas teorias de mudança no uso e na cobertura do solo, dado que sua ênfase está em atividades pontuais localizadas no espaço e não em uma área usada por várias atividades. A maioria dessas teorias é baseada em agentes, ou seja, deduzem mudanças na estrutura espacial, a partir do comportamento de unidades familiares ou empresas. (BRIASSOULIS, 2000)

Algo comum a todas as tarefas analíticas é a necessidade de se ter um veículo capaz de estruturar a concepção e explicação da realidade - uma teoria. A análise de mudança no uso e na cobertura do solo não constitui exceção. Entretanto, ainda que o uso de teorias na construção de modelos pareça indispensável, um número relativamente pequeno foi usado para servir de guia e de apoio à modelagem operacional. (LONERGAN e PRUDHAM, 1994)

A maior parte das teorias ainda carece de modelagem dedicada (não necessariamente matemática), e a recíproca é verdadeira. Vários são os modelos destituídos de fundamentações teóricas, em parte, porque a realidade considerada é extremamente complexa.

Cooke (1983) sugere que cada tradição teórica se especializa, *grosso modo*, em um dado nível espaço-temporal, embora seja difícil estabelecer qual o nível dominante para cada uma delas. Pode-se dizer, contudo, que quanto mais detalhado o nível de referência, mais elaborada se torna a teorização e são apresentadas considerações mais concretas e realistas do contexto e dos mecanismos de mudança. Em níveis mais abrangentes, a teorização é abstrata e pode ser muito difícil extrair, da teoria para o mundo real, suas contrapartes do contexto e dos mecanismos de mudança.

As teorias nas quais solo e uso do solo são tratados explicitamente – como áreas da superfície terrestre com propriedades variáveis espacialmente – são poucas, se comparadas ao grande número de esquemas teóricos disponíveis. O segundo milênio se esgotou e as teorias/modelos de von Thunen e de Alonso, apresentados no início desta Seção, permanecem como sendo os mais concisos e explícitos em termos de uso do solo.

¹⁸ - Em 1933, Christaller (1966) formulou a ‘teoria do lugar central’ para quantificar, em bases teóricas, o tamanho e a distribuição de estabelecimentos de varejo dentro de uma área urbana, empregando dois conceitos primordiais: o “alcance” de um bem e o “limite” de um bem.

Entretanto, mesmo essas duas teorias possuem limitações, na medida em que não cobrem todas as variedades existentes de tipos e padrões de uso do solo, nem toda a diversidade de indutores de mudança no uso e na cobertura do solo. Esta é uma característica da maioria das teorias nas quais o uso do solo é tratado explicitamente, ou seja, elas se referem de forma direta somente a tipos particulares de uso do solo – residencial, industrial, comercial, agrícola, florestal – e não a toda a gama existente e potencial de usos do solo.

O fato é que uma análise realista e holística do uso do solo e de suas mudanças necessita tratar o uso do solo e seus determinantes no mesmo nível de detalhamento. É interessante notar também o pequeno número de teorias que especificam padrões particulares de usos do solo que resultam no processo de mudança.

De fato, somente as teorias pioneiras – de von Thunen, de Alonso, da ecologia humana (sociológica), do determinismo ambiental – tratam esse aspecto de mudança no uso e na cobertura do solo. Essa talvez seja uma indicação dos limites de uma teorização significativa e da habilidade humana em prever configurações espaciais exatas, que decorram das intrincadas relações entre os indutores biofísicos e socioeconômicos do uso do solo.

Os indutores de mudança no uso e na cobertura do solo levados em conta pelas teorias aqui tratadas cobrem toda a gama de fatores biofísicos e socioeconômicos. Entretanto, os aspectos de particular importância são os mecanismos de mudança no uso e na cobertura do solo que essas teorias propõem.

A esse respeito, dois grupos de teorias podem ser discernidos. O primeiro é composto de teorias estáticas que assumem que o sistema de uso do solo atinge uma posição de equilíbrio em algum ponto no tempo quando certos fatores mudam. Essas teorias simplesmente mencionam quais são as mudanças, deixando para o usuário a racionalização do “porquê” dessas mudanças e o “como” elas produzem determinados resultados de uso do solo.

O segundo grupo de teorias não lida diretamente com as mudanças no uso e na cobertura do solo, mas com as mudanças em seus determinantes. Aparentemente, cabe ao usuário da teoria explicar o mecanismo específico de uso do solo no contexto particular de sua aplicação.

Cooke (1983) afirma ainda que, em geral, muito poucas teorias explicam mudanças no uso e na cobertura do solo. Algumas oferecem explicações para tipos particulares de mudança, tais como industrialização, urbanização, suburbanização e desmatamento. Ainda assim, as “explicações” por elas apresentadas carecem de uma qualificação mais apropriada.

Há diversas razões para essa pobreza explanatória dessas teorias. A explanação pode envolver fatores e processos que operem em vários níveis espaciais, seja na mesma escala temporal, ou não, dado um contexto espacial específico. A influência de fatores climáticos na mudança no uso e na cobertura do solo, por exemplo, requer uma “janela” temporal de séculos, em termos de zonas bioclimáticas, ao passo que a influência de mudanças no preço do produto ou de instrumentos de política possam ser examinados em períodos de tempo de apenas alguns anos e em escalas mais detalhadas.

Aparentemente, teorias capazes de levar em consideração essa complexidade espaço-temporal ainda não existem. Em suas presentes formas, muitas teorias não podem ser refletidas em modelos de mudança no uso e na cobertura do solo capazes de analisar os impactos de mudanças a partir de suas variáveis explicativas.

De acordo com Sayer (1976, 1979_a, 1984), isso se deve a três razões: (1) muitas teorias são concebidas em termos abstratos, que tornam sua operacionalização difícil; (2) a orientação epistemológica de várias teorias não é congruente com a modelagem formal; e (3) os modeladores não apresentam interesse em identificar modelos com teorias específicas, embora assunções teóricas sejam intrínsecas a todos os modelos construídos.

A análise de casos concretos de mudança no uso e na cobertura do solo pode ser conduzida de maneira significativa, em níveis espaciais e temporais ditados pelas circunstâncias particulares históricas e geográficas. Assim, parece ser mais adequada a utilização de uma síntese de teorias, ao invés de uma única, o que provocaria a inevitável perda de algumas dimensões do caso em estudo, ou torná-lo-ia exageradamente complexo para ser útil e facilmente entendido.

A literatura sobre mudança no uso e na cobertura do solo contém uma considerável variedade de teorias, no qual o uso do solo é tratado explicitamente e constitui o objeto de investigação teórica. Briassoulis (2000) informa que seis fontes inter-relacionadas para tais teorias podem ser enumeradas, em ordem decrescente de importância:

- a. Propósito do projeto teórico;
- b. Abordagem da tradição teórica;
- c. Escala espacial e nível de agregação espacial adotado;
- d. Tipos de uso do solo considerados como principais objetos de análise;
- e. Tipos de determinantes da mudança no uso e na cobertura do solo levados em consideração;

- f. Tratamento dado à dimensão temporal (inerente a qualquer caso de análise de mudança).

Assim, com base no critério de tradição teórica defendido por Briassoulis (2000), três são as classes genéricas de teorias de mudança no uso e na cobertura do solo, a saber:

1. Tradição teórica da economia urbana e regional
2. Tradição teórica sociológica e de política econômica
3. Tradição teórica natureza-sociedade

A partir dessas três categorias principais, as teorias podem ser classificadas mais detalhadamente, em função de outros critérios específicos. Em última instância, é muito difícil, senão impossível, prover uma classificação inequívoca, dado que mais de um critério pode ser usado para tal.

Para cada teoria, são examinados os seguintes tópicos principais: propósito (descritivo, preditivo, explicativo, prescritivo), modo de teorizar (suposições, tipo de uso do solo e os determinantes considerados, especialmente o mecanismo proposto para a mudança no uso e na cobertura do solo), escala espacial de referência, e dimensão temporal (duração, dinâmica).

1.3.1 - A tradição teórica da economia urbana e regional

Nesse conjunto de teorias, a realidade é representada por meio de conceitos e procedimentos de natureza econômica – fatores de produção, de produtos e de serviços, custos de transportes (ou transferência), custos marginais, economias de escala, externalidades e, acima de tudo, utilidade, que é a satisfação que um indivíduo obtém do consumo de um conjunto de bens e de serviços. (SIMON, 1982)

As teorias de mudança no uso e na cobertura do solo pertencentes a esta tradição se agrupam genericamente em teorias baseadas em aspectos microeconômicos e macroeconômicos.

As primeiras partem do comportamento individual do consumidor, para então agregar o comportamento de todos os consumidores, de maneira a elaborar padrões de uso do solo, produzidos quando a utilidade está sendo maximizada para todos eles (normalmente, maximização de lucros ou minimização de custo ou distância).

Em contraste, abordagens macroeconômicas tratam de comportamento agregado e indicam como padrões agregados podem ser produzidos.

Um terceiro grupo de teorias utiliza conceitos da economia e da sociologia. Briassoulis (2000) justifica sua inclusão nesta tradição teórica pela ênfase em fatores e processos econômicos de mudanças espaciais.

1.3.2 - A tradição teórica sociológica e de política econômica

De maneira geral, a teorização sociológica enfatiza a importância das ações humanas, relações sociais, redes sociais, e mudanças socioculturais em provocar mudanças políticas, econômicas, entre outras. O termo ‘social’ é usado aqui num sentido amplo, que engloba todas as manifestações da sociedade: modos de produção¹⁹, instituições, políticas, culturas, estilos de vida etc.

São cinco as teorias que compõem essa tradição teórica: teoria funcionalista-behaviorista; teoria estruturalista-institucionalista; teorias centro-periféricas; teorias sobre trocas desiguais; e teorias sobre lógica capital e desenvolvimentos desiguais. (BRIASSOULIS, 2000)

1.3.3 - A tradição teórica natureza-sociedade

Trata-se do conjunto mais aberto e diverso de dentre as três tradições teóricas consideradas, na medida em que incorpora a análise de mudança no uso e na cobertura do solo no contexto da mudança ambiental global. A principal questão formulada por esse conjunto de teorias é “como o homem se relaciona com a natureza”, que se traduz na questão mais comum e popular do “papel do homem como causador de mudanças ambientais”.

As teorias agrupadas nessa categoria (teorias baseadas em humanidades, teorias baseadas em ciências naturais e teorias baseadas em ciências sociais) têm, como característica comum, o trato com aquilo que Briassoulis (2000) denomina ‘sistema total’. Em outras palavras, a totalidade das interações entre natureza (meio ambiente), economia, sociedade (incluindo políticas e instituições) e cultura.

Considerações sobre o sistema total vêm desde a preocupação de Malthus com a relação entre disponibilidade de terras (em quantidade e qualidade) e crescimento populacional, no final do século XIII. O tema voltou a ganhar destaque com os acalorados debates ocorridos nas décadas de 1960 e de 1970, quando a crise ambiental ganhou as arenas científica e política.

¹⁹ - O conjunto de relações pelos quais uma sociedade estrutura e organiza atividades produtivas. É uma característica que distingue sociedades com base na sua organização socioeconômica. São modos representativos de produção: pré-capitalista, capitalista e socialista. (BRIASSOULIS, 2000)

Vale registrar que o conceito de sistema total existe desde a preocupação de Malthus com a relação entre disponibilidade de terra (em qualidade e quantidade) e crescimento da população ao final do século XVIII. Em 1864, George Perkins Marsh, em seu ensaio 'Homem e Natureza', descreveu objetivamente, como as pessoas usam e modificam o solo para servir a vários propósitos, alterando, assim, o ambiente. A influência dessas análises é evidente nas teorias que se desenvolveram nesta tradição e que alimentaram vivos debates, quando a crise ambiental ascendeu para as arenas científica e política.

Embora aplicáveis a vários níveis espaciais, as teorias multidisciplinares são mais relevantes em escalas mais abrangentes.

1.4 - MODELAGENS DO ESTUDO DA MUDANÇA NO USO E NA COBERTURA DO SOLO

Modelos podem ser considerados a representação formal de alguma teoria sobre um dado sistema de interesse. De forma ampla, modelos são abstrações, aproximações da realidade que é alcançada pela simplificação de relações complexas do mundo real, até um ponto em que essas relações possam ser entendidas e analiticamente manipuladas. A representação da realidade é expressa por meio do uso de símbolos.

A construção de modelos ajuda a dar suporte às teorias nas quais eles se baseiam. Entretanto, muito embora o uso de teorias pareça indispensável à construção de modelos, um número relativamente pequeno de teorias de mudança no uso e na cobertura do solo tem sido usado na orientação e no suporte para a construção desses modelos.

Modelos que tratam de uso do solo e de mudança no uso e na cobertura do solo de maneira explícita são basicamente aqueles nos quais o objeto direto da modelagem é a mudança no uso e na cobertura do solo. Esta, por sua vez, é o resultado de uma complexa rede de interações entre forças biofísicas e socioeconômicas sobre o tempo e o espaço. Lidar com essa complexidade para fins práticos é impossível sem alguma forma de simplificação dessas relações para dimensões compreensíveis e manipuláveis.

Como os problemas de mudança no uso e na cobertura do solo envolvem um número indefinido de variáveis, a redução e simplificação da diversidade existente no mundo real, com vistas a servir aos propósitos da construção de modelos ou é extremamente difícil, ou resulta em uma representação bastante grosseira da realidade.

Os principais usos passíveis de serem feitos com esses modelos são os seguintes:

- Suporte à decisão em contextos de política ambiental;
- Descrição das relações espaciais e temporais entre os indutores e os padrões resultantes de usos do solo e de suas mudanças;
- Veículos explanatórios das relações observadas;
- Predição de futuras configurações de padrões de uso do solo sob vários cenários de mudança biofísica e socioeconômica;
- Avaliação de impacto de atividades nas esferas ambientais e socioeconômicas;
- Prescrição de padrões ótimos de uso do solo para uso sustentado dos recursos naturais; e
- Avaliação associada aos três últimos usos mencionados – predição, avaliação de impacto e prescrição.

A concepção de modelos normalmente está associada com propósitos específicos do modelo, teorias que lhe dão suporte e tipos de uso do solo modelados, além dos níveis de análise espaço-temporal.

Baseada nesses critérios, Briassoulis (2000) fixou cinco tradições de modelagem – modelos estatísticos e econométricos, modelos de interação espacial, modelos de otimização, modelos integrados e modelos gerados a partir de abordagens mais recentes. Em cada uma dessas tradições, apresentadas a seguir, os modelos são sucessivamente classificados, de acordo com critérios particulares a cada caso.

1.4.1 - Modelos estatísticos e econométricos

A técnica estatística mais comumente utilizada é o Método de Regressão Linear Múltipla (MRLM), embora não seja incomum o uso de outras técnicas multivariadas (análise fatorial, análise canônica etc.).

A MRLM é uma técnica estatística para análise da relação matemática entre três ou mais variáveis. Uma delas é chamada dependente, na medida em que seus valores são considerados dependentes das mudanças nos valores das variáveis independentes (uma ou mais), também conhecidas como explanatórias ou explicativas. Quando a técnica envolve apenas uma variável independente associada à variável dependente, ela é conhecida como Método de Regressão Linear Simples. A aplicação do MRLM para a análise de problemas que envolvam procura e oferta propiciou os chamados “modelos econométricos”.

Modelos estatísticos, cujo objeto de análise é a mudança no uso e na cobertura do solo, existem desde 1960, pelo menos, e são bastante empregados em diversos estudos. Frequentemente, eles constituem componentes de modelos mais amplos, utilizados para a análise de mudança no uso e na cobertura do solo e de seus determinantes.

Uma distinção pode ser estabelecida entre modelos contínuos, que tratam do uso do solo como uma variável contínua (área dedicada a um determinado tipo de uso do solo), e modelos discretos, que tratam o uso do solo como uma variável discreta (diferentes tipos de uso do solo podem ser identificados).

Em um modelo estatístico de mudança no uso e na cobertura do solo, a área de estudo é geralmente subdividida em zonas, com seus tamanhos e formas dependentes do nível de agregação escolhido, bem como da disponibilidade de dados. No caso contínuo, para cada zona, é dada a distribuição dos tipos de uso do solo (variáveis dependentes), bem como os valores de outras variáveis preditoras ambientais e socioeconômicas (independentes). Uma equação de regressão múltipla para cada tipo de uso do solo é adequada a esses dados e apresenta segundo a forma geral:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i$$

onde: Y_i é o estado encontrado na unidade de superfície do solo e X_i é o preditor usado por cada variável. O termo ε_i é o erro do modelo estatístico.

Esse modelo pode ser usado para investigar as mudanças no uso e na cobertura do solo, em face de mudanças específicas em uma ou mais variáveis preditoras, por meio da substituição de seus valores na equação apresentada acima.

Modelos estatísticos discretos são usados para representar situações de escolha em geral. Na modelagem de uso do solo, cada uso é descrito como sendo a função de uma gama de características, usualmente diferentes de uma célula para outra. Para cada célula, a utilidade de cada tipo de uso do solo é investigada em função dessas características. (ANAS, 1982)

A probabilidade de se escolher um tipo de uso do solo particular em uma dada célula é calculado em função das utilidades associadas com os tipos de uso do solo considerados.

Outra técnica estatística, a análise de correlação canônica (CCA – *Canonical Correlation Analysis*), tem sido usada para explorar os determinantes socioeconômicos e ambientais, bem como para testar a estabilidade temporal dos padrões de uso do solo (HOSHINO, 1996). A CCA é uma técnica estatística multivariada usada para explorar a estrutura das relações entre

uma variável dependente e um conjunto de variáveis independentes e é particularmente adequada quando as variáveis independentes são correlacionadas entre si.

O propósito dos modelos estatísticos de mudança no uso e na cobertura do solo é a descrição, explanação e a predição (condicional e incondicional) dessas mudanças, como função de certos determinantes. Eles são modelos ditos estáticos, concebidos para operar em uma base anual.

Não há uma teoria específica por trás desses modelos, à exceção da teoria genérica, segundo a qual mudanças no uso e na cobertura do solo decorrem de mudanças em forças socioeconômicas e ambientais. Em outras palavras, eles adotam uma visão instrumentalista da teoria. (SAYER, 1979_b: 858)

A especificação da modelagem, na maioria dos casos, reside em variáveis quantitativas, o que significa dizer que aspectos qualitativos da mudança no uso e na cobertura do solo, que não possam ser medidos e quantificados, não entram no modelo.

A regressão múltipla revela correlações ou associações entre variáveis, um fato que não tem nada a ver com causalidade. Modelos causais requerem uma fundamentação teórica bem mais rigorosa do que aquela implícita nos modelos mencionados. A maior parte dos modelos estatísticos de mudança no uso e na cobertura do solo é composta de modelos de regressão linear múltipla, que assumem, portanto, linearidade entre as variáveis.

Uma forma de contornar essa fragilidade é utilizar técnicas de estatística espacial, ao invés das tradicionais técnicas convencionais. Isso porque dados espaciais quase sempre apresentam autocorrelação espacial²⁰, que deve ser levada em conta e corrigida por meio de técnicas de estatística espacial adequadas. (LeSAGE, 1999)

Além disso, deve-se ter em mente que modelos de regressão múltipla apresentam também multicolinearidade²¹, na medida em que as variáveis preditoras²², em geral, são correlacionadas entre si. Isso constitui um problema, quando se utiliza o modelo para explanação, o que não ocorre quando ele é usado para predição. (PAMPEL, 2000)

²⁰ - Presença de fortes relações entre as observações conduzidas em diferentes pontos no espaço. A autocorrelação espacial ocorre em coeficientes de regressão viesados. Técnicas especiais, conhecidas como estatística espacial e econometria espacial, precisam ser aplicadas para corrigir os problemas associados com a autocorrelação espacial.

²¹ - Em modelos estatísticos de regressão múltipla, quando as variáveis independentes são relacionadas entre si, surge o problema de multicolinearidade. Ele resulta em coeficientes de regressão que podem não ser estatisticamente significantes, na medida em que os coeficientes das variáveis independentes interrelacionadas refletem a influência que umas exercem nas outras, em maior ou menor grau.

²² - Variáveis utilizadas para explicar a variabilidade de outra variável (a dependente). Também são conhecidas como variáveis independentes.

No caso de modelos econométricos, a prática mais comum é o cálculo de mudanças de certos determinantes para o uso do solo para, então, converter essas estimativas em requisitos de uso do solo, por tipo de uso do solo. Para tanto, são usados coeficientes de atividade/uso do solo, que apresentam a razão entre a área de superfície e a magnitude de uma dada atividade de uso do solo, como, por exemplo, a razão de residentes por km², de vendas por m² quadrado, ou de áreas cultivadas por hectare.

1.4.2 - Modelos de interação espacial

Esses modelos se originam nas tentativas de modelar as atividades humanas no espaço físico, tomando por base a analogia com a lei da gravidade (física). Assim sendo, os modelos incluídos neste grupo são os modelos gravitacionais já consagrados, aliados às versões mais recentes, conhecidas como modelos de interação espacial.

Esses modelos têm sido usados para modelar uma variada gama de interações que surgem em decorrência de atividades humanas como o deslocamento para o trabalho, compras, circulação, e mobilidade, em geral. (HAYNES e FOTHERINGHAM, 1984)

Natural e logicamente, a força da interação entre tipos de uso do solo depende da magnitude e natureza da atividade associada. Conseqüentemente, mudanças em atividades - que também são refletidas em mudanças nas suas interações - podem causar algum tipo de mudança de uso e de cobertura do solo, seja essa mudança qualitativa, seja quantitativa, ou ambas as modalidades. O oposto também é verdadeiro. Mudanças no uso e na cobertura do solo podem induzir mudanças nas atividades associadas, bem como nas interações entre eles.

O propósito dos modelos gravitacionais (genericamente, modelos de interação espacial) é, basicamente, o de simular os fluxos entre zonas de origem e de destino, e de prever esses fluxos, quando ocorrem mudanças nos destinos e/ou nas origens e/ou quando muda a acessibilidade entre origens e destinos (principalmente por melhorias na rede de transporte).

Vale ressaltar que a maior parte das aplicações e usos desse tipo de modelo diz respeito a áreas urbanas e metropolitanas (não se aplica a áreas agrícolas e florestais).

1.4.3 - Modelos de otimização

Como sugere a denominação, esses modelos são orientados unicamente para produzir soluções que aperfeiçoem certos objetivos estabelecidos por tomadores de decisão. Em outras palavras, eles são ajustados para apoiar situações decisórias, nas quais deve-se escolher uma solução que satisfaça um ou mais objetivos, e que leve em conta restrições diversas.

Conseqüentemente, são modelos prescritivos, embora eles também possam ser usados como instrumentos de avaliação. Eles podem ter importantes aplicações na análise de uso do solo - especialmente aplicações voltadas ao planejamento desse uso. Recentemente, eles têm sido úteis na busca de soluções para o uso do solo capazes de contribuir para o desenvolvimento sustentável e para o uso de recursos ambientais e humanos.

De acordo com Briassoulis (2000), o principal critério para a classificação desse tipo de modelo parece ser a técnica específica de programação/otimização matemática empregada. Uma exceção é o grupo de modelos de maximização da utilidade, que é tratado como um grupo separado, na medida em que ele é expressamente baseado na teoria econômica.

1.4.4 - Modelos integrados

São diversos os modelos integrados de mudança no uso e na cobertura do solo, uma vez que o termo ‘integração’ possui diferentes significados em diferentes contextos. Eles também são chamados de modelos “abrangentes” ou “gerais”, embora o termo ‘integrado’ tenha passado a dominar a literatura a partir da década de 1980.

Modelos integrados são aqueles que consideram, de alguma forma, interações, relações e ligações entre dois ou mais componentes de um sistema espacial – sejam eles setores de atividade econômica, regiões, sociedade e economia, meio ambiente e economia – e as relacionam ao uso do solo e às suas mudanças, direta ou indiretamente. (WEGENER, 1986)

O propósito de construção da maioria dos modelos integrados não é modelar a mudança no uso e na cobertura do solo, mas algum outro aspecto de um sistema espacial de interesse.

De maneira geral, os modelos integrados surgiram na década de 1960. Os primeiros esforços incluíram explicitamente o uso do solo e essa característica continua até hoje. Entretanto, modelos integrados cujos propósitos sejam a análise de mudança no uso e na cobertura do solo são bastante recentes.

Uma característica comum aos modelos integrados, à parte de sua ênfase na integração, é que eles são, em sua maioria, modelos de larga escala. Em verdade, uma revisão da literatura (BATTEM e BOYCE, 1986; BOYCE, 1988; WEGENER, 1994) sobre modelos de larga escala revela que a maior parte deles é constituída de modelos integrados. Os níveis espaciais cobertos vão do global ao urbano/metropolitano.

O significado do termo ‘integração’ varia em função do propósito do modelo e é refletido na estrutura do modelo integrado. Cinco dimensões de integração podem ser genericamente categorizadas:

- Integração espacial – onde interações horizontais e/ou verticais entre níveis espaciais são enfatizadas, no que diz respeito ao fenômeno modelado;
- Integração setorial – onde o modelo representa as ligações e relações entre dois ou mais setores econômicos do sistema espacial de interesse, tal como varejo, imobiliário, transporte, indústria, agricultura etc.;
- Integração do uso do solo – nesta, o modelo considera as interações entre dois ou mais tipos de uso do solo, como residencial, comercial, produção, transporte etc. Esta dimensão pode ser equivalente, às vezes, à integração setorial;
- Integração economia-sociedade-meio ambiente – onde o modelo representa as ligações entre pelo menos dois dos diversos componentes do sistema espacial, tal como economia-meio ambiente, economia-sociedade, economia-energia etc.;
- Integração de sub-mercados – onde os modelos mostram quão diferentemente os sub-mercados se relacionam entre si (oferta e procura, por exemplo).

Briassoulis (2000) não considera a dimensão temporal como uma das dimensões de integração. Modelos que incorporam o tempo são denominados “dinâmicos”, que podem envolver de modelos simples até modelos integrados. Estes, contudo, são menos frequentes, dadas as muitas dificuldades associadas à concepção e à construção de modelos integrados dinamicamente operacionais.

As cinco dimensões principais de integração não são mutuamente excludentes. Na verdade, qualquer modelo integrado pode combinar mais de uma dessas dimensões. A tendência moderna na construção de modelos integrados é considerar várias dimensões, com ênfase na dimensão espacial – especialmente quando se modela o uso do solo ou assuntos ligados ao meio ambiente.

Note-se entretanto que, quanto mais alto é o grau de integração, maiores são as dificuldades para conceber e dar suporte teórico às relações de interesse, de um lado, e para operacionalizar e aplicar os respectivos modelos, de outro.

1.4.5 - Abordagens recentes

As quatro principais categorias de modelos apresentados até aqui cobrem, mais ou menos, a maioria dos modelos de uso do solo e de mudança no uso e na cobertura do solo. Entretanto, há diversas outras abordagens, as quais não se encaixam facilmente nessas categorias, mas que não configuram uma categoria específica, uma vez que sua aplicação é especializada, esparsa, ou muito recente.

Esses modelos podem ser agrupados, *grosso modo*, em: (a) abordagens orientadas pelas ciências naturais; (b) modelagem markoviana de mudança no uso e na cobertura do solo; e (c) abordagens baseadas em sistemas de informação geográfica.

A modelagem de mudança no uso e na cobertura do solo tem sido conduzida primariamente nas disciplinas de geografia, ciências regionais, e economia urbana e regional. Atividades ligadas ao planejamento têm-se utilizado dessas disciplinas, muito embora haja exceções, especialmente em anos recentes, quando os limites disciplinares de tornaram menos evidentes.

A modelagem de mudança no uso e na cobertura do solo, contudo, é tema de outras disciplinas, como ecologia, ecologia da paisagem, ciências florestais, ciências dos solos e ciências ambientais em geral. Os modelos desenvolvidos nessas disciplinas têm uma característica comum, que é a forte ênfase em aspectos biofísicos (determinantes e impactos) da mudança no uso e na cobertura do solo e, às vezes, uma certa desconsideração por determinantes socioeconômicos, institucionais, políticos, entre outros.

As Cadeias de Markov pertencem aos métodos analíticos dos processos estocásticos²³. Trata-se de uma técnica de simulação que tem sido aplicada à análise de mudança no uso e na cobertura do solo, mas sem a difusão de outras técnicas de simulação. (BRIASSOULIS, 2000)

A análise markoviana constitui um método estatístico bastante complexo e seu uso requer profundo entendimento dos conceitos matemáticos e estatísticos usados, bem como testes rigorosos das suposições assumidas (por exemplo, que os processos observados são processos markovianos e, em particular, estacionários).

São suposições básicas, mas difíceis de serem testadas. Uma das razões está na falta de séries históricas suficientes acerca das mudanças no uso e na cobertura do solo entre os

²³ - Processo cujos resultados dependem de elementos de chance (são expressos como probabilidades).

períodos de tempo, para ajustar as matrizes de probabilidade de transição que permitam testar a estabilidade dos coeficientes de transição no tempo.

Por fim, a abordagem proposta pelos modelos baseados em GIS encontra-se em desenvolvimento e suas aplicações ao estudo dessas mudanças ainda está por ser avaliada, comparativamente às técnicas de modelagem mais tradicionais já discutidas. Eles descrevem as localizações dos objetos no espaço e suas relações topológicas, bem como os atributos não espaciais dos objetos gravados. São denominados dados de atributo ou dados temáticos (FISCHER *et al.*, 1996).

A análise espacial oferece quantidade e diversidade de procedimentos para a análise de fenômenos espaciais. Os dois principais direcionamentos que podem ser destacados são a análise estatística de dados espaciais e a modelagem espacial, abordagem adotada neste trabalho.

1.4.6 - Considerações sobre os modelos apresentados

Os modelos apresentados podem não representar todo o universo de modelos que tem sido construído e usado para a análise de mudança no uso e na cobertura do solo. Diversos modelos nunca foram publicados em livros ou jornais científicos. Outros são feitos e jamais são usados, a despeito do benefício que trariam, caso fossem adotados.

Modelos são concebidos para uma variedade de propósitos, com diferentes recursos humanos, técnicos e financeiros disponíveis, e em períodos históricos nos quais o conhecimento básico requerido pode não estar bem desenvolvido ao longo de todos os aspectos existentes na estrutura do modelo. Assim, não é possível julgar se um modelo é “bom” ou “ruim”, na medida em que tudo depende das circunstâncias que levam à construção e ao uso do modelo.

Por essas razões, Briassoulis (2000) agrupa as principais características para a construção de modelos de mudança no uso e na cobertura do solo, segundo os seguintes aspectos:

- Propósito do modelo – modelos positivos (descritivos, preditivos e de avaliação de impactos) e modelos normativos (prescritivos).
- Objeto do estudo – áreas urbanas, rurais, florestais, regiões, nações, o globo etc.
- Nível de agregação²⁴ – espacial e funcional – sistema espacial de referência e setores representados com seus respectivos usos do solo.

²⁴ - A importância da escala – espacial, temporal, institucional etc. – é um tema transdisciplinar tanto para teorias quanto para modelos. Em modelos de ecologia da paisagem, escala é a primeira e mais importante preocupação na análise da mudança no uso e na cobertura do solo. (BRIASSOULIS, 2000)

- Dinâmica – relações temporais entre usos do solo e seus diversos determinantes.
- Teorias subjacentes – principais fundamentos que dão sustentação ao modelo.
- Especificação funcional – explicitude espacial.
- Questões ligadas aos dados – disponibilidade de conjuntos de dados adequados.
- Aplicações reais – sucesso operacional de um modelo.

1.5 - EXEMPLOS DE MODELOS DE MUDANÇA NO USO E NA COBERTURA DO SOLO - AMAZÔNIA

São escassas as abordagens preditivas para os modelos especificamente voltados ao estudo de mudança no uso e na cobertura do solo na Amazônia. Alencar, Solórzano e Nepstad (2004) realizaram um estudo acerca da relação espacial entre incêndios sob o dossel florestal (variável dependente) e uma série de variáveis independentes, por meio de análises de regressão logística.

Modelos de probabilidade de incêndio foram desenvolvidos para várias combinações das características dessas feições. Durante o período de dez anos do estudo, a maior parte (91%) da floresta considerada foi queimada nos anos denominados de Oscilação Sul do El Niño (*ENSO Years*), quando secas severas aumentaram a suscetibilidade a incêndios, por parte da floresta, e favoreceram a perda de controle de incêndios provocados com fins agrícolas.

De acordo com os autores, os incêndios na floresta amazônica estão fortemente associados ao empobrecimento florestal que ocorre em função do corte discriminado de madeira de lei. Contudo, o preditor de incêndio mais forte foi a percentagem do fragmento de floresta anteriormente cortada ou desmatada. Significa dizer que os incêndios sob o dossel florestal foram associados aos limites florestais, conforme foi reportado em estudos anteriores da Amazônia. Além disso, também estão associados a características da paisagem, tais como tamanho do fragmento florestal, distância de fontes de ignição (assentamentos e fornos de carvão) e distância de rodovias principais.

No estudo de Alencar, Solórzano e Nepstad (2004), modelos de regressão logística com base em informações de degradação de fragmentos florestais e distância de fontes de ignição previram com precisão mais de 80% dos incêndios florestais durante o evento ENSO de 1997-1998. Para anos ENSO, em geral, as variáveis que apresentaram correlação significativa com o fogo foram, em ordem decrescente:

- Perturbação anterior por exploração madeireira e incêndios;
- Distância de fornos de carvão;
- Distância da borda da floresta; e
- Tamanho do fragmento.

Soares-Filho *et al.* (2004) simularam resultados das mudanças de cobertura do solo ao longo da BR-163 (Cuiabá-Santarém), em função do eventual asfaltamento da rodovia e de aspectos de governança. Os autores reforçam o entendimento de que a distribuição espacial de atividades humanas em regiões de fronteira florestal é fortemente influenciado pela infraestrutura de transporte. A área estudada cobre os cerca de 1000 km ainda não-pavimentados, de um total de 6000 km da rodovia. O modelo de simulação por eles desenvolvido produz quatro cenários, divididos em dois submodelos.

O primeiro deles, denominado modelo espacial, está ligado a atividades populacionais, possui uma abordagem quantitativa e propõe cenários decorrentes de alto e médio crescimentos populacionais em torno do trecho analisado. Esse modelo, que usa o simulador 'Dinamica', foi executado sobre um banco de dados GIS, para um período de 30 anos, em etapas anuais.

O segundo, denominado modelo de cenários alternativos, é função das formas de intervenção governamental (conservativa e pessimista), possui uma abordagem semi-quantitativa e é baseado em agentes. Esse modelo foi concebido para incorporar as relações existentes entre desmatamento, população e renda *per capita*, integrando os efeitos das diversas variáveis, o que inclui tendências agrárias, grau de urbanização e respostas dos agentes aos cenários político-econômicos, suas percepções e demandas.

Em um estudo de mudança no uso e na cobertura do solo, mas que não pode ser considerado propriamente um estudo de modelagem, Roberts *et al.* (2002) descrevem a variação espaço-temporal da cobertura do solo na região central de Rondônia. Usando um processo multi-estágio, eles mapearam floresta primária, floresta secundária, pastagem, afloramento rochoso/savana, área urbana e água, por intermédio de 33 cenas Landsat, coletadas sobre três áreas contíguas, no período de 1975 a 1999.

Roberts *et al.* (2002) concluíram que as pastagens perduram vários anos sem abandono, deixando de resultar em florestas secundárias. Estas, quando ocorrem, raramente persistem mais de oito anos. A principal conclusão do estudo é a constatação de que a combinação de padrões de assentamentos governamentais, o desmatamento concentrado em faixas de cerca

de 50 km em torno das rodovias e a crescente atividade de ocupação promove uma intensa fragmentação florestal. Essa fragmentação promove o aumento da mortalidade nas bordas florestais e a perda de diversidade, entre outros fatores causadores de incêndios florestais.

Alguns estudos se limitam a modelar o processo de classificação de imagens, etapa essencial para os estudos de mudança no uso e na cobertura do solo. Campos (2005), por exemplo, propõem um procedimento para a modelagem do conhecimento multitemporal²⁵ e sua integração com outras formas de conhecimento num ambiente integrado para a interpretação automática de imagens de sensores remotos. A proposta consiste em utilizar diagramas de transição²⁶ de estado (BÜCKNER *et al.*, 1999; GROWE *et al.*, 2000) para expressar a relação entre a classe de um objeto num dado momento e a classe do mesmo objeto num instante anterior.

Campos *et al.* (2005) utilizam os valores de possibilidade de transição, estimados com algoritmos genéticos (DAVIS, 1990; PACHECO, 2006), como parâmetros das funções discriminantes²⁷. Por intermédio de ferramentas da lógica nebulosa (ZADEH, 1978; MENDEL, 1995), os resultados obtidos da aplicação do conhecimento multitemporal são combinados com outras formas de conhecimento, como o espectral²⁸, por exemplo. O texto sugere que, além de automatizar e acelerar a modelagem do conhecimento, o uso de algoritmos genéticos pode possibilitar resultados superiores aos parâmetros ajustados manualmente.

²⁵ - Modalidade de conhecimento que expressa a possibilidade de um dado segmento da imagem, no tempo t , pertencer a uma determinada classe, tendo em vista a classe à qual pertencia em $t-\Delta t$.

²⁶ - As transições estão associadas a um valor que expressa a possibilidade de ocorrência de alterações de classes. Com isso, pode-se restringir o número de classes a considerar para cada elemento no processo de classificação.

²⁷ - São superfícies de separação, ou superfícies de decisão, que determinam os limites entre duas ou mais classes. Essas superfícies têm dimensão $n-1$, onde n é a dimensão do vetor de características. (SILVA e TOZZI, 2006)

²⁸ - Modalidade de conhecimento que expressa relações entre a resposta espectral do segmento na imagem adquirida no tempo t e as classes idealizadas.

2 - MODELOS DE REGRESSÃO

Tendo em vista as teorias e os modelos sobre mudança no uso e na cobertura do solo abordados no Capítulo 1, este estudo propõe um modelo dinâmico de regressão logística, com ênfase nas relações espaciais entre as células analisadas, a ser implementado no âmbito de um sistema de informação geográfica com fins preditivos. Para melhor compreensão de suas principais características, são detalhados, a seguir, os modelos de regressão, em geral, e os de regressão logística em particular.

Como um dos vários modelos que compõem as análises estatísticas multivariadas, o método de regressão é largamente utilizado em negócios, ciências biológicas, ciências sociais e comportamentais, bem como em outras disciplinas. Sua aplicação serve a três propósitos principais: descrição, controle e predição²⁹. (NETER *et al.*, 1996)

Para os fins de descrição, entende-se que a informação deva ser confiável (replicável, passível de repetição) e válida (relevante). O controle, por sua vez, implica uma amostragem razoável das prováveis ocorrências, sempre que possível e apropriado, ao invés de uma aceitação passiva e oportunista dos dados, como forma de permitir o controle ou de contrabalançar o risco de um viés empírico. Finalmente, a predição pressupõe que a informação deva ser válida para observações passadas, presentes e futuras de um dado fenômeno. Na prática, descrição, controle e predição se superpõem com frequência.

O conceito de uma relação entre duas variáveis assume duas modalidades distintas: a funcional e a estatística. Uma relação funcional entre duas variáveis é expressa por uma equação matemática. Se X denota a variável independente e Y a variável dependente, a relação funcional adota a forma $Y = f(X)$.

No caso da relação estatística, dada por $Y = f(X) + \varepsilon$, sendo ε um termo de erro aleatório, as observações apresentam uma inexatidão intrínseca à sua natureza e não se enquadram diretamente na curva da relação. Apesar disso, as relações estatísticas podem ser muito úteis para uma ampla gama de aplicações.

²⁹ - Esses propósitos pertencem a um conjunto consensuado de características comuns ao método científico. Outras características incluem a observação, a “refutabilidade” (*falsifiability*) e a explanação causal.

Uma dessas aplicações é a análise de regressão, um método estatístico que utiliza a relação entre duas ou mais variáveis quantitativas, de maneira que uma variável possa ser predita a partir de outra, ou de outras variáveis.

O método foi originalmente desenvolvido por Sir Francis Galton, ao final do século XIX. Galton estudou a relação entre as alturas de pais e filhos e notou que as alturas dos descendentes, tanto de pais altos quanto de pais baixos, aparentavam “reverter” ou “regredir” para a média do grupo. Ele considerou ser essa tendência uma regressão à “mediocridade”. (NETER *et al.*, 1996)

Galton desenvolveu uma descrição matemática dessa tendência de regressão, que se tornou precursora dos modelos de regressão atuais. O termo ‘regressão’ persiste até hoje para descrever relações estatísticas entre variáveis. Sua modelagem expressa, formalmente, os dois ingredientes de uma relação estatística. Em primeiro lugar, há uma tendência de a variável de resposta Y variar com a variável preditora X sistematicamente. Em segundo lugar, nessa modelagem ocorre uma dispersão de pontos em torno da curva de relação estatística.

Essas características são incorporadas em um modelo de regressão, pelos seguintes postulados:

- Há uma distribuição de probabilidades de Y para cada X ;
- As médias dessa distribuição de probabilidade variam de maneira sistemática com X .

Considerando que a realidade deva ser reduzida a proporções manipuláveis, sempre que se constroem modelos, somente um número limitado de variáveis preditoras ou explanatórias (ou, ainda, explicativa) pode – ou deve – ser incluída em um modelo de regressão para qualquer situação de interesse.

Um problema central em muitos estudos exploratórios, portanto, é a escolha, para um modelo de regressão, de um conjunto de variáveis preditoras que seja “bom”, no sentido dos propósitos da análise. Uma consideração importante nessa escolha é o alcance da contribuição com vistas a reduzir a variação remanescente em Y , depois que as contribuições de outras variáveis incluídas tentativamente no modelo tenham sido consideradas.

A escolha da forma funcional da relação de regressão está presa à escolha das variáveis preditoras. Em geral, a forma funcional tem que ser decidida empiricamente, uma vez que os dados tenham sido coletados. Funções de regressão linear ou quadrática são frequentemente usadas como aproximações iniciais satisfatórias.

Ao se formular um modelo de regressão, normalmente é necessário restringir a cobertura do modelo a algum intervalo ou região de valores da variável preditora. O escopo pode ser definido tanto pela concepção do estudo, como pelos tipos de dados disponíveis.

Outra questão importante nas regressões deve ser destacada. A existência de uma relação estatística entre a variável de resposta Y e a variável preditora X não implica causalidade na dependência de Y em relação a X . Por mais forte que seja a relação estatística entre X e Y , não há, necessariamente, um padrão de causa e efeito ditado pelo modelo de regressão. Em função disso, há que se tomar muito cuidado ao estabelecer conclusões sobre relações causais, a partir de análises de regressão. Por não prover informação acerca de padrões de causalidade, a regressão precisa ser suplementada por análises adicionais, para se obter perspectivas sobre relações causais.

2.1 - REGRESSÃO LINEAR

2.1.1 - Regressão linear simples

Quando o modelo de regressão possui uma única variável explicativa e a função de regressão é linear, tem-se a seguinte relação estatística:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.1)$$

onde:

Y_i é o valor da variável de resposta para o i -ésimo evento;

β_0 e β_1 são parâmetros;

X_i é uma constante conhecida, mais precisamente o valor da variável preditora no i -ésimo evento;

ε_i é um termo de erro aleatório com média $E\{\varepsilon_i\}=0$ e variância $\sigma^2\{\varepsilon_i\}=\sigma^2$;

$i = 1, \dots, n$.

Esse modelo (2.1) é dito ‘simples’, ‘linear nos parâmetros’, e ‘linear na variável preditora’. É ‘simples’, no sentido de que há somente uma variável preditora. É ‘linear nos parâmetros’, porque nenhum parâmetro surge como um expoente ou é multiplicado ou dividido por outro parâmetro. Finalmente, é ‘linear na variável preditora’, porque essa variável aparece na primeira potência. Um modelo com essas características é também chamado de ‘modelo de primeira ordem’.

Os parâmetros β_0 e β_1 no modelo de regressão (2.1) são denominados ‘coeficientes de regressão’, sendo β_1 a declividade da linha de regressão – indica a mudança na média da probabilidade de distribuição de Y por unidade acrescida de X – e β_0 o intercepto em Y da linha de regressão.

Quando o escopo do modelo inclui $X=0$, β_0 dá a média da probabilidade de distribuição de Y para $X=0$. Quando o escopo do modelo não cobre $X=0$, β_0 não possui um significado particular, como um termo em separado do modelo de regressão.

2.1.2 - Regressão Linear Múltipla

Modelos de regressão também podem conter mais de uma variável preditora. Nesse caso, eles são denominados modelos de regressão múltipla – dado que contêm variáveis explicativas adicionais – e são mais úteis para fornecer previsões precisas da variável resposta.

Modelos de regressão múltipla também são bastante úteis em situações experimentais, onde é possível controlar as variáveis explicativas. Não obstante, tais modelos podem igualmente ser usados para dados observados aleatoriamente. Quando há duas variáveis explicativas, o modelo de regressão é dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (2.2)$$

Um modelo assim é denominado de modelo de primeira ordem com duas variáveis. Nesse caso, a função de regressão pode ser chamada de ‘superfície de regressão’ ou de ‘superfície de resposta’.

Quando se considera o caso onde há $p-1$ variáveis explicativas (X_1, \dots, X_{p-1}), o modelo possui a seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (2.3)$$

Aqui, a função de resposta é um hiperplano, que é um plano de mais de duas dimensões. Por conta disso, não é possível observá-lo graficamente.

Os modelos de regressão linear são aproximações satisfatórias para a maior parte das aplicações de regressão. No entanto, há situações em que um modelo não linear é mais apropriado, seja por indicações empíricas, seja por justificativas teóricas.

Nesses casos, quando a variável dependente de interesse possui apenas duas respostas qualitativas³⁰ possíveis, utiliza-se a regressão logística, tanto para fins descritivos, quanto para fins preditivos.

A regressão com variável dependente binária apresenta dois tipos de problema para os estimadores dos seus coeficientes – um de natureza conceitual e outro de natureza estatística – que tornam necessária a adoção de uma regressão alternativa à regressão linear.

i. Problema de ordem funcional

O problema conceitual com a regressão linear cuja variável dependente é dicotômica decorre do fato de que as probabilidades possuem, por definição, valores máximo e mínimo ‘0’ e ‘1’ respectivamente. Ainda assim, a linha de regressão linear pode se estender para cima ou para baixo, em direção a ‘ $+\infty$ ’ ou ‘ $-\infty$ ’. A possibilidade de predição de valores maiores que ‘1’ ou menores que ‘0’ para variáveis dependentes binárias não faz sentido e não possui aplicação preditiva.

Uma forma funcional da relação que possui mais sentido teórico é obtida, quando ela se aproxima da linearidade para valores intermediários e adota uma curvatura suave, na medida em que se aproxima dos extremos. Próximo a esses limites, são necessárias mudanças cada vez maiores nos valores de X para se obter variações em Y . (Figura 2.1)

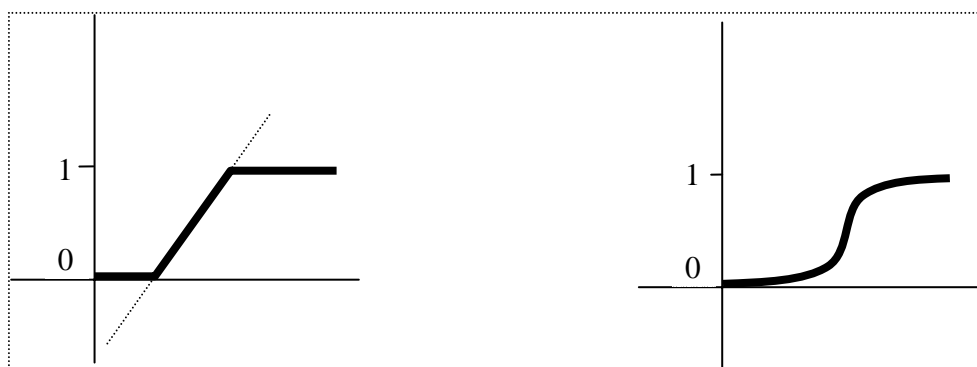


Fig. 2.1 – Problemas intrínsecos à regressão linear, quando a variável dependente é binária.

ii. Problema de inferência estatística

Ainda que uma linha reta se aproxime da relação não linear em alguns casos, surgem problemas que reduzem sua eficiência, mesmo em se conservando os estimadores não tendenciosos. Esses problemas envolvem o fato de que uma regressão com uma variável

³⁰ - Essas variáveis, também conhecidas como binárias, podem ser representadas pelos valores 0 e 1.

dependente binária viola as suposições de normalidade e de homocedasticidade³¹ e decorrem da existência de apenas dois valores observados para a variável dependente.

A heterocedasticidade apresenta implicações bem mais sérias do que a ausência de normalidade. Mesmo com amostras grandes, os erros-padrão, na presença de heterocedasticidade, são incorretos e os testes de significância, inválidos. Assim, o uso de regressão com uma variável dependente binária permanece inapropriado, sendo recomendado o uso da regressão logística, cujas principais características são apresentadas a seguir.

2.2 - REGRESSÃO LOGÍSTICA

2.2.1 - Origem do conceito

A definição de regressão logística deve levar em conta, separadamente, os termos ‘regressão’ e ‘logística’. O termo ‘curva logística’ é atribuído a Edward Wright (1558-1615) que, aparentemente, estava se referindo à curva logarítmica e não ao termo no sentido moderno. Galton, por outro lado, quando descobriu o fenômeno estatístico da regressão, utilizou originalmente o termo ‘reversão’.

Pierre François Verhulst (1804-1849) introduziu o termo ‘*logistique*’ como sendo aplicado à curva sigmóide. Acredita-se que ‘logístico’, na forma concebida por Verhulst, se refere às qualidades logarítmicas da curva. Em 1838, após ler o “ensaio sobre o princípio populacional”, de Thomas Malthus, Verhulst elaborou sua equação *logistique* para descrever o crescimento autolimitado de uma população biológica. A equação é também conhecida como a equação Verhulst-Pearl, em função de sua redescoberta em 1920.

Alfred Lotka deduziu-a novamente em 1925, chamando-a de lei do crescimento populacional. Aplicação típica da equação logística, o modelo de crescimento populacional estabelece que:

- A taxa de reprodução é proporcional à população existente, *ceteris paribus*;
- A taxa reprodução é proporcional ao montante de recursos disponíveis, *ceteris paribus*.

³¹ - A igualdade de variâncias sobre uma população é chamada **homogeneidade de variâncias** ou **homocedasticidade**. Alguns testes estatísticos assumem que o dados vêm de populações que têm a mesma variância. Se esta condição de homogeneidade de variâncias não é conhecida, o resultado do teste estatístico pode não ser válido. Nesse caso, a suposição de homocedasticidade é violada, porque o termo do erro da regressão varia com o valor de X. Por conta disso, a variância de ϵ não é constante. (NETER *et al.*, 1996)

Assim sendo, o segundo termo modela a competição por recursos disponíveis, que tende a limitar o crescimento populacional.

Sendo P o tamanho da população e t o tempo, esse modelo é formalizado pela seguinte equação diferencial:

$$\frac{\delta P}{\delta t} = kP(C - P) \quad (2.4)$$

onde a constante k define a taxa de crescimento e C é o limite máximo de crescimento (*carrying capacity*). A solução geral para essa equação é uma função logística, conforme mostrada na Figura 2.2.

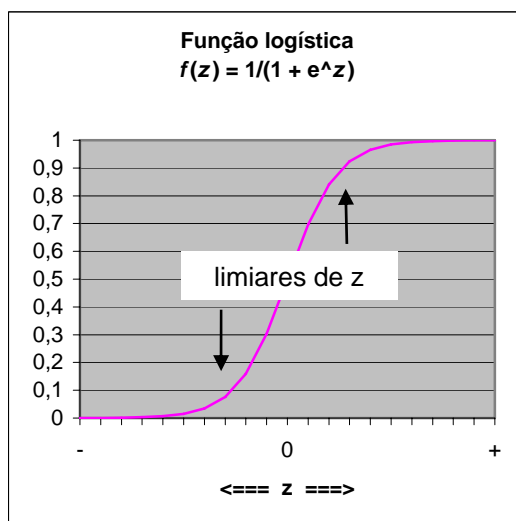


Fig. 2.2 – Gráfico da função logística.

O fato de a função logística $f(z)$ variar de ‘0’ a ‘1’ é a principal razão de sua popularidade (por exemplo, $0 \leq$ probabilidade de desmatamento ≤ 1). Outra razão está ligada à forma sigmóide da função logística. A sigmóide indica que o efeito de ‘ z ’ sobre um dado risco é mínimo para valores baixos de ‘ z ’, até que um determinado limiar seja atingido. O risco então sobe rapidamente durante uma faixa intermediária de valores de ‘ z ’, e se mantém próximo de ‘1’, tão logo se torne suficientemente grande.

Em suma, a curva sigmóide, que descreve o efeito combinado de vários fatores implicados no risco de ocorrência do fenômeno estudado, possui estimadas que devem variar entre ‘0’ e ‘1’.

2.2.2 - O Modelo Logístico

Tal como a regressão linear, a regressão logística pode ser simples (uma variável explicativa) ou múltipla (duas ou mais).

Uma forma simplificada de entendimento dos modelos de regressão logística – também denominados de modelo logito – é percebê-los como a regressão de variável dependente que transforma relações não lineares em relações lineares. Ao fazer essa transformação, a regressão logística passa a interpretação de coeficientes a partir de mudanças em probabilidades, para mudanças (menos intuitivas) das chances³² logaritmizadas – os logitos.

O modelo logito foi introduzido por Berkson em 1944, que cunhou o termo. Barnard, em 1949, criou o termo comumente conhecido por *log-odds* ou chances logaritmizadas. A *log-odd* de um evento é o logito da probabilidade desse evento. (NETER *et. al.*, 1996)

Sua forma mais simples é dada por logito (p_i) = $\beta_0 + \beta_i x_i$, onde x_i é alguma quantidade da qual pode depender o “sucesso” ou a “falha” no *i-ésimo* evento de uma seqüência de tentativas de Bernoulli, e p_i é a probabilidade de sucesso no *i-ésimo* evento.

Uma vez observados os valores de x em uma seqüência de casos, e ciente dos “sucessos” e “falhas” de cada caso, um estatístico freqüentemente estima os valores dos coeficientes ‘ β_0 ’ e ‘ β_i ’ pelo método da máxima verossimilhança. O resultado pode, então, ser usado para avaliar a probabilidade de “sucesso” em um caso subsequente, no qual o valor de x é conhecido. Estimação e predição por esse método são denominadas regressão logística. (THOMPSON, 1992: 1445, In: SILVA, 2003)

Uma característica importante do modelo logístico é o fato de ser orientado para estudos de acompanhamento (*follow-up study*³³), o que não ocorre com estudo de controle de caso (*case-control study*³⁴) e estudos transversais (*cross-sectional study*³⁵). No caso destes dois últimos, não se pode prever o risco individual, mas somente as estimadas de razão de chances (que também podem ser obtidas a partir de estudos de acompanhamento).

³² - Chance é a razão da probabilidade de que um evento ocorra pela probabilidade de que ele não ocorra ($P/(1-P)$). (NETER *et al.*, 1996)

³³ - Um estudo de acompanhamento descreve a probabilidade de um evento ocorrer, em função de variáveis independentes presumivelmente medidas ao início de um período fixado de acompanhamento (*follow-up*). (NETER *et al.*, 1996)

³⁴ - Trata-se de um método de estudo epidemiológico, no qual se comparam os fatores de risco de pessoas portadoras e não portadoras de uma doença de interesse. Disponível em http://www.biology-online.org/dictionary/case-control_study. Acesso em 6-12-05.

³⁵ - Também conhecido como estudo horizontal, é o estudo no qual a presença ou ausência de uma determinada doença de interesse ou de outras variáveis relacionadas à saúde são determinadas em cada indivíduo da população estudada, ou em uma amostra representativa, para um dado ponto no tempo. Difere do estudo longitudinal, que é conduzido por um certo período de tempo. Disponível em <http://www.biology-online.org/search.php?search=cross-sectional+study>. Acesso em 6-12-05.

Considerando a função logística

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2.5)$$

e fazendo

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k, \quad (2.6)$$

obtém-se

$$f(z) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}}. \quad (2.7)$$

Em essência, z é um índice que combina todos os X_i . Com o objetivo de descrever a probabilidade de que um dado fenômeno Δ , por exemplo, se desenvolva durante um determinado período (de t_0 para t_1 , por exemplo) para um elemento (indivíduo) com variáveis independentes de valores X_1, X_2, \dots, X_k , medidos em t_0 , $f(z)$ pode ser descrita em termos de probabilidade condicional e denotada por $P(\Delta=1 | X_1, X_2, \dots, X_k)$ que, por conveniência, recebe a notação $P(\Delta)$.

Assim, a equação do modelo é apresentada como:

$$P(\Delta) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \quad (2.8)$$

Os termos β_0 e β_i representam parâmetros desconhecidos, que precisam ser estimados com base nos valores de X_i e de Δ para um grupo de indivíduos.

O conjunto de dados é utilizado para ajustar o modelo, ou seja, para estimar os parâmetros desconhecidos β_0 e β_i . Os estimadores desses parâmetros são denotados por $\hat{\beta}_0$ e $\hat{\beta}_i$ e o método de estimação mais utilizado é o de máxima verossimilhança (Maxver).

2.2.3 - A transformação logito

O logito se inicia pela transformação de probabilidades em chances³⁶ que, ao contrário das probabilidades, não possuem um limite superior e são úteis para a interpretação dessas probabilidades.

³⁶ - A chance de que o evento seja observado é dada por meio da razão entre ' p_i ' e ' $1-p_i$ ', sendo p_i a probabilidade de se ter uma característica ou de experimentar um evento. Tal como a probabilidade, as chances apresentam um valor mínimo de '0', mas, ao contrário daquela, não possui limite máximo. É comum comparar duas chances como uma razão. Entretanto, deve-se distinguir razão de chances (*odds ratio*) e chances propriamente ditas. Chances são razões de probabilidades, enquanto *odds ratio* é a razão entre razões de probabilidades. (PAMPEL, 2000)

Como a probabilidade implica necessariamente um valor entre ‘0’ e ‘1’, pelo menos duas transformações são necessárias para que seja possível lidar adequadamente com a estrutura binária da variável dependente. Primeiramente, são calculadas as chances para, em seguida, se obter o logaritmo natural das chances.

Com essa transformação, é produzida uma variável que pode apresentar valores de ‘ $-\infty$ ’ a ‘ $+\infty$ ’. Tal transformação, denominada logito³⁷, adquire valores negativos (tendendo a ‘ $-\infty$ ’), quando a chance de ocorrência do evento varia de ‘1’ a ‘0’, e adquire valores positivos (tendendo a ‘ $+\infty$ ’), quando a chance desse evento varia de ‘0’ a ‘1’. Ao se utilizar o logaritmo natural da chance de $Y=1$ (expressão que fornece o logito) como variável dependente, a equação deixa de ficar limitada a valores situados entre ‘0’ e ‘1’.

As propriedades do modelo logito podem ser sintetizadas da seguinte forma:

- Ao contrário da probabilidade, o logito não possui limite inferior ou superior. As chances eliminam o limite superior, enquanto seu logaritmo elimina o limite inferior das probabilidades.
- A transformação logito é simétrica em torno da probabilidade de 0,5 – com sinais trocados.
- A mesma mudança em probabilidades traduz-se em diferentes mudanças nos logitos. Em outras palavras, pequenas diferenças em probabilidades resultam em diferenças cada vez maiores nos logitos, quando as probabilidades estão próximas dos valores limites ‘0’ e ‘1’.

Ao realizar a transformação logito, portanto, uma mudança em X provoca efeitos similares ao longo dos valores da transformação logito das probabilidades de Y . Assim, a linearidade proporcionada pelo logito define, com as probabilidades, relações não lineares bastante significativas.

A relação linear de X com o logito aparece em

$$L_i = \text{logito}(p_i) = \ln[p_i/(1-p_i)] = \beta_0 + \beta_i X_i. \quad (2.9)$$

Fazendo

$$p_i/(1-p_i) = e^{\beta_0 + \beta_i X_i} = e^{\beta_0} * e^{\beta_i X_i}, \quad (2.10)$$

significa dizer que a chance muda em função dos coeficientes tratados como expoentes.

³⁷ - Escreve-se $\text{logito}(Y)$.

Por dedução, tem-se

$$p_i = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}}. \quad (2.11)$$

Mas

$$L_i = \beta_0 + \beta_1 X_i \Rightarrow p_i = \frac{e^{L_i}}{1 + e^{L_i}}. \quad (2.12)$$

Uma mudança em L decorrente da mudança de uma unidade em X resulta em uma mudança maior nas probabilidades intermediárias de X do que nos limites inferior e superior. Essa não linearidade entre o logito e a probabilidade cria um problema fundamental de interpretação. Enquanto o efeito de X no logito obedece a uma relação linear, o mesmo não pode ser dito entre X e a probabilidade.

O fato é que $p_i/(1-p_i)$, cujo logaritmo fornece o logito, descreve a chance de ocorrer o fenômeno Δ em um dado indivíduo com variáveis independentes especificadas por X . Assim, a forma logito do modelo logístico fornece uma expressão para o logaritmo da chance de um indivíduo, dado um certo conjunto de variáveis X .

2.2.4 - Derivação da equação da razão de chances

Qualquer razão de chances, por definição, é uma razão entre duas chances, como, por exemplo, a chance de um grupo '1' (referente a um tempo t_1) dividida pela chance de um grupo '0' (referente a um tempo inicial t_0).

Dado o modelo logístico genérico

$$P(X) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}}, \quad (2.13)$$

pode-se escrever a chance para o grupo '1' como sendo $P(X_1)$ dividido por $1 - P(X_1)$ e para o grupo '0' como $P(X_0)$ dividido por $1 - P(X_0)$. À razão de chances obtida a partir da divisão entre as razões de chances dos dois grupos, dá-se o nome de razão de chances de risco (RCR), dado que as probabilidades de chances são todas definidas como riscos. Assim, tem-se que:

$$RCR_{X_1, X_0} = \frac{\frac{P(X_1)}{1 - P(X_1)}}{\frac{P(X_0)}{1 - P(X_0)}}. \quad (2.14)$$

Considerando

$$P(X) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \Rightarrow \frac{P(X)}{1 - P(X)} = e^{(\beta_0 + \sum \beta_i X_i)} \Rightarrow RCR_{X_1, X_0} = e^{\sum \beta_i (X_{1i} - X_{0i})} \quad (2.15)$$

Tem-se que,

$$\frac{e^a}{e^b} = e^{a-b} \Rightarrow RCR_{X_1, X_0} = e^{(\beta_0 + \sum \beta_i X_{1i}) - (\beta_0 + \sum \beta_i X_{0i})} \Rightarrow RCR_{X_1, X_0} = e^{\sum \beta_i (X_{1i} - X_{0i})} \quad (2.16)$$

Esta é a equação exponencial geral para a RCR de um modelo logístico que compare quaisquer dois grupos de indivíduos, especificados em termos de X_1 e X_0 . Observe-se que a equação envolve β_i , mas não β_0 .

Considerando que

$$e^{\sum_{i=1}^k z_i} = e^{z_1} \times e^{z_2} \times \dots \times e^{z_k}, \text{ e fazendo } z_i = \beta_i (X_{1i} - X_{0i}), \quad (2.17)$$

também se pode escrever, alternativamente,

$$RCR_{X_1, X_0} = \prod_{i=1}^k e^{\beta_i (X_{1i} - X_{0i})} = e^{\beta_1 (X_{11} - X_{01})} \times e^{\beta_2 (X_{12} - X_{02})} \dots \times e^{\beta_k (X_{1k} - X_{0k})}. \quad (2.18)$$

Significa dizer que a equação de RCR permite interpretar que cada uma das variáveis X contribui de forma multiplicativa para a razão de chances.

Dependendo de sua formulação, um modelo logístico pode implicar uma contribuição aditiva, por exemplo, das variáveis para com a razão de chances.

De acordo com Kleinbaum e Klein (2002), em termos computacionais, a formulação mais útil é a do exponencial da soma, dada por

$$RCR_{X_1, X_0} = e^{\sum_{i=1}^k \beta_i (X_{1i} - X_{0i})}. \quad (2.19)$$

Para variáveis categorizadas (0,1), pode-se obter uma razão de chances ajustada, por meio da exponencial do coeficiente correspondente a essa variável. Dada uma variável categorizada X_i , a sua razão de chances de risco ajustada se torna

$$RCR_{ajustada} = e^{\beta_i}. \quad (2.20)$$

2.2.5 - Interpretação de Coeficientes de Regressão Logística

Matematicamente, β_0 possui dois significados. Em primeiro lugar, dado que todas as variáveis X_i sejam nulas, o logito $P(X) = \beta_0$, o que não permite uma interpretação significativa,

dado que nenhum indivíduo teria como valor nulo variáveis que ocorrem naturalmente, como distâncias, por exemplo.

O segundo significado, no entanto, possui um apelo interessante. Nele, β_0 descreve o logaritmo da chance do *background* (ou *baseline*). Por chance do *background*, faz-se referência à chance resultante de um modelo logístico destituído das variáveis X_i .

Os efeitos das variáveis independentes na regressão logística possuem múltiplas interpretações, diferentes para probabilidades, chances e logitos. Cada uma delas possui vantagens e desvantagens.

Os efeitos das variáveis independentes nos logitos são lineares e aditivos, mas as unidades da variável dependente (logito) possuem pouco significado intuitivo.

Os efeitos das variáveis independentes nas probabilidades possuem significado intuitivo, mas não são lineares e nem aditivos, o que significa dizer que eles não podem ser simplesmente resumidos na forma de um coeficiente único.

A interpretação dos efeitos das variáveis independentes sobre as chances oferece um compromisso entre as alternativas anteriores. Os efeitos sobre as chances são multiplicativos, ao invés de aditivos, mas conservam uma interpretação direta.

i. Logitos

A interpretação dos logitos usa diretamente os coeficientes obtidos a partir dos estimadores da regressão logística, que simplesmente mostram a mudança nos logitos preditos ao se observar um evento ou de adotar uma característica por conta de uma mudança de uma unidade nas variáveis independentes.

A despeito da simplicidade de sua interpretação, os coeficientes da regressão logística carecem de uma métrica dotada de significado. Afirmações sobre os efeitos de variáveis nas mudanças dos logitos revelam pouco acerca das relações e pouco ajudam na explicação de resultados substantivos.

ii. Chances

Para achar os efeitos sobre as chances, basta retirar o expoente dos coeficientes da regressão logística, como no exemplo de duas variáveis a seguir:

$$\ln[p_i/(1-p_i)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \Rightarrow e^{\ln[p_i/(1-p_i)]} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2} \Rightarrow p_i/(1-p_i) = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2} \quad (2.21)$$

Deve-se notar que, nesse caso, um coeficiente 1 deixa as chances inalteradas; um coeficiente maior que 1 aumenta as chances; e um coeficiente menor que 1 diminui as

chances. Além disso, quanto mais distante de 1, em qualquer direção, estiver o coeficiente, maior será o efeito na mudança das chances.

Ao interpretar coeficientes exponenciais, há que se ter em mente que eles se referem a mudanças multiplicativas nas chances e não nas probabilidades.

iii. Probabilidades

Uma vez que as relações entre as variáveis independentes e as probabilidades não são lineares e nem aditivas, elas não podem ser completamente representadas por um único coeficiente. O efeito nas probabilidades tem que ser identificado para um valor específico ou conjunto de valores. A escolha de valores a usar na avaliação do efeito das variáveis nas probabilidades depende das inquietações do pesquisador e da natureza dos dados, mas uma estratégia inicial simples constitui-se da verificação do efeito na probabilidade para um caso típico.

No caso de variáveis independentes contínuas, uma forma rápida de dimensionar sua influência em probabilidades envolve o cálculo da declividade da tangente da curva não linear em qualquer ponto.

A mudança em probabilidade ou a declividade da tangente advém de uma simples equação para a derivada parcial, também conhecida como efeito marginal ou instantâneo, dada por:

$$\frac{\partial p}{\partial X_k} = \beta_k * p * (1 - p) \quad (2.22)$$

Basta multiplicar o coeficiente de regressão logística pela probabilidade selecionada ‘ p ’ e por ‘ $1-p$ ’.

A equação derivada parcial revela claramente os efeitos não lineares de uma variável independente nas probabilidades.

O efeito possui seu máximo, quando p é igual a 0,5, dado que $0,5 \times 0,5 = 0,25$; $0,6 \times 0,4 = 0,24$; $0,7 \times 0,3 = 0,21$; e assim por diante. Quanto mais p se aproxima dos limites superior (1) e inferior (0), menor é o produto $p \times (1-p)$ e menor é o efeito que a mudança de uma unidade em X tem na probabilidade.

Críticos desse procedimento entendem que o coeficiente resultante induz ao erro e é pouca coisa melhor do que a utilização de regressão linear. Ainda assim, a tendência de pesquisadores de pensar em termos de proporções ou probabilidades pode indicar o uso da

declividade na tangente da média da variável dependente ou em outros pontos da curva logística como um suplemento para outras interpretações.

Para variáveis independentes binárias, a mudança relevante ocorre de 0 a 1, e a linha tangente para pequenas mudanças em X faz menos sentido. Ao invés disso, é possível computar variáveis preditas para cada grupo, e então medir as diferenças dos grupos em termos de probabilidade.

Como uma alternativa à derivada parcial, probabilidades preditas para variáveis independentes contínuas podem ser usadas para variáveis contínuas, da mesma forma que para variáveis binárias. Mudanças nas probabilidades preditas indicam o efeito real de uma mudança discreta em X (como uma unidade, por exemplo) e não o efeito na linha de tangente, implicada por uma mudança instantânea ou infinitesimal em X . Por essa razão, alguns preferem o uso de probabilidades preditas para uma mudança discreta na derivada parcial. (KAUFMAN, 1996; LONG, 1997)

iv. Coeficientes padronizados

Normalmente em regressão, multiplicam-se coeficientes não padronizados pela razão do desvio-padrão de X pelo desvio-padrão de Y para obter coeficientes padronizados. O produto fornece coeficientes idênticos àqueles que seriam obtidos se, antes de serem colocadas no programa de regressão, as variáveis fossem primeiramente transformadas em valores-padrão (média zero e desvio-padrão igual a '1').

Ao contrário de programas de regressão múltipla, os programas de regressão logística não computam coeficientes padronizados com frequência. Esse problema decorre, em parte, da ambigüidade do significado de valores-padrão ou unidades-padrão para as variáveis binárias. Variáveis binárias padronizadas meramente traduzem valores de '0' e '1' em outros dois valores. Se a média da variável dependente é igual à probabilidade p , a variância é igual a $p \times (1-p)$.

Assim, valores de Y iguais a '1' possuem valores z (padronizados) iguais a $(1-p)/\sqrt{p \times (1-p)}$ e valores de Y iguais a '0' apresentam valores de z iguais a $(0-p)/\sqrt{p \times (1-p)}$.

Com apenas dois valores para a variável binária padronizada, a alusão a uma mudança de desvio-padrão carece de referência concreta. Pelo fato de uma mudança de desvio-padrão em uma variável binária não ter tipicamente o mesmo significado que uma mudança de desvio-

padrão em uma variável contínua, é comum se evitar o uso de coeficientes padronizados para variáveis binárias.

2.3 - O PROBLEMA MULTIVARIADO DA PESQUISA

Conforme mencionado anteriormente, os modelos de regressão logística estão entre as diversas técnicas, compreendidas pela análise multivariada, dedicadas à análise de conjuntos de dados que contenham mais de uma variável. Várias dessas técnicas – como é o caso do modelo de regressão logística aqui adotado – foram desenvolvidas recentemente, em parte porque requerem a capacidade computacional dos computadores modernos para sua implementação.

O problema multivariado típico pode ser descrito da seguinte forma:

Qual é a relação entre uma ou mais variáveis de estudo e um determinado efeito resultante?

Ao aplicar esse problema ao efeito dicotômico do desmatamento, no qual se atribui ‘0’ para áreas não desmatadas e ‘1’ para áreas desmatadas, uma coleção de variáveis independentes é usada para descrever ou prever a variável dependente *IPT* (impacto ambiental – desmatamento). Essa coleção é formada pela variável de exposição *PAD* (proximidade da área desmatada), juntamente com uma série de variáveis de controle C_1, C_2, \dots, C_n . No caso desta pesquisa, são as seguintes as variáveis C_i usadas na formulação inicial do modelo de regressão logística:

C_1 – ROD (proximidade de rodovias);

C_2 – HID (proximidade de hidrovias);

C_3 – EPA (existência – ou não – de proteção ambiental no polígono em torno da célula);

C_4 – DEC (declividade do terreno);

C_5 – ALT (altitude do terreno);

C_6 – IFR (índice de fragmentação); e

C_7 – APE (área de floresta no polígono envolvente);

Todas as variáveis independentes (*PAD* e C_i) podem ser denotadas por $X_1, X_2 \dots X_k$, onde k é o número considerado de variáveis.

Kleinbaum e Klein (2002) descrevem os casos especiais do modelo logístico geral, que incluem a análise simples (matriz 2×2) e as análises que envolvem interação multiplicativa

entre as variáveis independentes. Para cada caso, são consideradas as definições de variáveis no modelo e a fórmula para a razão de chances que descreve a relação entre *IPT* e *PAD*.

2.3.1 - Análise Simples

Como a variável *IPT* é considerada no modelo logístico como sendo dicotômica, sua relação com uma única variável de exposição (proximidade da área desmatada = *PAD*), por exemplo, pode ser expressa por uma matriz 2×2 , capaz de permitir a caracterização dessa análise de uma forma convencionada como simples.

Por conveniência, define-se a variável de exposição como uma variável (0,1) e dispõe-se seus valores nas duas colunas da matriz. O mesmo é feito com a variável *IPT*, sendo que seus valores são distribuídos nas linhas. Com isso, as frequências das células da matriz são denotadas por 'a', 'b', 'c' e 'd', conforme o quadro abaixo.

	<i>IPT=1</i>	<i>IPT=0</i>
<i>PAD=1</i>	<i>a</i>	<i>b</i>
<i>PAD=0</i>	<i>c</i>	<i>d</i>

O modelo logístico para a análise simples pode ser definida pela expressão

$$P(IPT) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 PAD)}} \quad (2.23)$$

onde *PAD* assume o valor '1' para áreas próximas da área desmatada e '0' para áreas distantes.

O termo $P(IPT)$ denota a probabilidade de que o desmatamento assumo o valor '1', a partir de um dado valor para *PAD*. Em outras palavras, essa probabilidade expressa o risco de impacto ambiental – desmatamento, dada a proximidade da célula de floresta à área desmatada.

Quando *PAD* é igual a '1', tem-se o risco R_1 , que é a probabilidade condicional de $IPT=1$, dado $PAD=1$. Quando *PAD* é igual a '0', tem-se o risco R_0 , que é a probabilidade condicional de $IPT=1$, dado $PAD=0$.

É possível usar esse modelo de análise simples para comparar áreas desmatadas com áreas não desmatadas. Utilizando os termos R_1 e R_0 , pode-se escrever a razão de chances de risco como sendo:

$$RCR_{PAD=1 \text{ vs. } PAD=0} = \frac{\frac{R_1}{1-R_1}}{\frac{R_0}{1-R_0}} \quad (2.24)$$

Para se computar a razão de chances em termos dos parâmetros do modelo logístico, substitui-se sua expressão na fórmula da razão de chances. Assim,

$$\begin{aligned} PAD = 1 &\Rightarrow R_1 = \frac{1}{1 + e^{-(\beta_0 + [\beta_1 \times 1])}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1)}} \\ PAD = 0 &\Rightarrow R_0 = \frac{1}{1 + e^{-(\beta_0 + [\beta_1 \times 0])}} = \frac{1}{1 + e^{-\beta_0}} \\ RCR &= \frac{\frac{R_1}{1-R_1}}{\frac{R_0}{1-R_0}} = \frac{\frac{1}{1 + e^{-(\beta_0 + \beta_1)}}}{\frac{1}{1 + e^{-\beta_0}}} = e^{\beta_1} \end{aligned} \quad (2.25)$$

onde β_1 é o coeficiente da variável de exposição PAD.

Em suma, para o modelo de análise simples, que envolve uma variável de exposição (0,1), têm-se as seguintes definições:

$$P(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 PAD)}} \quad e \quad RCR = e^{\beta_1} \quad (2.26)$$

Por sua construção simples, a estimativa da razão de chances pode ser obtida pela divisão do produto das frequências ad/bc nas células da matriz 2×2 gerada pelo modelo. Assim,

$$\hat{RCR} = e^{\hat{\beta}} = ad/bc \quad (2.27)$$

Nesse caso de análise simples, não é necessário o uso de computadores para estimar a razão de chances. Para casos especiais mais complexos, que envolvem mais de uma variável independente, no entanto, há que se lançar mão de programas computacionais para tais cálculos.

2.3.2 - Interação Multiplicativa

A investigação de casos com interação multiplicativa entre duas variáveis independentes envolve uma equação com razões de chances de riscos (RCR), que correspondem a diferentes combinações entre as variáveis.

Nesses casos, as razões de chances são definidas em termos de riscos. Seja R_{AB} o risco de impacto ambiental – desmatamento, dados valores específicos para as variáveis independentes A e B, ou, em outras palavras, seja R_{AB} a probabilidade condicional de $IPT=1$, dados A e B.

Como A e B são variáveis dicotômicas, há quatro valores possíveis para R_{AB} , conforma a matriz 2x2 abaixo.

	$B=1$	$B=0$
$A=1$	R_{11}	R_{10}
$A=0$	R_{01}	R_{00}

Quando A e B são iguais a ‘1’, o risco R_{AB} é dado por R_{11} . De maneira similar, para $A=1$ e $B=0$, tem-se R_{10} ; para $A=0$ e $B=1$, tem-se R_{01} . Por fim, para $A=0$ e $B=0$, obtém-se R_{00} .

Há que se ter em mente que a matriz 2x2 gerada nesses casos não descreve uma análise simples, pois os cabeçalhos da linha e da coluna denotam duas variáveis independentes, ao invés da relação entre uma variável independente e uma variável dependente. Mais do que isso, a informação contida nessa matriz é uma coleção dos quatro riscos correspondentes às diferentes combinações das duas variáveis independentes, e não de quatro frequências de células, correspondentes a diferentes combinações de ‘exposição – desmatamento’.

2.3.3 - Efeitos de confusão e de interação - o Modelo EVW

O modelo logístico que leva em conta os efeitos de diversas variáveis e, ainda, permite o controle para efeitos de confusão (*confounders*) e de interação é denominado modelo EVW. (KLEINBAUM e KLEIN, 2002)

Seja ‘PAD’ uma variável de exposição contínua e ‘p’ é um conjunto de controle C_1, C_2 e assim por diante, até C_p , podendo as variáveis C_1 a C_p tanto ser contínuas quanto categorizadas. No caso desta tese, o desmatamento $IPT_{(0,1)}$ é a variável-resposta (dependente) e a proximidade da área desmatada ($PAD_{(0,1)}$) é a variável de exposição. As variáveis de controle são: proximidade de rodovias (ROD), proximidade de hidrovias (HID), existência de proteção ambiental (EPA), declividade (DEC), altura (ALT), índice de fragmentação (IFR) e área do polígono envolvente (APE).

Dentre as variáveis de controle, as variáveis ‘ROD’, ‘HID’, ‘DEC’, ‘ALT’, ‘IFR’ e ‘APE’ são contínuas, enquanto ‘EPA’ é dicotômica. Embora sejam ‘p’ variáveis de controle, somadas à variável de exposição PAD, o modelo possui ‘p + q’ variáveis, na medida em que ‘q’ variáveis multiplicativas são acrescentadas, a saber W_1 (PAD x DEC) e W_2 (PAD x ALT).

Assim, o logito do modelo se torna

$$\begin{aligned} \text{logito}(IPT) = & \alpha_0 + \beta_1 PAD \\ & + \gamma_1 C_1 + \gamma_2 C_2 + \gamma_3 C_3 + \gamma_4 C_4 + \gamma_5 C_5 + \gamma_6 C_6 + \gamma_7 C_7 && \text{(efeitos principais)} && (2.28) \\ & + \delta_1 PAD \times C_4 + \delta_2 PAD \times C_5 && \text{(efeitos de interação)} \end{aligned}$$

onde:

α_0 = intercepto;

β_1 = parâmetro da variável de exposição;

γ_i = parâmetros das variáveis de confusão; e

δ_i = parâmetros das variáveis de interação potencial.

Os parâmetros γ e δ são aqui usados, além dos tradicionais β , para distinguir diferentes tipos de variáveis no modelo. Enquanto o parâmetro β é utilizado exclusivamente para caracterizar o coeficiente da variável de exposição *PAD* (*E*), γ indica os coeficientes de confusão potencial (*V*), e δ indica os coeficientes de interação potencial (*W*) no modelo.

Além da variável de exposição (*PAD*), o modelo geral contém p_1 variáveis denotadas por V_1, V_2, \dots, V_{p_1} , que são os potenciais *confounders*. [Ex: $V_1 = C_1, V_2 = (C_2)^2, \dots$ também podem ser idênticos ($V = C$), enfim, $V = f(C)$]

O termo *W* resulta de produtos da variável de exposição com as variáveis *C*, gerando os potenciais modificadores de efeito com a variável de exposição *PAD*. [Ex: $W_1 = PAD \times C_1, W_2 = PAD \times C_2$ etc.]

O modelo EVW considera o efeito de uma única variável de exposição submetida ao controle de efeitos potenciais de confusão e de interação e contém $p_1 + p_2 + 1$ variáveis, onde p_1 é o número de confundidores potenciais, p_2 é o número de termos de interação potenciais e 1 denota a variável de exposição. Neste trabalho, há $p_1 = 7$ variáveis de confusão em potencial, além de $p_2 = 2$ variáveis de interação, perfazendo um total de 10 variáveis. A forma geral do modelo é dada por:

$$\text{logito}(Y) = \alpha_0 + \beta_1 E + \gamma_1 V_1 + \gamma_2 V_2 + \dots + \gamma_{p_1} V_{p_1} + \delta_1 W_1 + \dots + \delta_2 W_{p_2} \quad (2.29)$$

onde:

α_0 = intercepto;

β_1 = coeficiente de *E*;

γ_i = coeficientes de *V*; e

δ_i = coeficientes de *W*.

2.3.4 - Estratégia de modelagem

Para se obter uma estimativa válida referente à relação entre as variáveis dependente e de exposição (independente), levando em conta as modificações de efeito e de confusão (variáveis independentes), são necessárias informações sobre como as variáveis são escolhidas para o modelo inicial, como elas são preservadas no modelo final, bem como os modificadores de efeito e os *confounders* são investigados, com vistas a determinar seu papel no modelo final.

Sem informação dotada de significado acerca da estratégia de modelagem utilizada, torna-se difícil investigar a validade dos resultados obtidos. Assim, há necessidade de guias referentes à estratégia de modelagem, para que o pesquisador saiba quais informações buscar para a construção do modelo.

Na prática, a maior parte das estratégias de modelagem é *ad hoc*; em outras palavras, o pesquisador cria a estratégia, na medida em que evoluem suas análises. Esses guias também são importantes para outros procedimentos de modelagem que não a regressão logística.

Inicialmente, são especificadas as variáveis dependente e de exposição, juntamente com o conjunto de variáveis C_1 a C_p , a ser considerado para fins de controle. Essas variáveis são definidas e medidas pelo pesquisador, tomando-se por base os objetivos do estudo e uma revisão da literatura e/ou corpo teórico do assunto pesquisado.

A seguir, é necessário especificar as variáveis do tipo V , que são funções das C_i que, por sua vez, são alimentadas no modelo como potenciais *confounders*. Genericamente, recomenda-se (KLEINBAUM e KLEIN, 2002) que a escolha dos V seja baseada primariamente em pesquisas anteriores ou no corpo teórico existente, levando-se em conta possíveis problemas estatísticos, como multicolinearidade, por exemplo, que possam resultar de certas escolhas.

A escolha mais simples para V é o conjunto de C propriamente dito. Se o número de C é muito grande, pode ser apropriado considerar um subconjunto de C de maior relevância e facilidade de interpretação, com base em conhecimento anterior.

Uma vez escolhidos os V , o próximo passo é determinar os W . Esses são os modificadores de efeito que vão para o modelo como termos de produtos com a variável de exposição E . Kleinbaum e Klein (2002) recomendam, ainda, que a escolha dos W seja restrita ou aos próprios V ou a produtos que envolvam dois V . Assim, os termos de produtos no modelo assumem a forma de $E \times V$ e $E \times V_i \times V_j$, onde V_i e V_j são diferentes entre si.

Para a maior parte das situações, recomenda-se que tanto V quanto W sejam os próprios C , ou mesmo um subconjunto de C . A razão para isso é que produtos do tipo EV_iV_j são muito difíceis de interpretar, mesmo quando são estatisticamente significativos. Além disso, esses termos tendem a provocar o fenômeno da multicolinearidade. Assim sendo, a melhor forma de evitar esses problemas é usar termos do tipo EV_i .

Ao se escolherem as variáveis V e W que serão incluídas no modelo inicial, é preciso assegurar que o modelo possua uma certa estrutura, a fim de evitar resultados enganosos. Tal

estrutura é denominada modelo hierarquicamente bem formulado – *HWF*. (KLEINBAUM e KLEIN, 2002)

O *HWF* deve satisfazer a seguinte característica: dada qualquer variável do modelo, todos os componentes de ordem inferior da variável deve também ser contido no modelo. Para entender essa definição, são dados a seguir dois exemplos de modelo, um bem formulado e um mal formulado hierarquicamente:

- Exemplo de modelo **bem** formulado hierarquicamente

$$\text{logito } P(X) = \alpha + \beta E + \gamma_1 V_1 + \gamma_2 V_2 + \delta_1 EV_1 + \delta_2 EV_2 \quad (2.30)$$

Observação: note-se que tanto os componentes de EV_1 (E, V_1) como os de EV_2 (E, V_2), que são de ordem inferior estão no modelo.

- Exemplo de modelo **mal** formulado hierarquicamente

$$\text{logito } P(X) = \alpha + \beta E + \gamma_1 V_1 + \gamma_2 V_2 + \delta_1 EV_1 + \delta_2 EV_2 + \delta_3 EV_1 V_2 \quad (2.31)$$

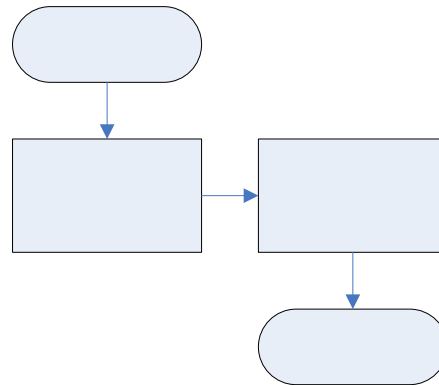
Observação: embora E, V_1, V_2, EV_1 e EV_2 estejam no modelo, o produto $V_1 V_2$, que é de uma ordem inferior a $EV_1 V_2$, não está contido no modelo.

A justificativa para se formular um modelo *HWF* está na necessidade de independência na codificação das variáveis, quando da realização de testes estatísticos. Um modelo mal formulado hierarquicamente pode apresentar diferentes resultados para uma variável de exposição dicotômica, em função de ela estar codificada como (0,1) ou (-1,1), por exemplo. O mesmo não acontece com um modelo *HWF*.

A partir de um modelo completo – *HWF* – o próximo passo é a eliminação de variáveis consideradas desnecessárias. A abordagem adotada para esta tese, conforme sugerida por Kleinbaum e Klein (2002), segue o diagrama de fluxo de dados apresentado na Figura 2.3.

As eliminações dessas variáveis adotam uma abordagem denominada eliminação reversa hierárquica de variáveis, porque se parte de um modelo maior para um menor, ao mesmo tempo em que as variáveis são tratadas em diferentes ordens (também da maior para a menor) em diferentes passos.

Na medida em que se avança no processo, algumas variáveis são eliminadas e outras são preservadas. Para aqueles termos preservados, há um princípio a ser observado na identificação de componentes de ordem inferior, que é o princípio da hierarquia. Em geral, ele estabelece que, se uma variável-produto é mantida no modelo, então todos os componentes de ordem inferior dessa variável devem ser mantidos.



Modelo inicial

Fig. 2.3 – Eliminação de variáveis no modelo EVW proposto por Kleinbaum e Klein (2002).

Como testes estatísticos para componentes de ordem inferior podem levar a diferentes conclusões, em função da codificação adotada, esses testes não são apropriados para os componentes de ordens inferiores. Em suma, a questão de interesse é descrever a relação entre a variável dependente e a variável independente, levando-se em conta possíveis efeitos de confusão e de modificação de efeitos provocados por variáveis de controle que, combinadas, formam as variáveis V e W do modelo EVW .

Eliminação dos termos EV_iV_j

Para se obter uma estatística válida da relação exposição-efeito que leve em conta efeitos de modificação e confusão, há que se iniciar com um modelo EVW . Como forma de considerar a interação das variáveis, usa-se uma eliminação reversa hierárquica de variáveis, tomando-se o cuidado de tratar primeiramente os termos de ordem superior.

O processo de eliminação reversa de variáveis demanda testes estatísticos para os termos de interação, de maneira a retirar termos pouco relevantes e preservar os demais. Esses testes devem ser aplicados em um único “pacote” para os termos de uma dada ordem.

Se o teste de pacote não resulta significativo, pode-se optar pela eliminação de todos os termos contidos no pacote. Por outro lado, caso o teste seja significativo, isso quer dizer que alguns, mas não necessariamente todos os termos do pacote, são significativos e que devem ser preservados.

3 - CONCEPÇÃO DO MODELO – PROCEDIMENTOS METODOLÓGICOS

Conforme foi mencionado anteriormente, há um consenso em torno do entendimento de que o conhecimento integrado dos processos de mudança de **uso** do solo e dos fatores que afetam tais processos possibilitará a elaboração de modelos preditivos confiáveis de médio e de longo prazo, em diferentes escalas. Entretanto, é também consenso que a compreensão absoluta desses processos – sintetizada por meio de técnicas de modelagem – depende de uma abordagem transdisciplinar, o que implica uma alta complexidade de concepção e de implementação.

Em função disso, esta tese procura tão somente abordar mudanças na **cobertura** do solo, por meio da modelagem de uma única externalidade do uso do solo amazônico – o desmatamento – dado que “não há meios de se estudar diretamente as mudanças no uso do solo” (BRIASSOULIS, 2000).

Para tanto, o modelo estatístico adotado neste trabalho de tese, com vistas a estudar essas mudanças na cobertura do solo, adota técnicas de análise multivariada – é um modelo de regressão logística. Por questões de conveniência, foi denominado de ‘Análise Multivariada Aplicada a Zoneamento (AMAZON)’.

Em função da ênfase dada às relações espaciais entre as células analisadas, esta concepção implica necessariamente sua implementação no âmbito de um sistema de informação geográfica.

Ocorre que, em geral, os trabalhos que envolvem modelos de regressão usam pacotes estatísticos – SAS, SPSS e STATA, entre outros – que já são preparados para receber os dados de entrada em formatos especificamente concebidos para esses aplicativos. Neles, as operações são feitas entre tabelas mais ou menos extensas – a depender do número de eventos e de variáveis.

Para o AMAZON, entretanto, as variáveis concebidas para a regressão logística teriam que ser geradas a partir de imagens digitais, cujo formato é bastante distinto daquele processado nesses pacotes estatísticos. Além do formato distinto, tais imagens possuem milhões de pixels que, ainda que sejam reamostrados, tornariam a adoção de um pacote estatístico algo inviável do ponto de vista computacional.

Via de regra, a literatura consultada apresentou modelos nos quais um dado número de variáveis (de quatro a doze) eram aplicados a um número de eventos estatísticos que não

ultrapassava duas dezenas de milhares. Assim, fenômenos espacialmente distribuídos em uma dada área de estudo e suas respectivas observações foram tabulados para serem analisados separadamente, em algum aplicativo estatístico. Em alguns casos verificados, por exemplo, boa parte da Amazônia Legal teve seus dados tabulados (ALENCAR *et.al.*, 2004), enquanto em outros, a mudança no uso do solo foi estudada para três estados inteiros – Rondônia, Mato Grosso e Pará (SILVA, 2006).

No AMAZON, onze variáveis (uma dependente e dez independentes, sendo uma de exposição, sete de controle e duas de interação) têm seus efeitos verificados em um número de cerca de 2,5 milhões de eventos (células de floresta *versus* células de desmatamento). Esse volume de dados torna seu tratamento impraticável, se conduzido pelos meios normalmente apresentados na literatura consultada. Em suma, ao invés de tratar uma amostra estatística, o AMAZON lida com todo o universo amostral considerado.

De maneira a permitir um melhor entendimento da formulação do AMAZON, a Figura 3.1 apresenta uma visão geral das etapas cumpridas em sua concepção.

As duas primeiras etapas – ‘Concepção inicial’ e ‘Organização dos dados de entrada’ – são tratadas neste Capítulo, enquanto as duas últimas – ‘Formulação do modelo’ e ‘Implementação do modelo’ – são objeto de discussão do Capítulo 4.

3.1 - COLETA DE DADOS

3.1.1 - Concepção inicial

Os modelos de mudanças no uso e na cobertura do solo tratam do “resultado de uma complexa rede de interações entre forças biofísicas e socioeconômicas sobre o tempo e o espaço”, que é consequência dessas mudanças. No caso da Amazônia brasileira, tal complexidade atinge patamares críticos, não só pelas características intrínsecas à região, como extensão territorial e dificuldades de acesso, entre outras, mas pela extrema dificuldade em se obter dados consistentes, quando a proposição da pesquisa envolve uma abordagem dos aspectos socioeconômicos.

Assim, a escala exerce um favor preponderante, na medida em que dela depende a escolha entre elaborar um modelo detalhado para uma microrregião ou comunidade, e conceber um modelo simples (mas não necessariamente fácil), que se dedique a analisar eventos que ocorram em ambientes mais amplos. Tomando por base essa premissa, buscou-se adotar a segunda postura.

A partir da idéia de que é possível obter uma indicação apriorística dos fatores que exercem influência nas mudanças na cobertura do solo, optou-se pela utilização de variáveis de ordem biofísica, cuja obtenção é bem menos complexa relativamente às variáveis socioeconômicas.

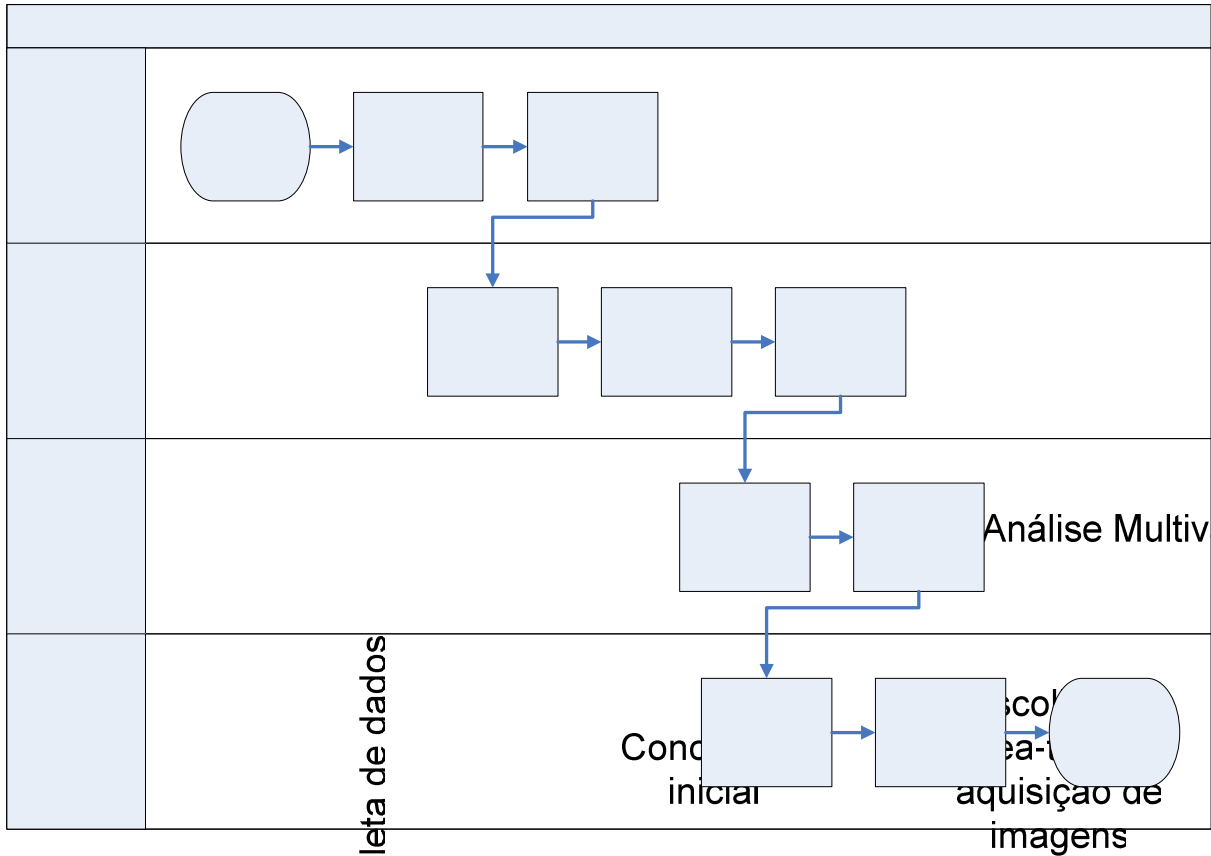


Fig. 3.1 – Esquema geral adotado na concepção do modelo AMAZON.

3.1.2 - Escolha da área-teste e aquisição de imagens

Tendo em vista a necessidade de analisar variáveis biofísicas diversas, capazes de influenciar ações de desmatamento na região amazônica, buscou-se identificar uma área na qual essa dinâmica estivesse sob atividade intensa no momento atual, ao mesmo tempo em que apresentasse um grau mínimo de desmatamento no início do período avaliado.

A “Terra do Meio”, denominação atribuída à região do estado do Pará situada “no meio” da faixa de terra que separa os rios Xingu e Araguaia, possui tais características, como de resto as demais regiões localizadas na vasta região que ficou conhecida por “Arco do Desmatamento”.

Embora a área-teste escolhida possua as características necessárias à condução deste trabalho de tese, ela, como de resto qualquer outra área da região amazônica, não é

Organização dos dados de entrada - SPRING e ARGIS -
 Formulação do modelo - ARGIS -

Georreferenciamento e registro de imagens

Realização do trabalho (so...

representativa para toda a Amazônia em termos de dinâmica de desmatamento. Não obstante, o AMAZON foi formulado de maneira tal a prevenir que as peculiaridades locais exerçam influência no desempenho do modelo, conforme as próximas seções buscam deixar claro.

Para que fosse possível estudar a evolução dessas variáveis de forma consistente e sistemática, seria necessária uma área grande o suficiente para abarcar as principais variáveis em todas as suas grandezas possíveis. A melhor relação custo-benefício para uma abordagem como esta foi encontrada no satélite Landsat 5. Em uma única cena, é possível cobrir uma área de 185km x 185km com um conjunto de resoluções que pode ser considerado excelente para aplicações florestais (pixel de 30m x 30m, sete bandas espectrais, revisita a cada 26 dias e vida útil superior a duas décadas).

Como forma de observar o fenômeno do desmatamento por um período de tempo longo o suficiente para inferir relações de dependência entre variáveis, foram selecionadas cenas do período que se estende de 1985 a 2004, via de regra espaçadas a cada quatro anos (exceção feita à primeira, separada em três anos da próxima cena).

3.1.3 - Realização do trabalho de campo

Como forma de facilitar as ações subsequentes de interpretação e de classificação das imagens, um extensivo trabalho de campo foi planejado.

Para tanto, a Força Aérea Brasileira destacou dois de seus esquadrões baseados em Belém-PA – 1º Esquadrão de Transporte Aéreo (1º ETA) e 1º Esquadrão do 8º Grupo de Aviação (1º/8º GAV) – para apoiarem a condução do trabalho.

Como não há, em São Félix do Xingu, instalações de apoio para a operação rotineira das aeronaves da FAB, o 1º ETA dedicou uma aeronave de transporte C-95 “Bandeirante” para trasladar o combustível necessário à operação deslocada do 1º/8º GAV. Este, por sua vez, manteve uma tripulação completa do helicóptero H-1H “Iroquois”, na duração da missão que lhe foi atribuída.

O esforço aéreo despendido nas missões de helicóptero foi da ordem de 15 horas de voo, distribuídas entre o deslocamento (ida e volta de Belém a S. F. do Xingu) e a missão de sobrevôo propriamente dita, cujos trajetos podem ser identificados no plano de missão superposto a um mosaico de cenas Landsat, apresentado na Figura 3.2. Ambos os circuitos foram conduzidos no sentido horário, porém com mudança de visada. Ora o circuito era imageado da porta lateral direita do helicóptero, ora da esquerda, de maneira a cobrir áreas maiores e diversas.

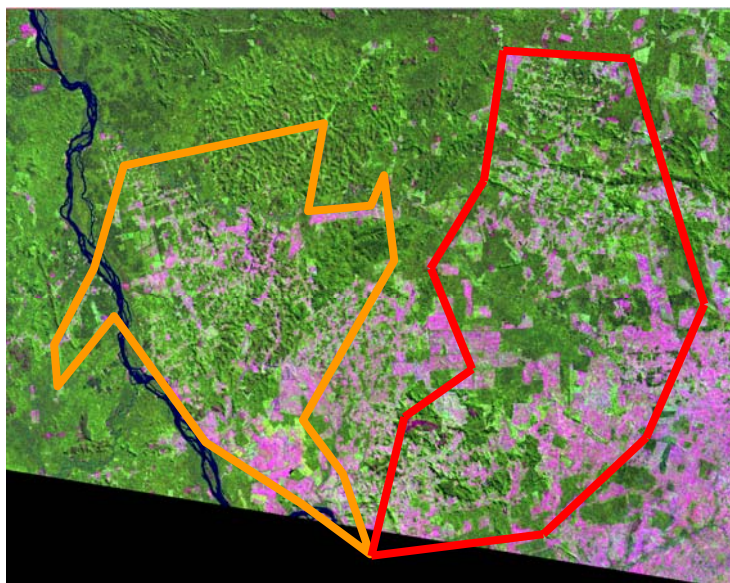


Fig. 3.2 – Rotas percorridas pela aeronave FAB 8530 do 1º ETA, a partir do aeródromo de São Félix do Xingu.

3.2 - ORGANIZAÇÃO DOS DADOS DE ENTRADA

A área-teste escolhida para o estudo, ilustrada pelas Figuras 3.3 e 3.4, é delimitada pela “órbita/ponto” 225/64 do satélite Landsat 5 (cena de 185km x 185km), localizada no município de São Félix do Xingu, região Sul do estado do Pará. Trata-se de uma área situada no “Arco do Desmatamento”, onde se verifica um intenso processo de ocupação da floresta nativa para atividades agropecuárias.

Antes que o modelo possa ser formulado e testado, os dados que irão compô-lo precisam ser adequadamente preparados, em função da etapa de modelagem considerada. O AMAZON leva em conta parâmetros que permitem verificar o relacionamento da variável dependente com as variáveis independentes, de exposição e de controle.

As operações que permitem conhecer tal relacionamento são conduzidas em um ambiente computacional conhecido como sistema de informação geográfica. Para este estudo, foi adotado o aplicativo ArcGIS 9.2.

Os parâmetros mencionados acima são extraídos de planos de informação (PI) que compõem o projeto sobre o qual o modelo é processado. Parte desses PI advêm de dados vetoriais colhidos junto a fontes secundárias, como IBGE e Imazon, por exemplo, e são alimentados diretamente no ArcGIS.

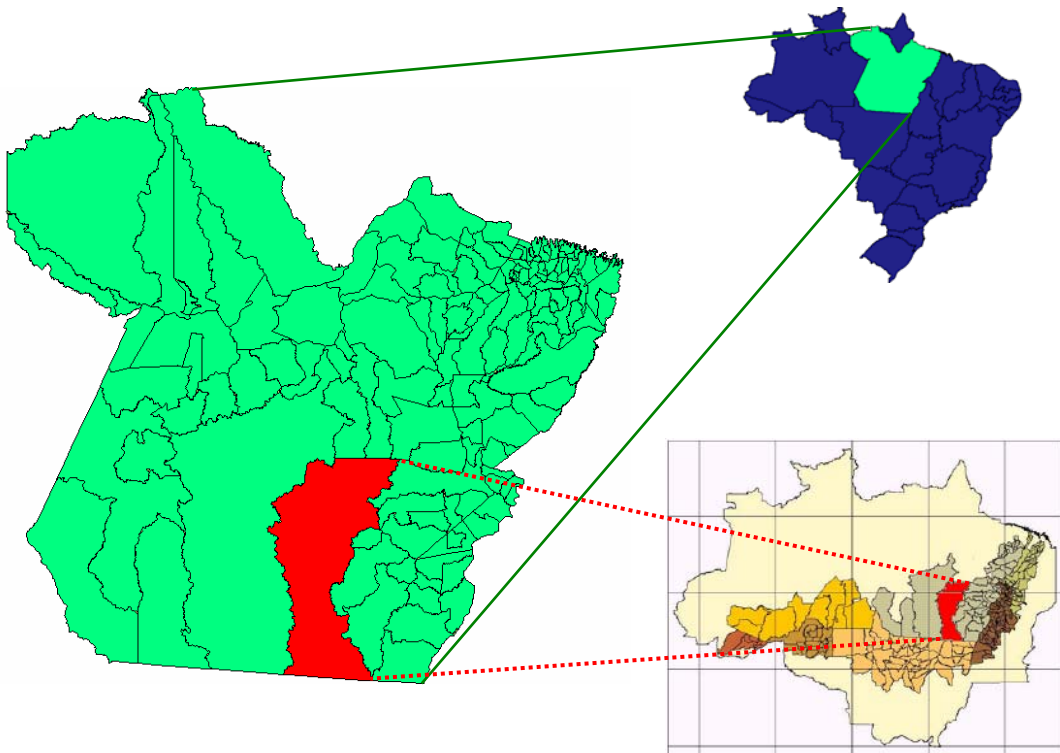


Fig. 3.3 – A área de estudo em relação ao estado do Pará e ao Arco do Desmatamento.

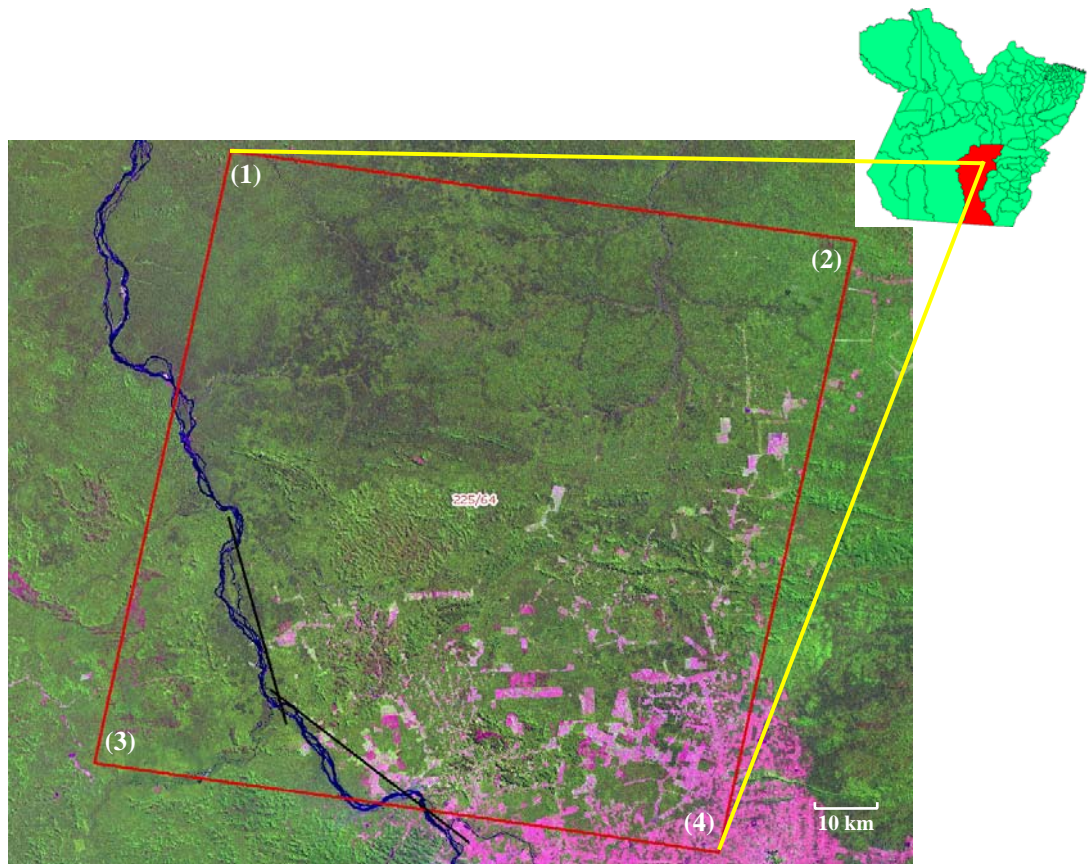


Fig. 3.4 – Área – teste, formada pelas coordenadas (1) $052^{\circ}40'W/05^{\circ}49'S$, (2) $052^{\circ}06'W/05^{\circ}53'S$, (3) $052^{\circ}46'W/06^{\circ}22'S$, e (4) $052^{\circ}12'W/06^{\circ}27'S$.

Outros PI têm que ser preparados a partir de um tratamento específico dado às imagens orbitais, responsáveis por retratar as mudanças biofísicas que ocorrem na área de estudo e que são o alvo desta tese. Essa preparação foi conduzida no aplicativo SPRING, do INPE.

Por se tratar de um estudo que envolve, antes de mais nada, mudanças na cobertura do solo, analisou-se um período de aproximadamente duas décadas, de 1985 a 2004 – praticamente a vida útil do satélite de observação terrestre Landsat 5, lançado em março de 1984.

Esse período foi amostrado, a partir da cena de 1988, a cada quatro anos, totalizando seis cenas: 1985 (10 de junho), 1988 (18 de junho), 1992 (29 de junho), 1996 (8 de junho), 2000 (30 de junho) e 2004 (26 de maio). A fim de minimizar efeitos típicos de sazonalidade, foram selecionadas cenas da estação seca, que apresentaram boa qualidade (ausência de cobertura de nuvens). O intervalo de quatro anos entre as cenas se deve à combinação de uma amostragem que permite a montagem de uma série temporal suficientemente extensa, com uma quantidade de dados razoavelmente reduzida, passível de ser tratada em tempo hábil.

A preparação das imagens no ambiente SPRING, bem como a subsequente geração dos PI no ArcGIS são descritas neste Capítulo.

3.2.1 - Georreferenciamento e registro de imagens

Tendo em vista o fato de que as imagens Landsat, quando de sua aquisição, estão geometricamente corrigidas, o georreferenciamento das imagens é feito registrando-se a imagem de satélite a uma carta topográfica digitalizada, por meio da aquisição de alguns pontos de controle.

No registro, associam-se as coordenadas da imagem (linha, coluna) com as coordenadas geográficas (latitude, longitude) das cartas topográficas, para o ano-base (1985), sendo as imagens dos anos subsequentes registradas entre si na forma imagem x imagem, de maneira a coincidirem espacialmente, conforme ilustra a Figura 3.5.

Em seguida, atribui-se ao conjunto uma projeção cartográfica e seu respectivo *datum* (UTM 22, SAD 69) para permitir as operações estatísticas exigidas pelo modelo em um ambiente SIG.

3.2.2 - Classificação temática

Considerando os parâmetros necessários à geração do modelo, as imagens foram classificadas em ‘floresta’, ‘não-floresta’, ‘desmatamento’ e ‘hidrografia’.

São inúmeros os procedimentos metodológicos existentes na literatura concebidos para possibilitar classificações temáticas mais precisas de uso e de cobertura do solo. De maneira geral, o estabelecimento desses procedimentos é dependente de dois fatores principais: o sistema sensor utilizado na coleta dos dados e a natureza do alvo classificado. Por exemplo, imagens de radar conduzem a procedimentos metodológicos bastante distintos daqueles desenvolvidos para imagens ópticas. O mesmo pode ser dito em relação à classificação de áreas urbanas e áreas florestais.

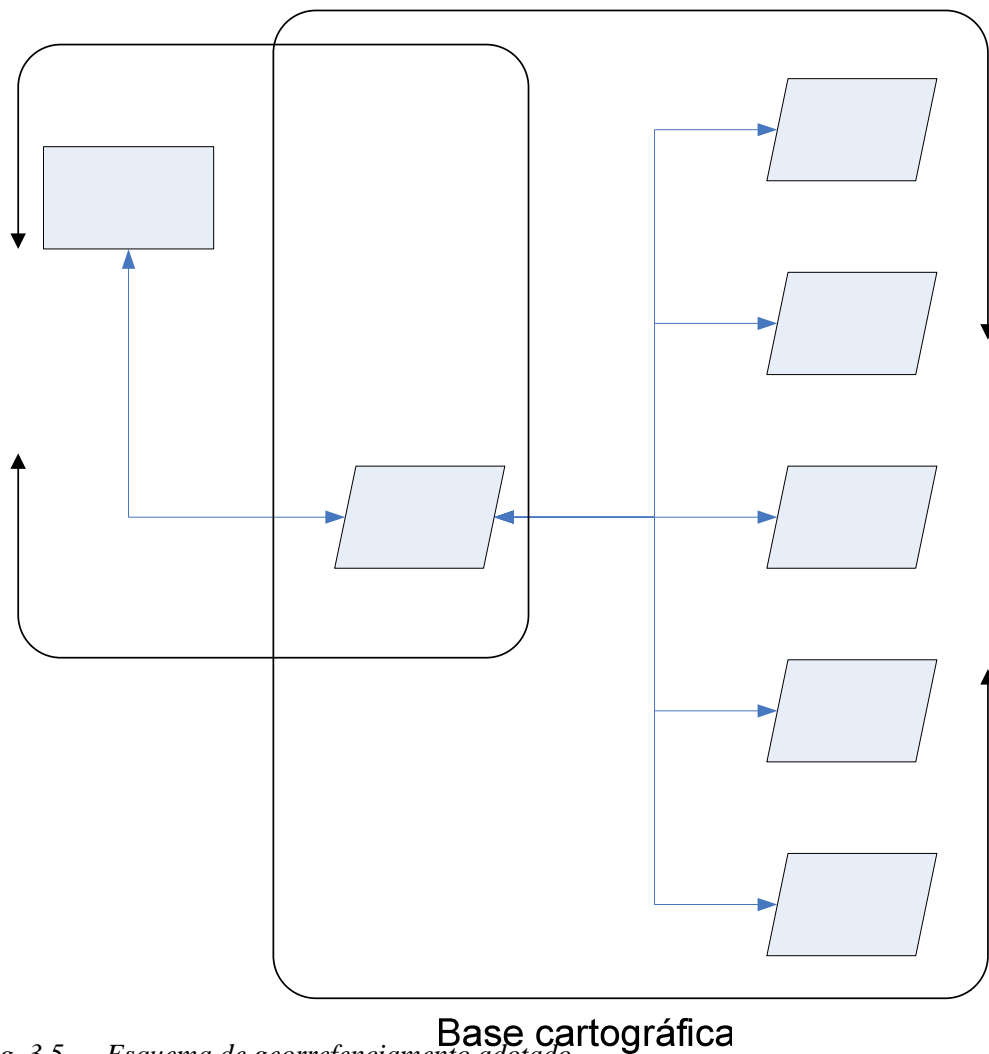


Fig. 3.5 – Esquema de georreferenciamento adotado.

Esses fatores ainda possuem características subordinadas, por assim dizer. No caso dos sistemas sensores, cabem destacar a resolução e abrangência espectral, resolução espacial e polarização, por exemplo. No caso dos alvos, há que se considerar sua extensão, existência ou não de sazonalidade, as classes de uso e de cobertura do solo nas quais se tem interesse, entre outras coisas. Uma vez combinadas, essas características geram possibilidades virtualmente irrestritas de procedimentos de classificação.

No caso deste estudo, considerando o fato de que a classificação, ainda que importante, não passa de uma etapa intermediária para a formulação do modelo a ser testado, buscou-se um procedimento metodológico consagrado para o binômio ‘sensor óptico multiespectral x área de floresta úmida’.

Assim sendo, foi escolhido o procedimento desenvolvido pelo INPE para o PRODES DIGITAL (SHIMABUKURO *et. al.*, 1998). Trata-se de um método operacionalmente consagrado para a classificação do desmatamento na região amazônica, composto de cinco etapas, a saber: (i) geração das imagens-fração; (ii) segmentação das imagens-fração; (iii) geração do arquivo de contexto e extração de regiões; (iv) classificação da imagem segmentada; e (v) mapeamento da imagem segmentada.

i) Geração das imagens-fração

Com o objetivo de estimar a proporção de solo, vegetação e sombra para cada pixel da imagem, a partir da resposta espectral das bandas 3 (Vermelho \Rightarrow 0,63 – 0,69 μm), 4 (Infravermelho Próximo \Rightarrow 0,76 – 0,90 μm) e 5 (Infravermelho Médio \Rightarrow 1,55 – 1,75 μm), do Landsat-TM, utilizou-se o modelo linear de mistura espectral de Shimabukuro e Smith (1991).

O modelo de mistura espectral pode ser escrito como:

$$r_i = a * veg_i + b * solo_i + c * sombra_i + e_i, \quad (3.1)$$

onde r_i é a resposta do pixel na banda i da imagem Landsat-TM; ‘ a ’, ‘ b ’ e ‘ c ’ são proporções de vegetação, solo e sombra (ou água) que compõem o pixel; veg_i , $solo_i$ e $sombra_i$ correspondem às respostas espectrais de cada uma dessas componentes citadas para cada banda i ; e e_i é o erro de estimação intrínseco para cada banda i .

Com isso, são geradas as imagens-fração correspondentes a solo, vegetação e sombra. Dessas três, usa-se a fração-sombra para a classificação propriamente dita, pois ela proporciona um maior contraste entre áreas florestadas (média quantidade de sombras) e áreas desflorestadas (baixa quantidade de sombras), bem como provê uma boa separabilidade de temas tais como hidrografia, queimadas e áreas de capoeira.

A imagen-fração ‘solo’, por sua vez, mostra-se útil para realçar as áreas recém-cortadas e para permitir a análise das áreas ocupadas com cerrado. Por fim, a imagen-fração ‘vegetação’ realça as áreas de cobertura vegetal.

A Figura 3.6 apresenta as três imagens-fração geradas para o ano de 1996. Esta cena foi escolhida para ilustrar os passos de preparação dos dados usados na formulação do modelo por se tratar de uma data intermediária no estudo e para a qual a dinâmica de desmatamento já se encontrava em estágio razoavelmente avançado, com uma clara distribuição temática.

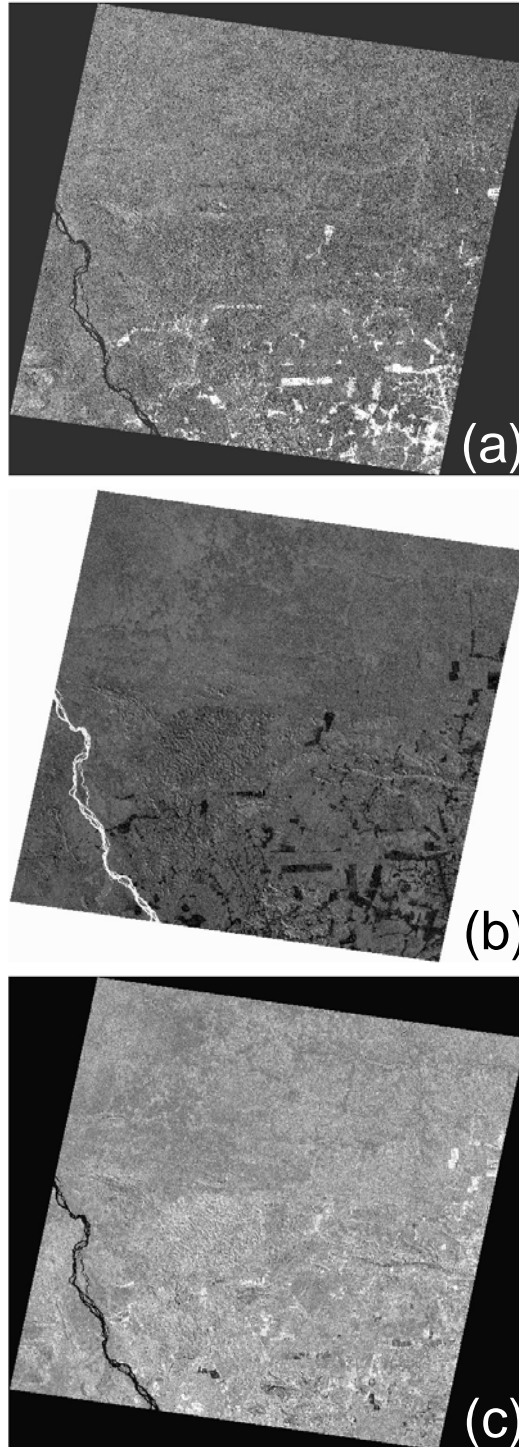


Fig. 3.6 – Imagens-fração de solo (a), sombra (b) e vegetação (c) geradas a partir do modelo linear de mistura espectral para o ano de 1996.

ii) Segmentação das imagens-fração

De acordo com Crepani *et al.* (2001), o procedimento usado para a segmentação de imagens é baseado no algoritmo de “crescimento de região”, onde um conjunto de pixels homogêneos é ligado de acordo com suas propriedades. Uma descrição detalhada do processo de segmentação pode ser encontrada em Batista *et al.* (1994).

A segmentação exige o estabelecimento de dois parâmetros. O primeiro deles é o limiar de similaridade, definido pela diferença mínima de níveis de cinza, abaixo da qual dois segmentos são considerados similares e, portanto, agrupados em uma única região. O segundo parâmetro é o limiar de área, definido pela área mínima a constituir uma região e dado em número de pixels.

Neste trabalho, foram estabelecidos os limiares ‘8’ e ‘16’ para similaridade e área, respectivamente, por serem esses os valores mais utilizados no PRODES DIGITAL. Um exemplo do resultado obtido pode ser visto na Figura 3.7.

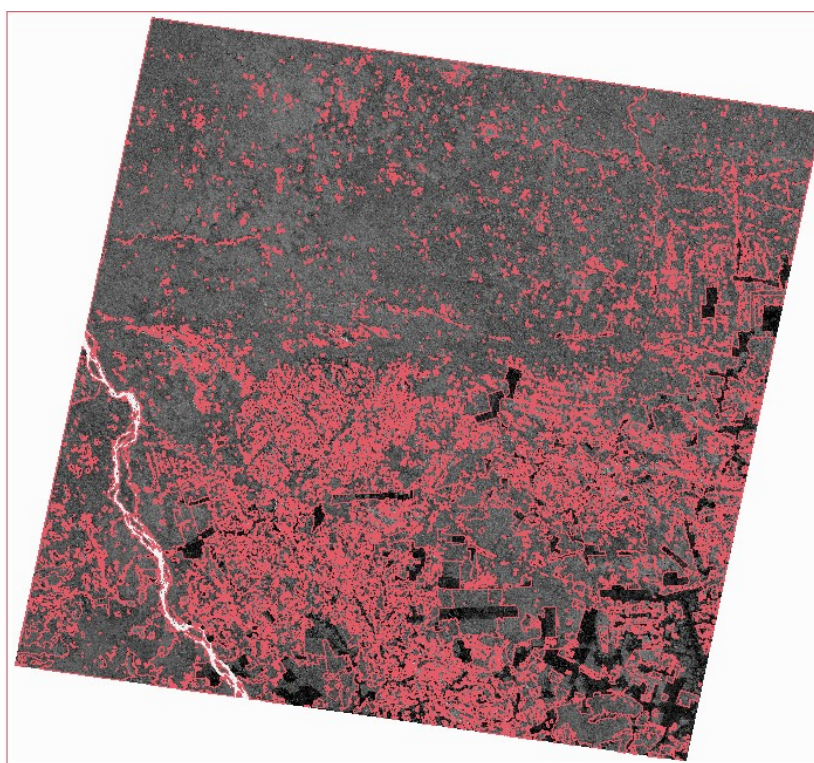


Fig. 3.7 – Imagem-fração de 1996 (sombra) segmentada pelo algoritmo de crescimento de região.

iii) Geração do arquivo de contexto e extração de regiões

Um arquivo de contexto deve ser gerado com vistas à classificação. Nele, constam as bandas ou imagens utilizadas, o tipo de classificação por região e a imagem segmentada. A seguir, são extraídos os atributos estatísticos (médias e matrizes de covariância) do conjunto de regiões definido pela segmentação.

iv) *Classificação da imagem segmentada*

Seguindo os procedimentos metodológicos propostos por Shimabukuro *et al.* (1998), uma classificação não-supervisionada baseada em um algoritmo de agrupamento (*clustering*) foi aplicada nas imagens-fração segmentadas.

A classificação não-supervisionada por isodados calcula a média da classe uniformemente distribuída no espaço de dados para, então, agrupar iterativamente os pixels restantes, por meio de técnicas de distância mínima. Cada iteração recalcula a média e reclassifica os pixels em relação às novas médias.

A divisão, fusão e exclusão iterativa de classes é feita com base nos parâmetros dos limiares utilizados. Todos os pixels são dispostos na classe mais próxima, a menos que um desvio-padrão ou limiar de distância seja especificado, situação na qual alguns pixels podem não ser classificados se não atenderem os critérios selecionados. Esse processo continua até que o número de pixels em cada classe mude a uma razão menor que o limiar de mudança de pixel selecionado ou que o número máximo de iterações seja alcançado. Informações técnicas mais detalhadas podem ser encontradas em Tou e Gonzalez (1974).

A Figura 3.8 apresenta a classificação obtida para a cena de 1996.

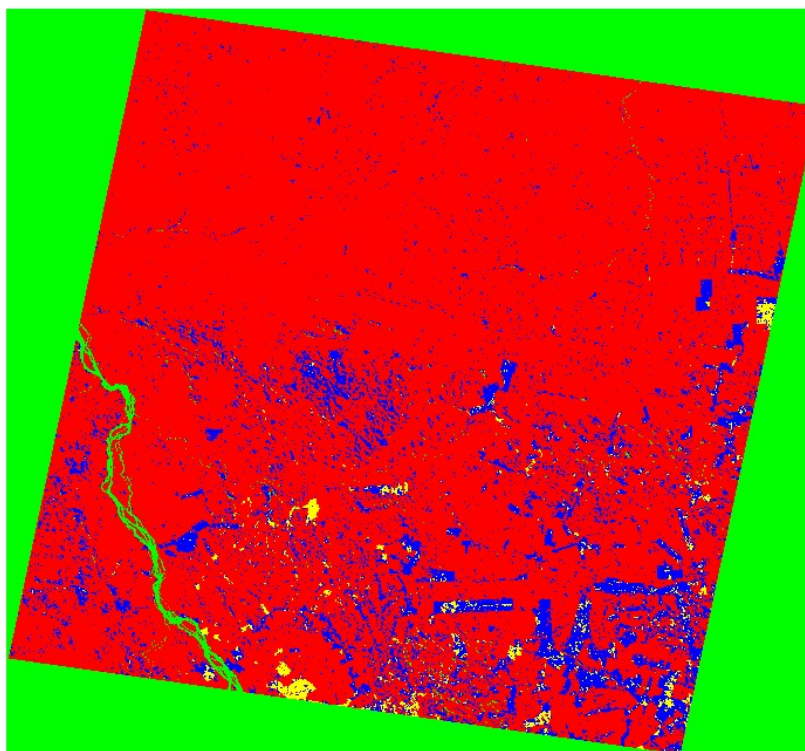


Fig. 3.8 – Imagem classificada (1996) com base em algoritmo de agrupamento.

v) *Mapeamento da imagem segmentada*

Uma vez classificada a imagem segmentada, os temas obtidos pelo classificador são associados às classes pré-definidas para o modelo a ser formulado, conforme ilustra a Figura 3.9.

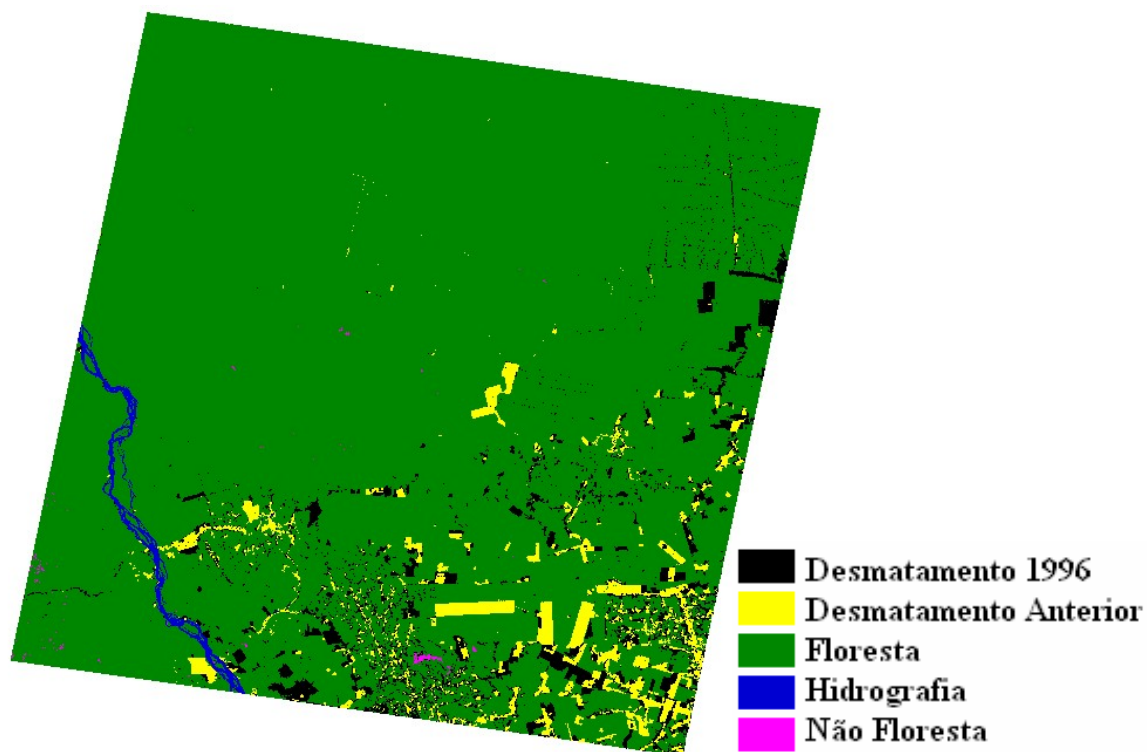


Fig. 3.9 – Mapeamento final para o ano de 1996.

A tarefa de georreferenciamento e registro de imagens, bem como a de classificação temática dessas imagens, encerrou a etapa conduzida no ambiente SPRING. Os arquivos ali gerados foram exportados, em formato vetorial (polígonos, linhas e pontos), para o ArcGIS, onde foram manipulados, em formato *raster* (grade numérica), com vistas à formulação do modelo concebido para o estudo.

3.2.3 - Escolha e geração das variáveis

A concepção de um modelo relativamente simples como o AMAZON pressupõe um número limitado de variáveis, dado o volume computacional a ser gerado no processo de regressão. Não se deve perder de vista o fato de que uma cena Landsat possui um número superior a 40 milhões de pixels, que devem ser analisados para cada uma das variáveis escolhidas e para cada uma das datas consideradas.

Essa quantidade de dados é de tal monta que não foi possível computá-los todos em sua máxima resolução. Assim, esse número foi reduzido em 16 vezes, por meio do agrupamento

dos pixels em células de formato 4 x 4, manobra que tornou possível executar a regressão logística do AMAZON no ambiente computacional empregado³⁸.

Além da variável dependente ‘impacto ambiental’ IPT, que é binária (0,1), oito variáveis independentes – PAD, ROD, HID, EPA, DEC, ALT, IFR e APE, que são detalhadas a seguir – foram inicialmente criadas com base na percepção de quais fatores têm o potencial de exercer influência na dinâmica do desmatamento e, naturalmente, tendo em vista a literatura consultada.

A partir dessas variáveis independentes, outras duas foram geradas a partir da interação multiplicativa entre a variável independente de exposição e duas variáveis independentes de controle, resultando nas variáveis descritas a seguir.

Cada variável foi traduzida em um PI (plano de informação ou *layer*), cujos valores são dispostos célula-a-célula, de maneira a permitir operações algébricas entre os diversos PI.

i) Variável de exposição PAD – proximidade da área desmatada

Essa variável é a essência da hipótese lançada no início deste trabalho de tese, na medida em que, dos fatores de ordem biofísica, PAD é a variável que mais influência exerce sobre a dinâmica do desmatamento, uma vez que o avanço dessa frente pressupõe sua proximidade com as áreas de floresta.

A obtenção da PAD foi feita por meio da interrogação, célula-a-célula, sobre a distância dessa célula para o polígono identificado com a classe de desmatamento (distância euclidiana), conforme ilustra o diagrama de fluxo de dados da Figura 3.10.

Uma vez obtidos os valores de PAD para cada célula, essa variável foi categorizada em faixas de 1 km. Isso porque, uma vez que essa variável fosse mantida em seu formato contínuo original, de metro em metro, não haveria como estabelecer a contribuição incremental da variável independente PAD na explicação da variável dependente IPT, pois tal contribuição seria próxima de ‘0’.

³⁸ - O computador usado possui a seguinte configuração: *Pentium 4* de 3.4 GHz, 2 GHz de RAM e 80 GB de espaço em disco.

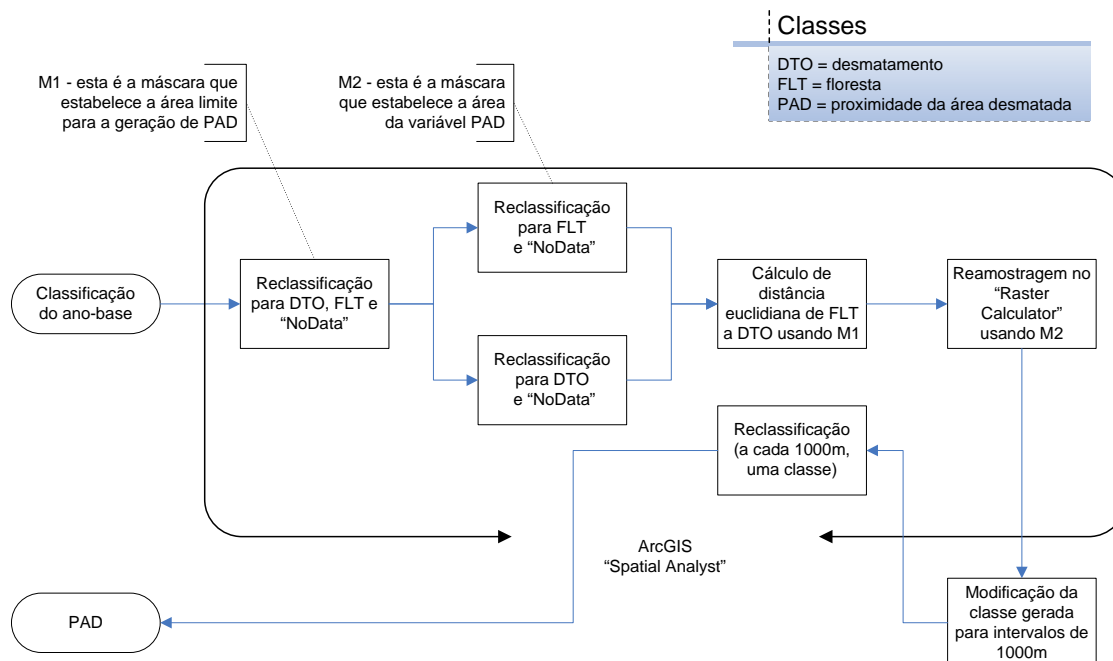


Fig. 3.10 – Esquema geral para obtenção da variável PAD.

ii) Variável de controle ROD – proximidade de rodovias

A proximidade de rodovias possui uma ligação quase tão estreita com o desmatamento quanto a proximidade deste com a área desmatada. A literatura consultada estabelece uma forte dependência entre a construção de rodovias e a dinâmica de desmatamento na Amazônia, em alguns casos fixando um “ponto de corte” de 10 km em torno das rodovias, em outros de 50 km, a partir do qual o desmatamento é minimizado.

No caso do AMAZON, contudo, o que faz a variável ROD exercer uma influência menor no exemplo estudado é o fato de ser ela uma variável obtida a partir de dados secundários. Enquanto PAD é extraída diretamente da interpretação da imagem Landsat, ROD é proveniente dos dados disponibilizados por terceiros – neste caso o Imazon. Como o estudo foi espaçado em intervalos de quatro anos, são grandes as chances de que os vetores que representam as rodovias não coincidam minimamente com as datas de coletas das imagens a elas correspondentes.

Para sua obtenção, um procedimento similar àquele seguido para a geração da variável PAD foi adotado, tendo sido o resultado final categorizado também em distâncias de 1 km entre as faixas subsequentes, a partir da rodovia considerada. A Figura 3.11 ilustra as etapas seguidas para a obtenção dessa variável.

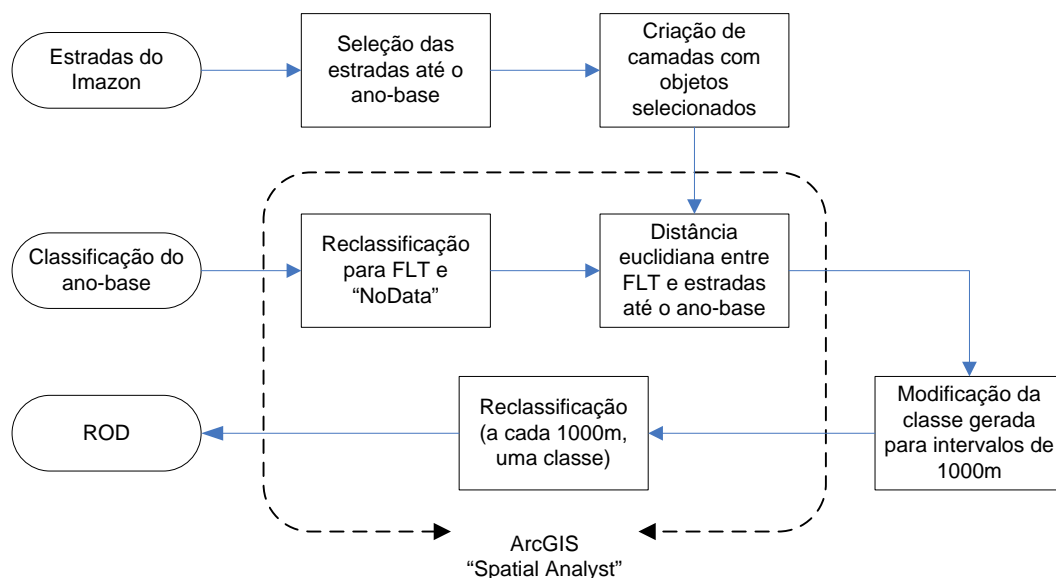


Fig. 3.11 – Esquema geral para obtenção da variável ROD.

iii) Variável de controle HID – proximidade de hidrovias

As hidrovias constituem um eficiente canal de distribuição de madeira e sua proximidade facilita sobremaneira o escoamento da produção madeireira amazônica. A obtenção desta variável, conforme ilustrada pela Figura 3.12, segue praticamente os mesmos passos adotados para a obtenção de ROD.

Entretanto, a quase inexistência de rios na área investigada fez com que esta variável fosse considerada nula para o AMAZON aqui implementado, uma vez que, para a cena Landsat 225/64 isoladamente, a utilização da variável HID agrega pouco ou nenhum valor ao modelo.

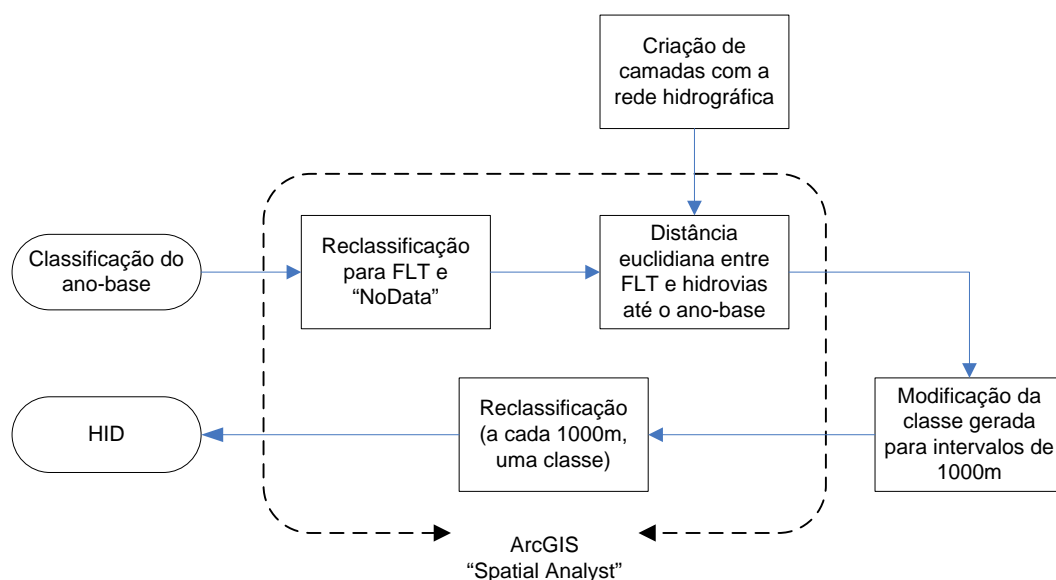


Fig. 3.12 – Esquema geral para obtenção da variável HID.

iv) Variável de controle EPA – existência de proteção ambiental

A variável EPA, cujas etapas de obtenção são apresentadas na Figura 3.13, é uma variável binária (0,1), gerada a partir do mapeamento de áreas protegidas, sejam elas unidades de conservação (UC) – PARNA da Serra do Pardo e FLONA do Itacaí – sejam terras indígenas (Trincheira/Bacaj e Arawet/Ipixuna). Os dados foram obtidos junto ao Imazon (2005), por ocasião do trabalho de campo.

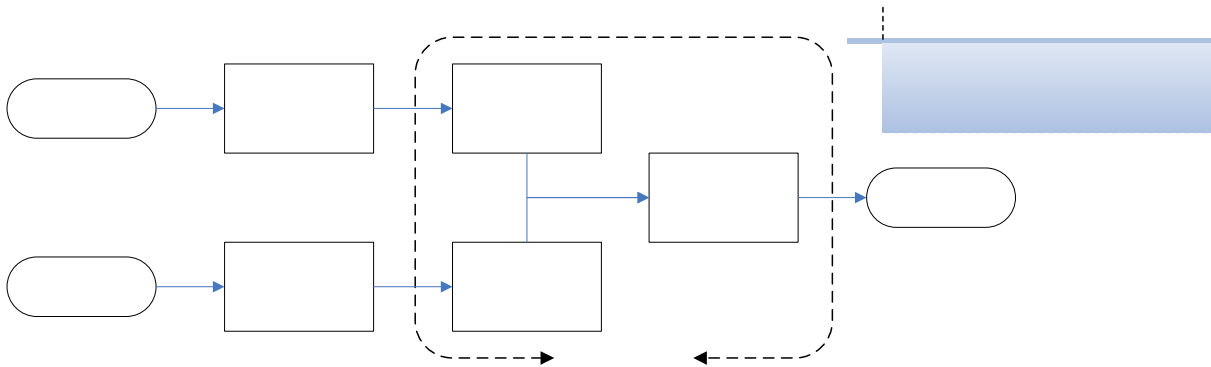


Fig. 3.13 – Esquema geral para obtenção da variável EPA.

Ainda que não possam ser consideradas “fatores de ordem biofísica”, essas áreas protegidas foram agregadas ao modelo, por conta da facilidade de obtenção desse tipo de dado, princípio que sempre norteou o trabalho. Isso foi feito à medida em que as UC foram criadas, o que ocorreu a partir de 1996. No ano de 2004, embora constasse a demarcação da terra indígena Apyterre, a área correspondente na imagem utilizada pelo AMAZON ainda não havia sido homologada naquela data, o que implicou a sua não utilização na composição da variável EPA de 2004.

v) Variável de controle DEC – declividade

Presume-se que a atividade de desmatamento seja mais favorável em relevos mais suaves do que o contrário, dada a maior facilidade de remoção das imensas toras de madeira em terrenos planos. Assim, a partir do modelo de elevação SRTM (Shuttle Radar Topography Mission), disponibilizado na Internet pela NASA, foi possível gerar o mapa de declividade da área de estudo (ferramenta *slope* do ArcGIS 9.2), com uma resolução espacial de 90m na dimensão horizontal e de aproximadamente 16m na vertical.

A classificação para todas as cenas foi feita com um espaçamento equidistante de 5° entre a superfície mais plana e a mais íngreme. As etapas seguidas para a obtenção de DEC,

bem como da variável ALT, detalhada no próximo item, são ilustradas na Figura 3.14. A espacialização de DEC, igual para todas as datas, é apresentada a seguir, na Figura 3.15.

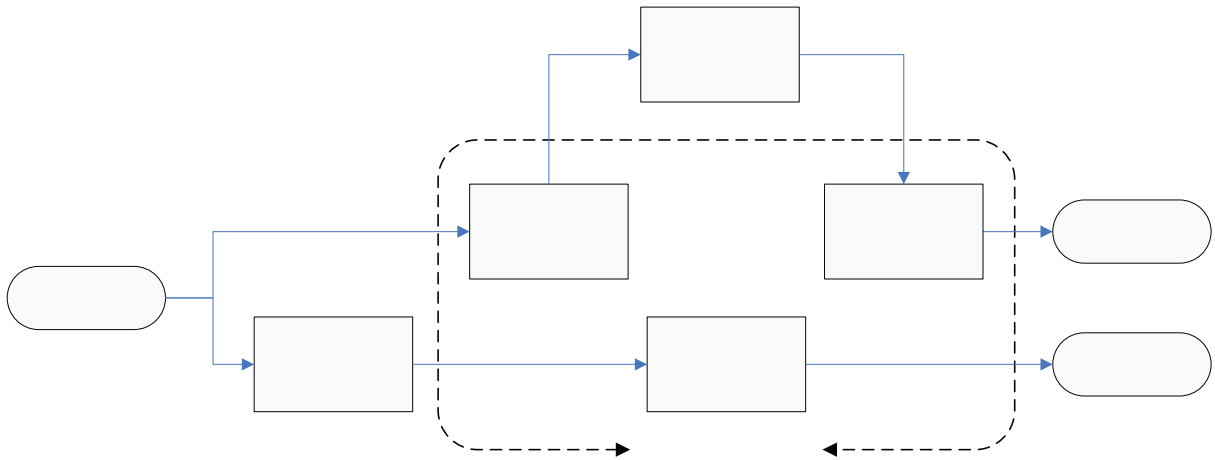
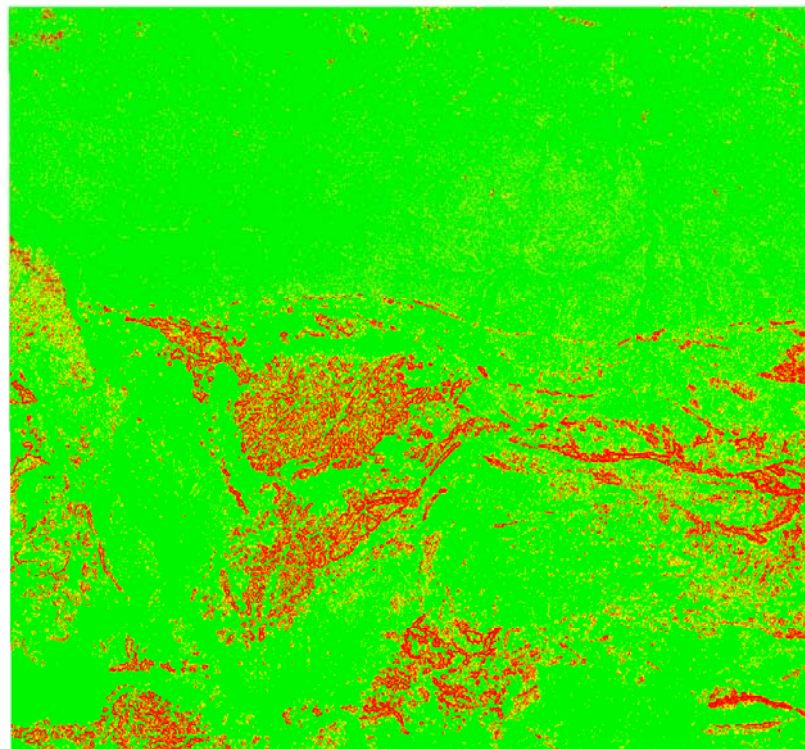


Fig. 3.14 – Esquema geral para obtenção das variáveis DEC e ALT.



Geração
declividade
contínua em



Fig. 3.15 – Espacialização da variável DEC no entorno da área de estudo.

vi) *Variável de controle ALT – altitude*

De maneira similar à variável DEC, também com ALT se espera um comportamento da variável dependente IPT em relação à condição do terreno, neste caso em termos de altura. Presume-se que, quanto mais alto for o terreno, mais árduo se torna o manejo florestal, especialmente em locais onde as vias são de difícil acesso, como é o caso daquela região.

Assim, o modelo SRTM foi “fatiado” em perfis verticais separados 16m entre si, a mesma resolução do modelo. Também como no caso da variável DEC, a variável ALT permanece inalterada em todas as datas. Sua distribuição espacial, também igual para todas as datas, é apresentada pela Figura 3.16.

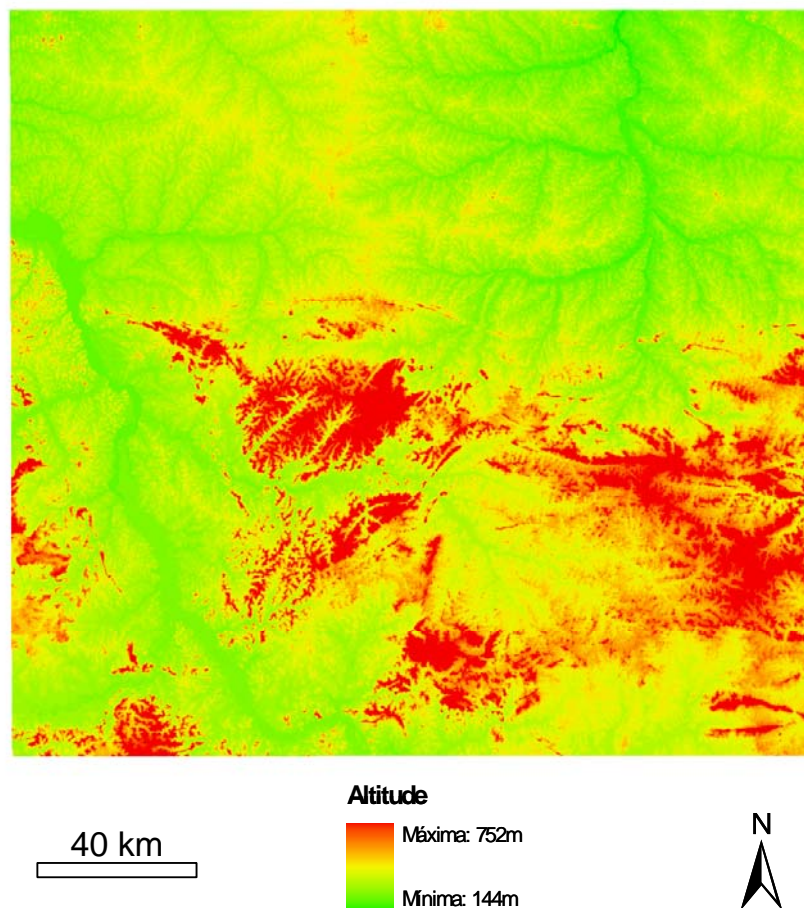


Fig. 3.16 – Espacialização da variável ALT no entorno da área de estudo.

vii) *Variável de controle IFR – índice de fragmentação*

É de se esperar que a forma do polígono no qual se inserem as células de floresta desempenhe um papel de peso na dinâmica do desmatamento. Quanto mais alongado for esse

polígono, mais frágil estará a floresta nele contida. A recíproca é verdadeira e se faz representar por uma forma circular, na qual se alcança o menor perímetro relativamente à área do polígono.

Algebricamente, pode-se obter um indicador da fragmentação para uma célula florestal por meio da seguinte equação:

$$IFR = \frac{4\pi A}{P^2} \quad (3.2)$$

onde A é a área do polígono e P é o seu perímetro.

Quanto mais próximo de '1' for o valor desse indicador, mais próximo da forma circular estará o polígono. Por outro lado, quanto mais alongada for a forma do polígono, mais próximo de '0' será o indicador.

Como é mais intuitivo associar valores crescentes para o índice de fragmentação com estados mais avançados de fragmentação florestal, é feita uma conversão desse índice para valores percentuais, sendo o valor resultante subtraído de '100'. Assim, uma área perfeitamente circular apresenta um IFR igual a 0%. Uma vez gerados os índices, esses são classificados em intervalos naturais³⁹ e remapeados, de maneira a permitir sua operação com outras variáveis.

Ainda que existam na literatura outras maneiras, que não a simples consideração da forma para se estabelecer índices de fragmentação, a maneira acima descrita foi escolhida por sua objetividade de implementação.

O resultado da espacialização da variável IFR para o ano-base de 2000 pode ser visualizado na Figura 3.17, enquanto a Figura 3.18 ilustra as etapas seguidas tanto na formulação do IFR como da variável APE, detalhada na seqüência.

³⁹ - O método de classificação por "intervalos naturais" particiona os dados em classes, por meio de um algoritmo, comumente denominado de 'Otimização Jenks'. O algoritmo calcula agrupamentos de valores dos dados e, eventualmente, um número de classes baseado no menor erro total possível – a soma de desvios absolutos em torno da mediana da classe ou, alternativamente, a soma do quadrado dos desvios em torno da média. (ESRI, 2004)

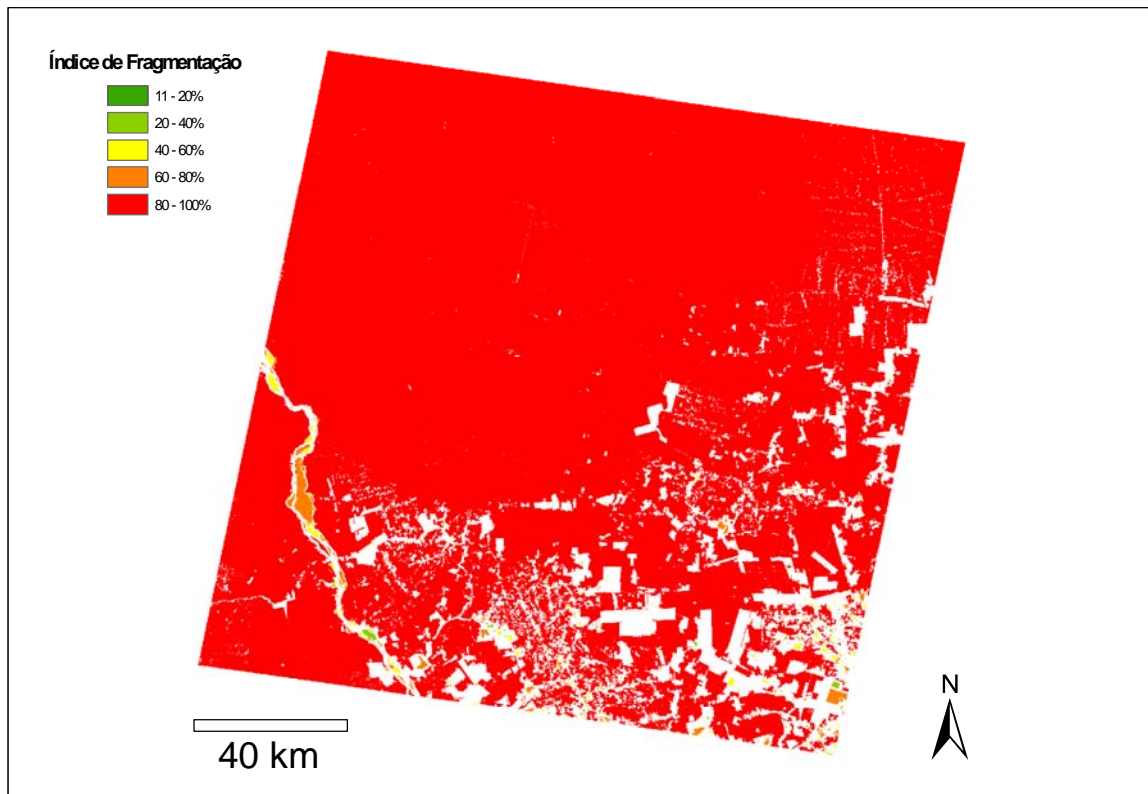


Fig. 3.17 – Espacialização da variável IFR para o ano-base de 2000.

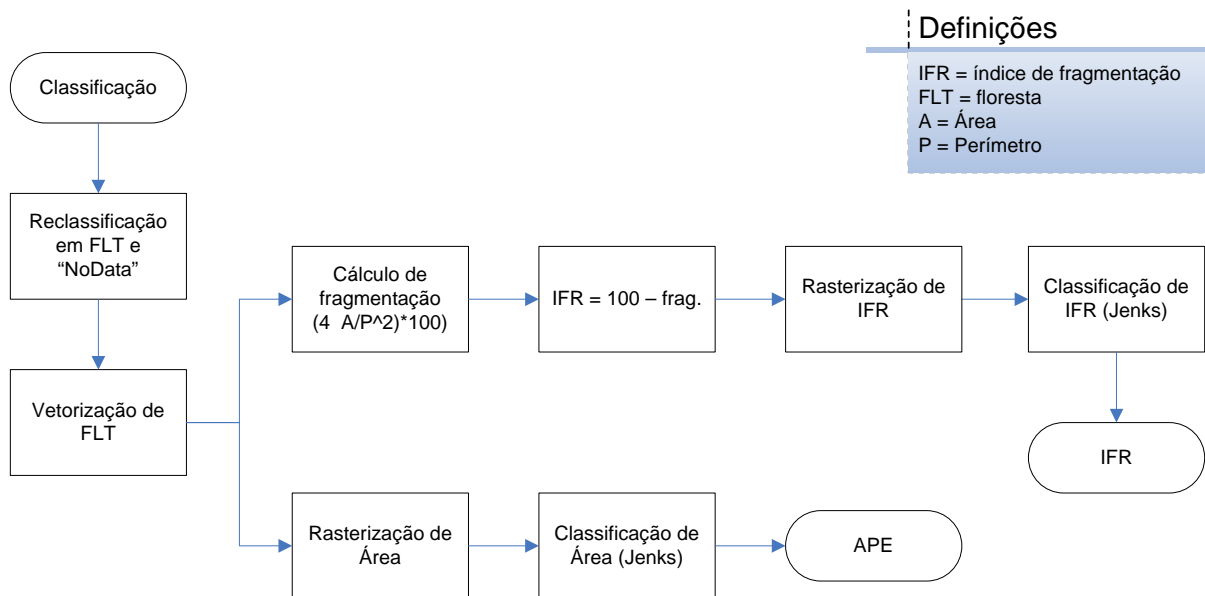


Fig. 3.18 – Esquema geral para obtenção das variáveis IFR e APE.

viii) Variável de controle APE – área do polígono envolvente

A variável APE relaciona o tamanho da célula à sua capacidade de preservação, na medida em que uma pequena área florestal certamente possui um maior risco de

desmatamento do que uma área de maior porte. A Figura 3.19 apresenta a variável APE referente ao ano-base de 2000.

ix) *Variáveis de interação $W_1 = PAD \times DEC$ e $W_2 = PAD \times ALT$*

Variáveis multiplicativas não possibilitam interpretações intuitivas, dado que seus componentes podem ter unidades completamente distintas. Entretanto, elas são usadas em função do potencial que têm em evidenciar as contribuições dos seus componentes isolados, quando associados entre si.

Sua classificação é feita de maneira equitativa e os valores resultantes da multiplicação são adimensionais. Os resultados para o ano-base de 2000 são apresentados na Figura 3.20, enquanto a Figura 3.21 apresenta as etapas na construção dessas variáveis.

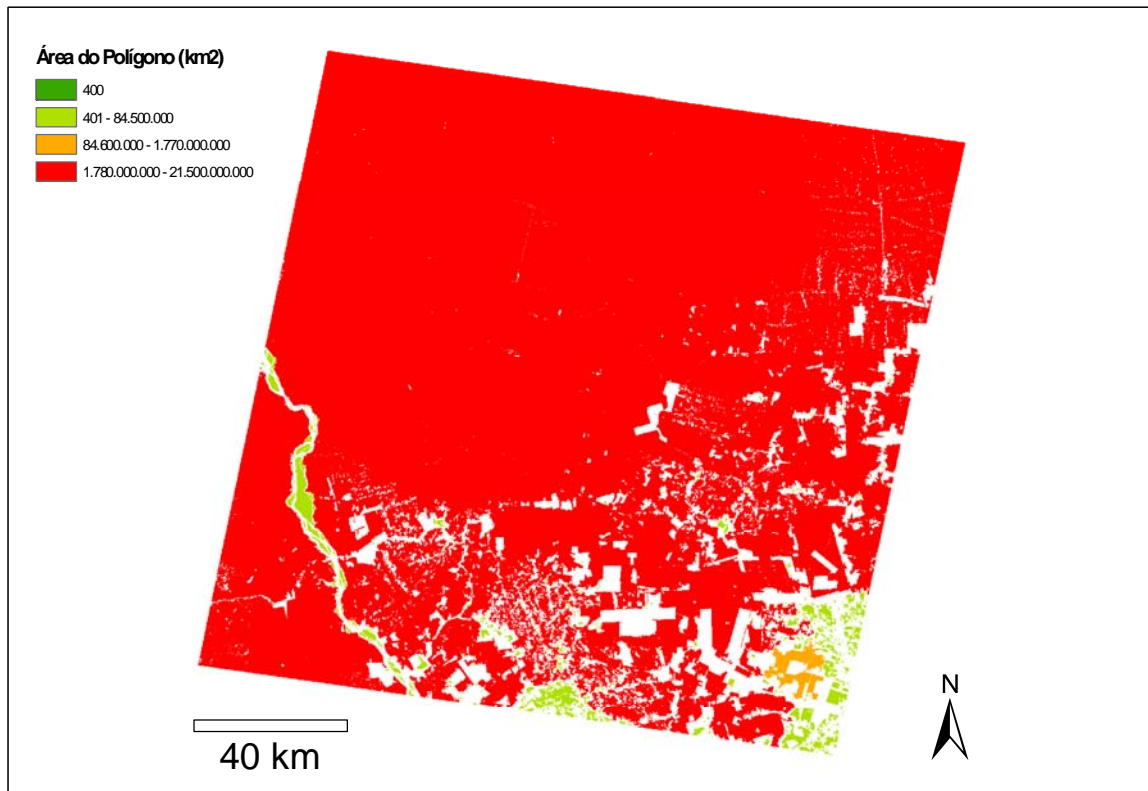


Fig. 3.19 – Espacialização da variável APE para o ano-base de 2000.

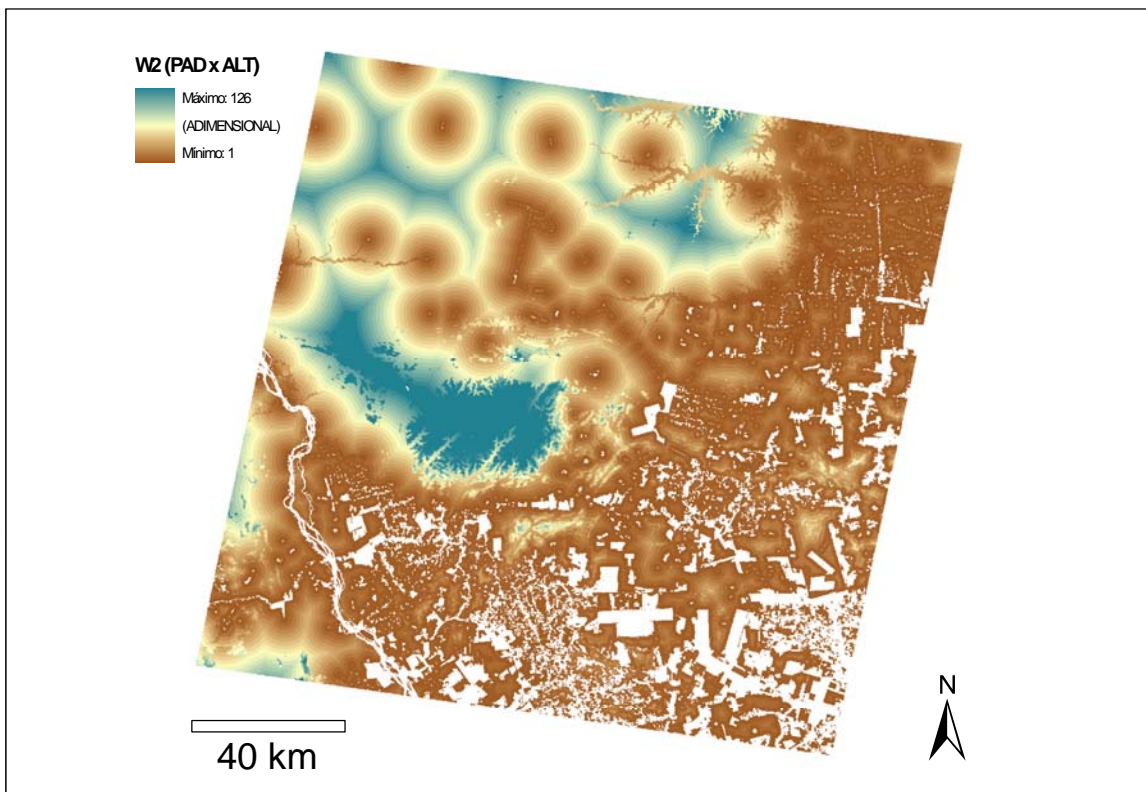
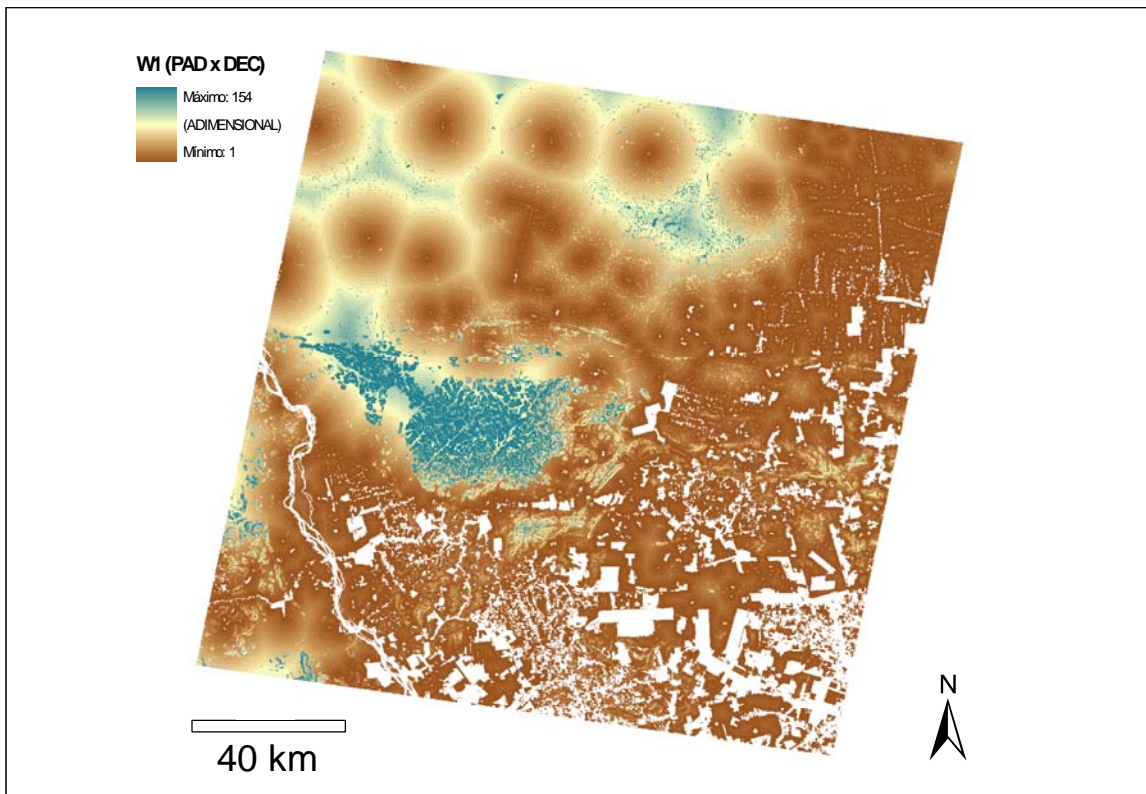


Fig. 3.20 – Exemplos de espacialização das variáveis W_1 (acima) e W_2 (abaixo) para o ano-base de 2000.

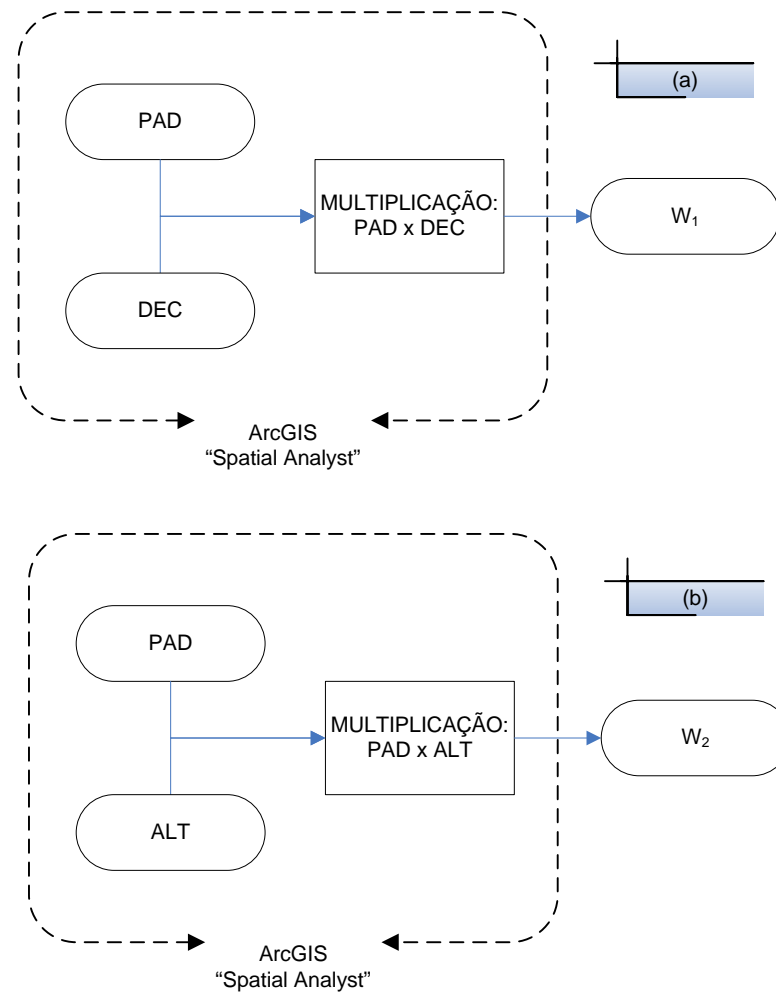


Fig. 3.21 – Esquema geral para obtenção das variáveis W_1 (a) e W_2 (b).

O quadro apresentado pela Figura 3.22 a seguir sumaria as variáveis criadas para o AMAZON e suas respectivas características funcionais.

Variáveis	Tipo	Observações	
IPT	Dependente	Binária (0 = floresta; 1 = desmatamento)	
Variáveis independentes	PAD	Var. de exposição	Variável categorizada, com espaçamento equidistante e intervalos de 1km.
	ROD	Var.de controle	Variável categorizada, com espaçamento equidistante e intervalos de 1km.
	HID	Var.de controle	Variável categorizada, com espaçamento equidistante e intervalos de 1km.
	EPA	Var.de controle	Binária (0 = protegida; 1 = não protegida)
	DEC	Var.de controle	Variável categorizada, com espaçamento equidistante, com intervalos de 5°.
	ALT	Var.de controle	Variável categorizada, com espaçamento equidistante, com intervalos de 90m.
	IFR	Var.de controle	Variável categorizada, obtida por otimização de Jenks, de 5 a 9 classes ⁴⁰ .
	APE	Var.de controle	Variável categorizada, obtida por otimização de Jenks, de 3 a 6 classes.
	W ₁ (PAD x DEC)	Var.de interação	Variável categorizada, com valores únicos adimensionais.
	W ₂ (PAD x ALT)	Var.de interação	Variável categorizada, com valores únicos adimensionais.

Fig. 3.22 – Quadro das variáveis criadas para utilização no AMAZON.

⁴⁰ - O número de classes obtido pelo método de Jenks é dependente de características intrínsecas aos dados, variando em função de tais características.

4 - FORMULAÇÃO E IMPLEMENTAÇÃO DO MODELO – RESULTADOS E DISCUSSÃO

4.1 - FORMULAÇÃO DO MODELO

O método estatístico de regressão – linear ou logística – não pressupõe necessariamente uma relação temporal entre a variável dependente e as variáveis independentes. Em geral, busca-se relacionar, para uma dada amostra de tamanho ‘n’, alguns fatores “explicativos” – as variáveis independentes – para um outro fator “explicado” – a variável dependente. A partir do estabelecimento de uma relação estatística entre esses fatores, torna-se possível “predizer” o parâmetro explicado do evento ‘n+1’, tomando-se por base seus atributos de caráter explicativo.

Com o AMAZON-PD, entretanto, tomando esse evento como sendo uma célula da imagem Landsat, não se pretende predizer o parâmetro explicado da célula ‘n+1’ da imagem, mesmo porque tal célula não existe, mas sim todas as ‘n’ células cobertas por essa imagem. Dessa forma, a relação entre a variável dependente – IPT – e as variáveis independentes – PAD, ROD, HID, EPA, DEC, ALT, IFR, APE, W_1 e W_2 – só faz sentido se houver um lapso temporal entre IPT e as demais variáveis, após o qual o efeito das variáveis explicativas sobre a variável explicada possa ser verificado.

Para tanto, aos “anos-base” foram atribuídos os parâmetros das variáveis independentes, cuja regressão logística permite predizer, em termos probabilísticos, a variável dependente da data subsequente, conforme será detalhado a seguir.

4.1.1 - Regressão logística entre diferentes datas – o modelo inicial

Uma vez que, das quatro classes inicialmente geradas – floresta, desmatamento, hidrografia e não-floresta – as duas últimas permaneceram virtualmente inalteradas ao longo de todo o período estudado, elas foram reclassificadas e passadas à categoria de *NoData*⁴¹. Com isso, criou-se uma seqüência de imagens divididas nas classes ‘floresta’ e ‘desmatamento’.

Como a variável-resposta (probabilidade de impacto ambiental – desmatamento) deve ser verificada na data seguinte àquela do ano-base, é essencial que a área na qual essa variável

⁴¹ - O aplicativo ArcGIS atribui a cada célula (x,y), de cada classe criada (floresta, desmatamento etc.), um valor (z), que pode ser usado algebricamente em diversas formas de operação entre bandas (planos de informação – PI). Assim, quando se deseja excluir determinada classe de uma operação, basta categorizá-la com ausência de valor (*NoData*), que não é o mesmo que zero.

é válida seja idêntica àquela do ano-base, onde são “interrogadas” as variáveis independentes. Assim, evita-se a comparação de áreas de tamanhos diferentes.

Para tanto, a “máscara⁴²” do ano-base é composta somente pela classe ‘floresta’. Na data seguinte, as células dessa área são confrontadas com as mesmas células, agora divididas em ‘floresta’ e ‘desmatamento’. Quando essa data se torna o ano-base em relação à data subsequente, a classe ‘desmatamento’ é reclassificada como *NoData* e assim por diante. Ao final, as seis datas consideradas no estudo (1985, 1988, 1992, 1996, 2000 e 2004) geram cinco modelos, referenciados nos anos-base.

No modelo inicial, dado pela equação abaixo, todas as variáveis concebidas para o AMAZON-PD, em um dado ano-base, são usadas na regressão logística que estabelece a probabilidade de ocorrência de impacto ambiental (sob a forma de desmatamento) relativa àquele ano-base, a ocorrer na próxima data.

$$\begin{aligned} \text{Logito}[P(IPT_{pd})] = \ln \left[\frac{P(IPT_{pd})}{1 - P(IPT_{pd})} \right] = \beta_0 + \beta_1 PAD_{ab} + \beta_2 ROD_{ab} + \beta_3 HID_{ab} + \\ + \beta_4 EPA_{ab} + \beta_5 DEC_{ab} + \beta_6 ALT_{ab} + \beta_7 IFR_{ab} + \beta_8 APE_{ab} + \beta_9 W_{1ab} + \beta_{10} W_{2ab} \end{aligned} \quad (4.1)$$

onde $P(IPT_{pd})$ é a probabilidade do impacto ambiental provocado pelo desmatamento na próxima data;

PAD_{ab} é a proximidade da área desmatada no ano-base;

ROD_{ab} é a proximidade de rodovias no ano-base;

HID_{ab} é a proximidade de hidrovias no ano-base (não implementada nesta tese);

EPA_{ab} é a existência (ou não) de área protegida no ano-base;

DEC_{ab} é a declividade no ano-base;

ALT_{ab} é a altitude do terreno no ano-base;

IFR_{ab} é o índice de fragmentação no ano-base;

APE_{ab} é a área do polígono envolvente da célula de floresta no ano-base;

W_{1ab} é o produto $PAD \times DEC$ no ano-base;

W_{2ab} é o produto $PAD \times ALT$ no ano-base; e

β_i são os coeficientes das variáveis independentes.

A discussão acerca do desempenho de cada uma das variáveis, apresentada a seguir, revela a primeira e mais marcante característica do AMAZON, que é a mudança na

⁴² - Área na qual são feitas as operações algébricas. Não pertence necessariamente ao polígono envolvente da imagem, uma vez que, no interior desse polígono, pode haver células com o atributo *NoData*.

representatividade de suas variáveis, à medida que evolui a cobertura do solo. Uma percepção da mudança na cobertura do solo para a área de estudo pode ser obtida a partir da seqüência de imagens classificadas em hidrografia (azul), não-floresta (magenta), floresta (verde) e desmatamento (amarelo), apresentadas na Figura 4.1.



Fig. 4.1 – Evolução das classes ‘floresta’ e ‘desmatamento’, de 1985 a 2004.

4.1.2 - Eliminação de variáveis – o modelo ideal

Partindo do modelo inicial, as variáveis são eliminadas uma a uma (para tanto, basta fazer o respectivo parâmetro β igual a ‘0’), no intuito de verificar qual é a configuração que adere melhor à predição de áreas desmatadas.

A configuração mais eficiente é obtida com base em tabelas de classificação, nas quais os valores esperados para ‘floresta’ e ‘desmatamento’ são confrontados com os valores observados para essas mesmas classes. Com isso, são gerados os termos de ajuste do modelo⁴³ (AM), de verdadeiros positivos⁴⁴ (VP) e de falsos positivos⁴⁵ (FP), que permitem escolher o modelo preservado para o ano-base em questão, com base no desempenho do modelo (DM).

A Tabela 4.1 sintetiza os parâmetros envolvidos na construção das tabelas de classificação empregadas em cada um dos modelos e apresentadas mais adiante⁴⁶. Os Anexos de I a V, por outro lado, apresentam todas as tabelas de classificação geradas ano a ano, com vistas a permitir a escolha de um modelo a preservar para cada ano-base, adotando como o critério o valor ‘DM’.

TABELA 4.1 – Síntese das tabelas de classificação utilizadas para avaliar o AMAZON

NOME DO MODELO PRESERVADO PARA O ANO-BASE		
Total de células da máscara [(++) + (+-) + (-+) + (--)]	Desmatamento observado (+)	Floresta observada (-)
Desmatamento esperado (+)	Número de células ‘++’	Número de células ‘-+’
Floresta esperada (-)	Número de células ‘+-’	Número de células ‘--’
$AM = \frac{(++)+(--)}{(++)+(+-)+(-+)(--)}$	$VP = \frac{(++)}{(++)+(+-)}$	$FP = \frac{(-+)}{(-+)(--)}$

Lesschen *et. al.* (2005) recomendam as tabelas de classificação para comparar resultados de modelos. Embora não existam regras gerais para se julgar os valores considerados aceitáveis, ajustes de modelos com percentuais superiores a 50% são estatisticamente melhores do que a aleatoriedade.

⁴³ - Percentual resultante do somatório das predições corretas para as classes ‘floresta’ e ‘desmatamento’, dividido pelo total de células da máscara analisada.

⁴⁴ - Percentual resultante da divisão do número de células corretamente preditas para a classe ‘desmatamento’ pelo número total de células dessa classe.

⁴⁵ - Percentual resultante da divisão do número de células erroneamente preditas para a classe ‘desmatamento’ pelo número total de células da classe ‘floresta’.

⁴⁶ - As tabelas de classificação aparecem identificadas pela letra ‘f’ das figuras 4.9 a 4.14.

Pontius e Schneider (2001), por sua vez, defendem que um valor superior a 70% é considerado aceitável para a modelagem de uso e de cobertura do solo, enquanto Hosmer e Lemeshow (2000) consideram percentuais superiores a 80% como excelentes, e superiores a 90% como excepcionais.

Para os fins deste estudo, o modelo a preservar é aquele que apresenta o maior índice resultante da operação abaixo (desempenho do modelo – DM), tomando-se o cuidado de não considerar modelos com valores de ‘AM’ e ‘VP’ inferiores a 70%, assim como com valores de ‘FP’ superiores a 30%. Enquanto o critério para os valores de ‘AM’ segue as orientações de Pontius e Schneider (2001), o critério de verdadeiros positivos superiores a 70% e de falsos positivos inferiores a 30% é adotado nesta tese por similaridade de raciocínio. Em consequência, estipula-se que DM seja superior a 0,28 [(0,7 * (0,7 - 0,3)].

$$DM = AM \times (VP - FP) \quad (4.2)$$

No Anexo 1, é apresentada a listagem de todas as tabelas dos modelos formulados para todos os anos-base usados nesta tese, tanto os modelos preservados quanto os não preservados.

i) Comportamento da variável PAD

Duas são as razões principais que tornam PAD uma variável de exposição: (1) o caráter contínuo da frente de desmatamento, que torna árvores próximas as candidatas mais fortes a ceder caminho para o avanço dessa frente; e (2) a constatação inequívoca dos parâmetros da variável em relação ao ano-base a que pertencem (Tabela 4.2), proveniente da interpretação, da segmentação e da classificação da imagem de satélite, cujos erros podem ser considerados desprezíveis na resolução adotada.

Essas razões, aliadas ao comportamento consistente do seu coeficiente (β_1) em todos os anos-base, fazem com que PAD seja mantida em cada um dos modelos preservados. A Figura 4.2 apresenta a evolução de PAD ao longo dos anos-base.

TABELA 4.2 – Comportamento isolado da variável de exposição PAD (1985 a 2000)

Ano-base	β_0	β_1 (PAD)
1985	-2,377	-0,161
1988	-2,374	-0,387
1992	-1,238	-0,439
1996	-1,330	-0,716
2000	-1,033	-0,390

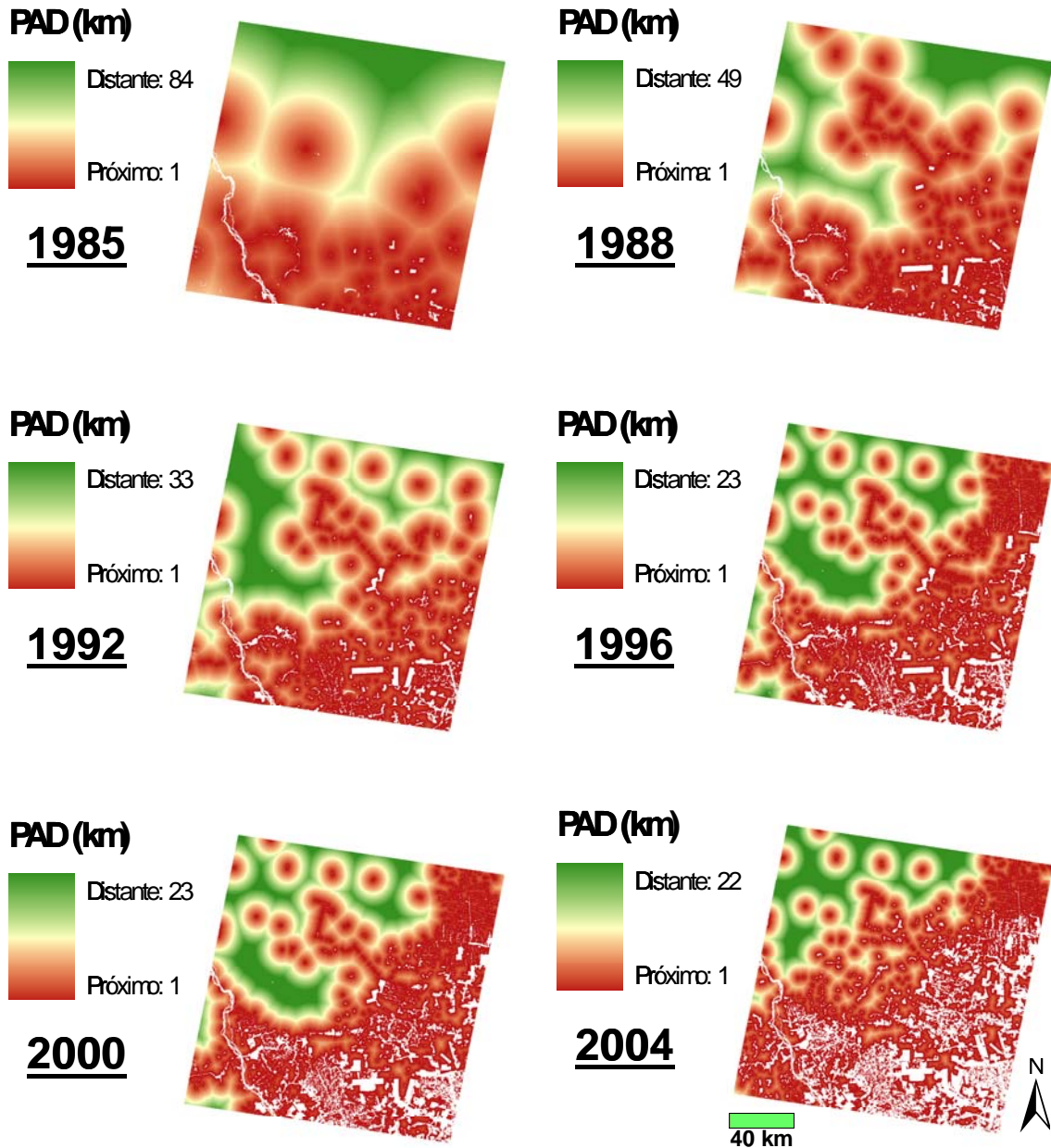


Fig. 4.2 – Evolução da Proximidade da Área Desmatada (PAD) de 1985 a 2004. No intuito de se manter um padrão de cores constante, a escala de distâncias para PAD assume um caráter relativo nessas imagens. Em consequência, as distâncias classificadas com a cor verde em 1985, por exemplo, são bem maiores do que aquelas de 2004.

ii) Comportamento da variável ROD

Seria normal se esperar que ROD apresentasse um desempenho ainda mais eficiente do que PAD, uma vez que a construção de rodovias em geral precede a atividade de desmatamento em larga escala. Entretanto, ao contrário de PAD, ROD mostrou-se uma variável bastante imprecisa, especialmente por se tratar de uma variável proveniente de fonte

secundária. O arquivo vetorial obtido junto ao Imazon, em 2005, possui uma taxa de atualização que não coincide com as datas das imagens usadas no estudo.

Deve-se ressaltar que a iniciativa do Imazon em preparar os dados vetoriais relativos à construção de rodovias economizou alguns meses de um árduo trabalho de interpretação de imagens. Contudo, sua baixa taxa de atualização, ao menos nesse caso específico, faz com que a verdadeira influência sendo exercida pela construção de novas rodovias seja omitida na data em que o evento ocorre, para ser revelada em uma data posterior. Isso fica particularmente claro na imagem de 1992, que confirma a existência de novas rodovias, embora elas ainda não tenham sido catalogadas.

Essa falta de sincronismo entre a construção de rodovias e sua catalogação provoca oscilações no grau de contribuição da variável, conforme se pode depreender da Tabela 4.3, que faz uso do modelo que contém apenas a variável de exposição PAD e a variável de controle ROD. Esta abordagem, usada para verificar o comportamento da variável ao longo do período observado, também é adotado nas tabelas posteriores, em relação às demais variáveis de controle.

Observe-se que a falta de registros confiáveis de rodovias ao longo dos anos do estudo faz com que os coeficientes de ROD apresentem grandes variações. São obtidos desde valores positivos, segundo os quais quanto maior for a distância da rodovia, maior será o desmatamento, até um coeficiente negativo de grande impacto efeito na variável dependente ($\beta_2 = -4,281$). Isso ocorre porque atividades de desmatamento são observadas em determinadas áreas onde provavelmente existem rodovias que não estão registradas no modelo.

TABELA 4.3 – Comportamento de ROD frente à variável de exposição PAD (1985 a 2000)

Ano-base	β_0	β_1 (PAD)	β_2 (ROD)
1985	-2,699	-0,173	+0,009
1988	-2,428	-0,390	+0,001
1992	-1,004	-0,338	-0,054
1996	-0,980	-0,699	-4,281
2000	-0,886	-0,216	-0,073

As figuras 4.3 e 4.4 tornam mais clara essa constatação. Por um lado, a Figura 4.3 denota a grande influência das rodovias nessa atividade, na medida em que apresenta uma forte coerência entre o crescimento da “mancha” de rodovias e o avanço da frente de

desmatamento. Por outro lado, a Figura 4.4 atesta o descompasso entre as atualizações dessas rodovias e suas identificações na imagem Landsat.

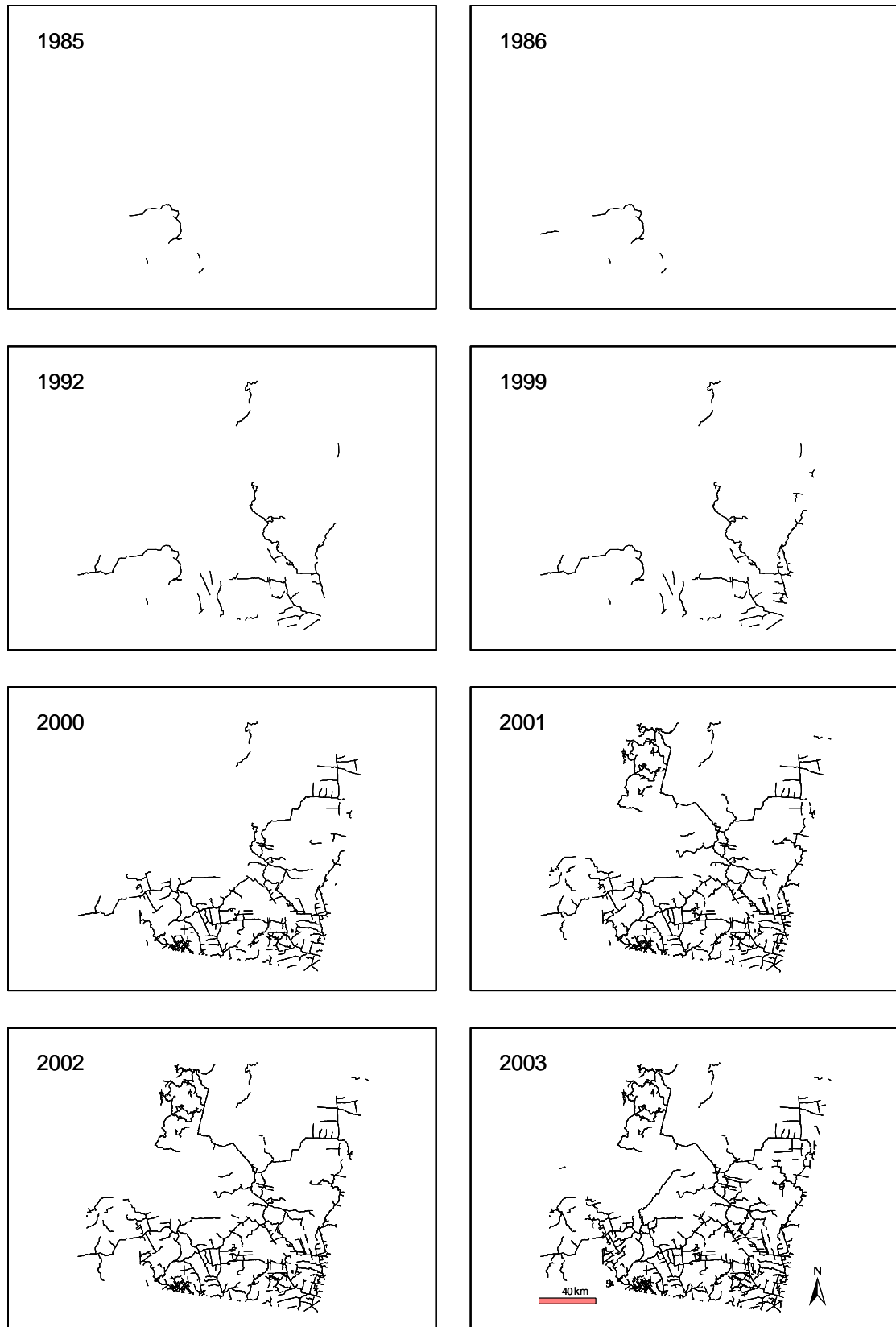


Fig. 4.3 – Datas de catalogação pelo Imazon de novas rodovias na área de estudo.

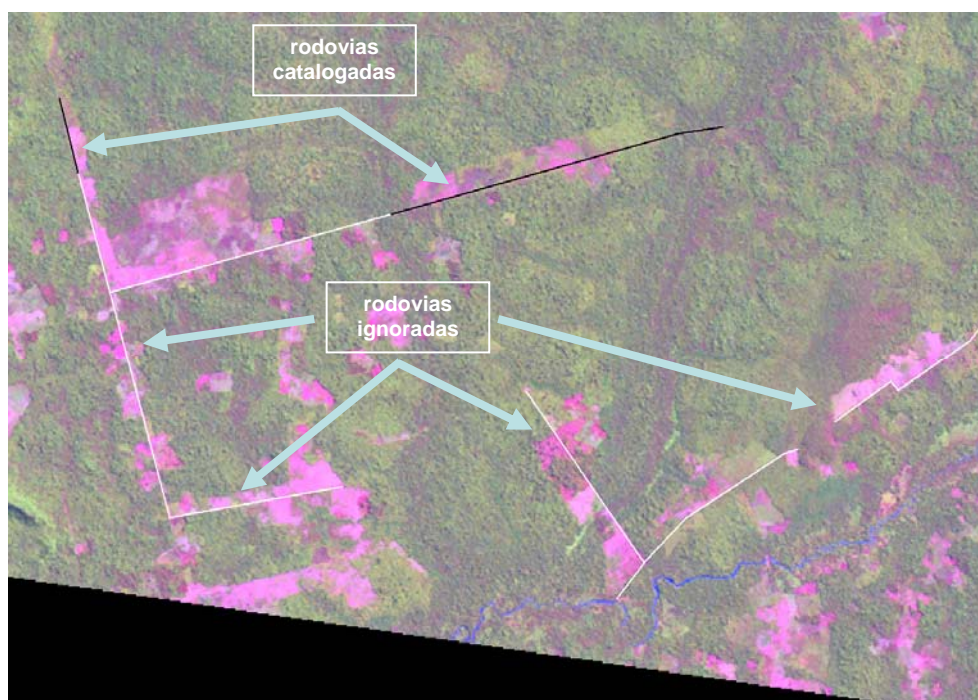


Fig. 4.4 – Exemplo de descompasso entre novas rodovias (brancas) que ainda não se encontram catalogadas e rodovias mais antigas (pretas), estas já catalogadas.

iii) Comportamento da variável EPA

Em 1985, 1988 e 1992, ainda não havia unidades de conservação (UC) na área de estudo, o que torna EPA uma variável irrelevante para essas datas.

Não obstante, uma forma inequívoca de validar o desempenho de EPA é a simples constatação visual de seu efeito de contenção de crescimento do desmatamento em uma área protegida, quando sua homologação se dá posteriormente à consolidação do desmatamento.

Um exemplo claro disso é demonstrado nos desmatamentos assinalados na Figura 4.5, nas cores vermelho e laranja, áreas onde UC/terras indígenas seriam homologadas mais tarde.

Observe-se que a criação da FLONA do Itacaí, assinalada em laranja na porção média da imagem de 2000, não impediu a progressão do desmatamento, verificada já na imagem seguinte, de 2004. Por outro lado, a homologação das terras indígenas de Trincheira/Bacaj, destacadas em vermelho na parte superior direita da imagem, bastou para conter o avanço da frente de desmatamento lá instalada desde antes da cena de 1996, que ficou praticamente intocada até 2004.

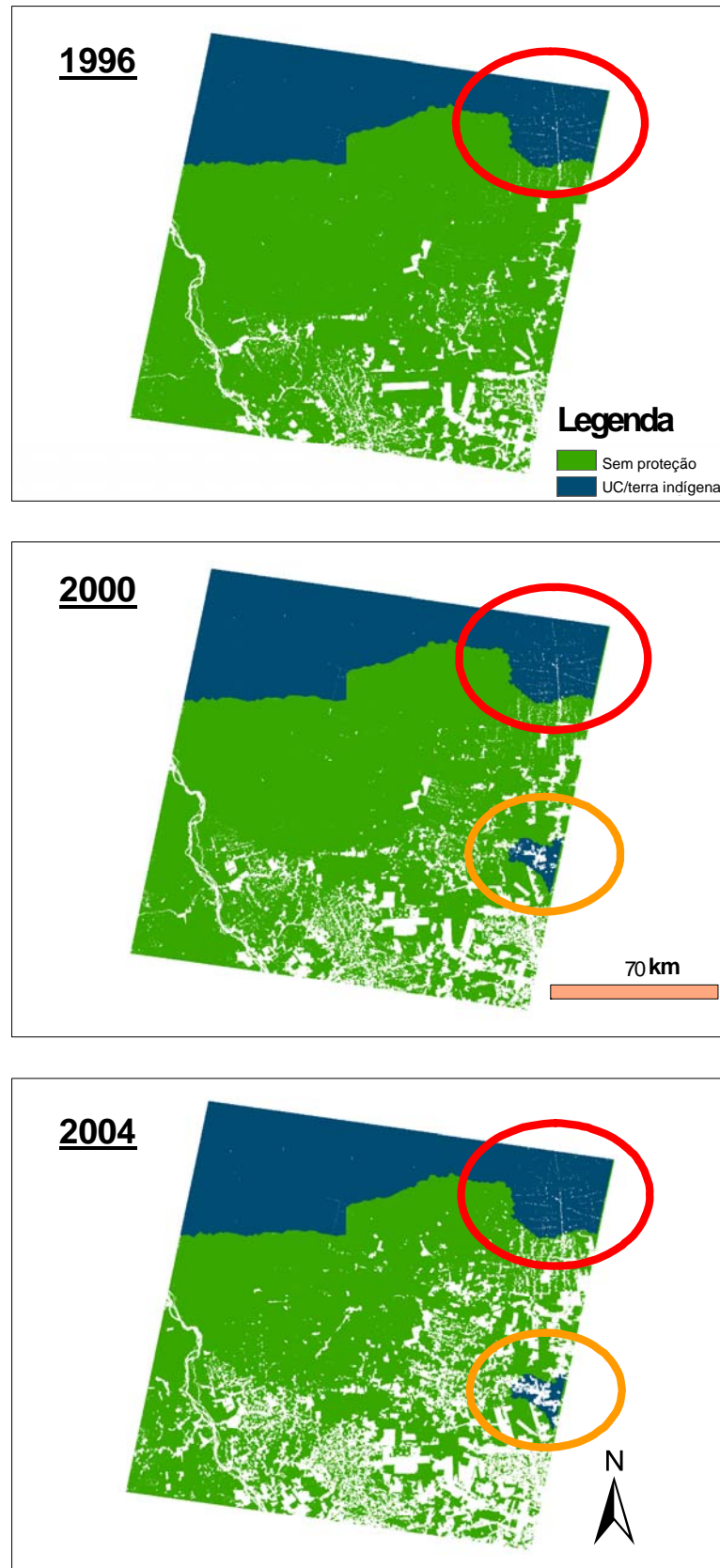


Fig. 4.5 – Evolução da variável EPA de 1996 a 2004.

A partir de 1996, foram homologados o Parque Nacional da Serra do Pardo e a FLONA do Itacaí, bem como as terras indígenas de Trincadeira/Bacaj e de Arawet/Ipixuna). Como resultado, EPA mostrou ser uma variável de controle relevante, uma vez que, nos anos-base de 1996 e 2000, mostrados na Tabela 4.4, os coeficientes não somente são representativos, mas também apresentam um sinal compatível com o esperado. Em outras palavras, ao assumir o sinal negativo, EPA “quer dizer” que, quanto menor for a proteção ambiental, maior será a probabilidade de desmatamento.

TABELA 4.4 – Comportamento de EPA frente à variável de exposição PAD (1996 a 2000)

Ano-base	β_0	β_1 (PAD)	β_2 (EPA)
1985	---	---	---
1988	---	---	---
1992	---	---	---
1996	-1,209	-0,699	-4,281
2000	-0,924	-0,356	-2,322

A terra indígena Apyterewa foi criada somente em 2004, mas ainda assim não havia sido homologada, o que fez com que sua área não fosse incluída neste estudo.

Observa-se, por fim, que a comparação visual entre os modelos gerados a partir de 1996 dão conta de que EPA é a principal variável responsável pela baixa degradação ambiental observada no setor Noroeste da área-teste investigada.

iv) Comportamento da variável DEC

Durante o processo de escolha das variáveis, DEC sugeria ser uma variável de controle de desempenho previsível. A razão estava no fato de que, para uma região de difícil acesso, como é a Amazônia, um relevo acidentado agrava o uso de máquinas (correntes ou motosserras) para o escoamento da produção madeireira, ao contrário de um terreno plano, mais propício para esse fim.

No trabalho de campo, contudo, a repetida visão de áreas recém-queimadas ou ainda em chamas jogou por terra aquela expectativa, conforme exemplifica a Figura 4.6. Obviamente, o fogo não se subordina à declividade do terreno. Ao contrário, quanto mais íngreme for a encosta florestal, mais rapidamente avançará a frente de incêndio “morro acima”.



Fig. 4.6 – Exemplo de queimada “morro acima”, diversas vezes observada no trabalho de campo.

Assim sendo, à medida que a ocorrência de áreas desmatadas por fogo se intensifica ao longo dos anos, a contribuição de DEC para o modelo regride sistematicamente, sugerindo até mesmo uma inversão do sinal a partir de 2004 (quanto mais acidentado o terreno, maior o desmatamento), conforme se pode observar na Tabela 4.5.

TABELA 4.5 – Comportamento de DEC frente à variável de exposição PAD (1985 a 2000)

Ano-base	β_0	β_1 (PAD)	β_2 (DEC)
1985	-1,592	-0,163	-0,602
1988	-1,903	-0,392	-0,349
1992	-1,029	-0,442	-0,146
1996	-1,140	-0,720	-0,132
2000	-0,931	-0,391	-0,071

Isso não quer dizer que DEC seja uma variável pouco útil para modelos como o AMAZON. Ao contrário, seu comportamento reflete rigorosamente as observações feitas no tratamento dos dados estudados e ratifica a prática comum de uso do fogo para “limpar” a floresta, quando as madeiras de maior valor comercial já foram extraídas das áreas próximas às encostas pelos

métodos tradicionais, com correntes ou com motosserras. Dessa maneira, encaixa-se no conhecido ciclo “exploração madeireira predatória/assentamento/pecuária extensiva”.

v) *Comportamento da variável ALT*

De maneira similar à variável DEC, esperava-se que o fator altitude tivesse influência previsível sobre o desmatamento, uma vez que as populações amazônicas tendem a viver próximo às margens dos rios, que constituem locais consistentemente mais baixos. Assim, quanto maior fosse a altitude da célula de floresta, mais difícil seria o acesso e, em consequência, menos intensa seria a atividade de desmatamento. O modelo também falseou essa hipótese secundária, novamente em função do desmatamento provocado por incêndios.

O coeficiente β_1 apresentou uma tendência de comportamento similar à tendência apresentada por DEC. No entanto, a provável inversão de sinal sugerida pela análise de DEC, a partir de 2004, ocorreu efetiva e consistentemente com ALT, conforme se pode verificar na Tabela 4.6, já a partir de 1996. Seguindo um padrão de tendência bem caracterizado, ALT tem seu coeficiente reduzido – até ser invertido – com o passar dos anos, à medida que se intensifica o uso do fogo como forma de manejo florestal.

TABELA 4.6 – Comportamento de ALT frente à variável de exposição PAD (1985 a 2000)

Ano-base	β_0	β_1 (PAD)	β_2 (ALT)
1985	-1,824	-0,161	-0,234
1988	-1,957	-0,392	-0,172
1992	-0,897	-0,445	-0,139
1996	-1,330	-0,716	0,000
2000	-1,282	-0,386	+0,103

vi) *Comportamento da variável IFR*

Conforme foi mencionado no Capítulo 3, o índice de fragmentação sugere que, quanto mais circular for a área contínua de floresta (IFR tende a ‘0’), menor será a probabilidade de desmatamento. A recíproca é verdadeira, ou seja, quanto mais fragmentada estiver a área (IFR tende a ‘1’), maior será a probabilidade de desmatamento.

Na prática, porém, os resultados não se mostraram tão coerentes com aquela expectativa. Os locais que apresentaram os menores IFR realmente apresentaram formas arredondadas, mas certamente não estavam protegidos da atividade de desmatamento por conta dessa característica, mas por se tratarem de ilhas fluviais, conforme mostra a Figura 4.7.

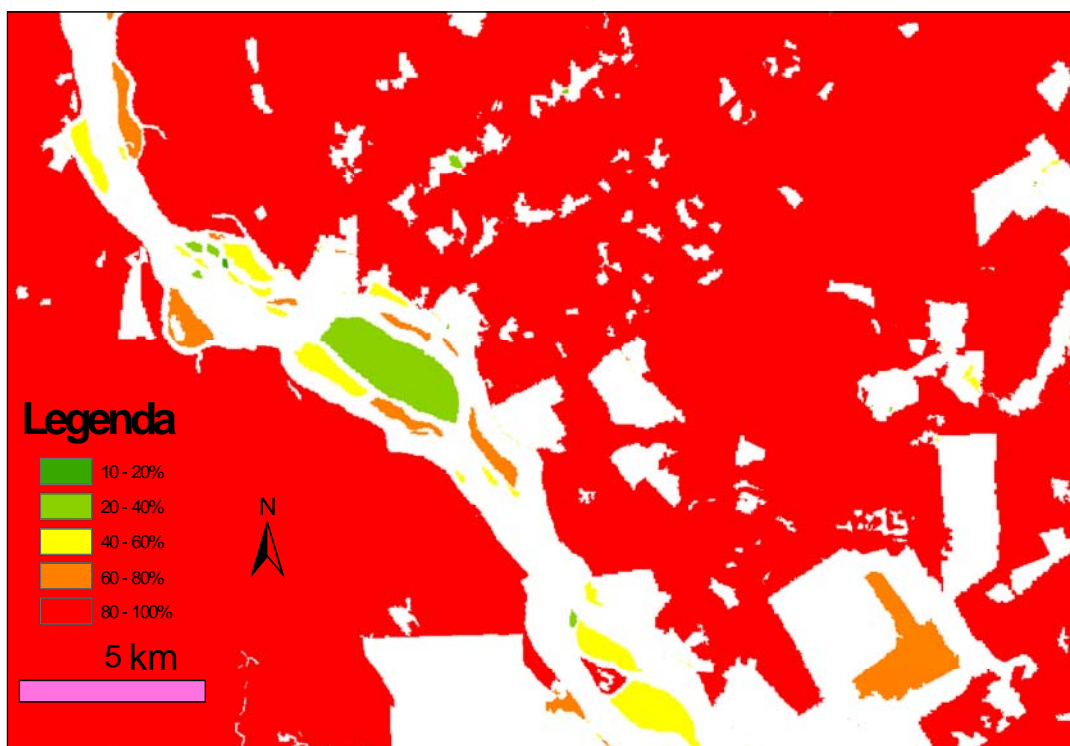


Fig. 4.7 – Exemplo de células (ano-base de 2000) com baixo índice de fragmentação (verde).

Por outro lado, a grande vastidão florestal, localizada na porção superior da imagem constituía praticamente um único polígono e os inúmeros recortes vizinhos ou internos a esse polígono aumentaram muito o seu perímetro relativamente à área, o que fez com que quase a totalidade da imagem apresentasse um alto índice de fragmentação.

Em consequência, a variável IFR agregou pouco ou nenhum valor ao modelo, conforme ficou claro a partir da interpretação de que a variável IFR atuava como se fosse uma constante, o que faz sentido, quando se observa sua espacialização (Capítulo 3, Figura 3.16), que nos mostra muito pouca variação nesses índices, pelo menos até o ano-base de 2000.

Uma ressalva a ser feita, entretanto, é de que IFR agregou pouco valor ao AMAZON para esta cena Landsat, o que não quer dizer que IFR não tivesse influência no modelo em alguma outra circunstância. Uma avaliação mais específica teria que ser realizada em outras áreas, de maneira a dimensionar melhor a contribuição dessa variável.

vii) *Comportamento da variável APE*

O mesmo fenômeno que praticamente anulou IFR foi responsável pela ineficiência da variável APE. Em tese, áreas pequenas significam alta probabilidade de desmatamento no

momento seguinte. Entretanto, as pequenas ilhas florestadas do rio Xingu permaneceram intactas, mais em função de suas localizações do que de suas dimensões. As grandes áreas de floresta contidas em alguns poucos polígonos, entretanto, tornaram APE uma variável capaz de agregar muito pouco ao AMAZON, pelo menos até a fase final do período estudado. Inicialmente, como ocorreu com IFR, APE apresentou um comportamento praticamente constante.

viii) Comportamento das variáveis W_1 (PAD x ALT) e W_2 (PAD x DEC)

Ainda que não forneçam uma interpretação intuitiva, as variáveis multiplicativas podem agregar valor ao modelo, quando seus componentes possuem aspectos convergentes em termos de contribuição para com a variável dependente, na medida em que evidenciam as contribuições isoladas desses componentes.

No caso deste estudo, tanto W_1 quanto W_2 tiveram bons desempenhos na composição dos modelos testados. No processo de escolha dos modelos preservados, entretanto, somente W_1 se tornou parte de um modelo preservado, o que ocorreu no ano-base de 1996.

4.2 - IMPLEMENTAÇÃO DO MODELO

Nesta Seção, são discutidos os resultados com o AMAZON ano a ano. De maneira a permitir um melhor entendimento do esquema adotado nessas figuras, a Figura 4.8 sintetiza os componentes utilizados nessa parte da operação, bem como ilustra graficamente os conceitos de ajuste do modelo (AM), de verdadeiros positivos (VP) e de falsos positivos (FP). As imagens que representam esses resultados estão todas dispostas seqüencialmente nas figuras 4.9, 4.10, 4.11, 4.12 e 4.13, que são relativas aos anos-base de 1985, 1988, 1992, 1996 e 2000, respectivamente.

Essa abordagem permite esclarecer, por exemplo que, apesar de contar sempre com um número maior de células, os resultados obtidos para FP são sempre menores do que VP em termos percentuais. Isso ocorre porque a tabela de classificação que gera os valores de AM, VP e FP está baseada em quatro categorias geradas a partir de duas classes – floresta (-) e desmatamento (+) – com contagens bastante distintas, que variam de cerca de 98,5% e 1,5% para o ano-base de 1985, até cerca de 91% e 9% para o ano-base de 2000, respectivamente.

Como VP decorre dos acertos estimados para a classe ‘desmatamento’, seus valores absolutos são pequenos, em contraste com valores percentuais altos. Por outro lado, FP decorre dos erros dessa estimação, que são verificados na classe ‘floresta’, muito maior, o que gera contagens altas em termos absolutos, mas bastante reduzidas em termos percentuais

(<30%) e que, naturalmente, são os valores de interesse para avaliação de desempenho do modelo por meio da técnica de tabela de classificação.

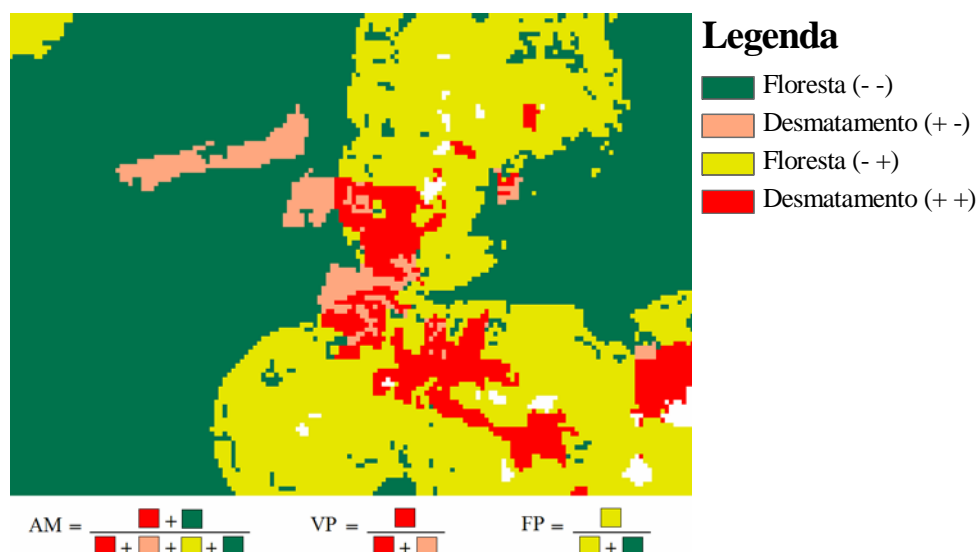


Fig. 4.8 – Ilustração dos componentes envolvidos na geração das tabelas de classificação.

Tomando como exemplo o ano-base de 1985 e partindo do modelo gerado na ferramenta ArcGrid do ArcGIS 9.2 (Figura 4.9a), um mapa de probabilidade de desmatamento é gerado (Figura 4.9b) e posteriormente categorizado em duas classes (valores > 0,5 ⇒ ‘1’; valores < 0,5 ⇒ ‘0’). A essa imagem (Figura 4.9c), é somada a variável dependente ‘IPT’ (Figura 4.9d) da data subsequente. O resultado, dividido em quatro classes (Figura 4.9e) – floresta (- -), desmatamento (+ -), floresta (- +) e desmatamento (+ +) – permite a geração da tabela de classificação (Figura 4.9f) que quantifica o desempenho do modelo. O mesmo padrão é utilizado nas figuras subsequentes, de 4.10 a 4.14.

As letras e cores atribuídas aos cabeçalhos dos diversos modelos são detalhadas nos Anexos de I a V. Basicamente, elas indicam as variáveis utilizadas no melhor resultado alcançado para aquele ano-base. Assim, no modelo referente ao ano-base de 2000 (Figura 4.13), por exemplo, são atribuídas as letras ‘P-R-E-D-A’, o que significa dizer que aquele modelo faz uso das variáveis ‘PAD’, ‘ROD’, ‘EPA’, ‘DEC’, e ‘ALT’.

Como foi dito anteriormente, o critério para a escolha do melhor modelo referente a cada ano-base é arbitrado pela Equação 4.2. Para cada um dos anos considerados, também é possível observar outros resultados promissores, respeitados os critérios de ‘DM’ superior a 0,28, ‘AM’ e ‘VP’ superiores a 70% e ‘FP’ inferior a 30%. É possível que, em outras situações, o modelo preservado seja referente a um desses resultados próximos aos aqui escolhidos, em função da possibilidade de se obter variáveis de controle mais precisas do que foi praticável alcançar com esta tese.

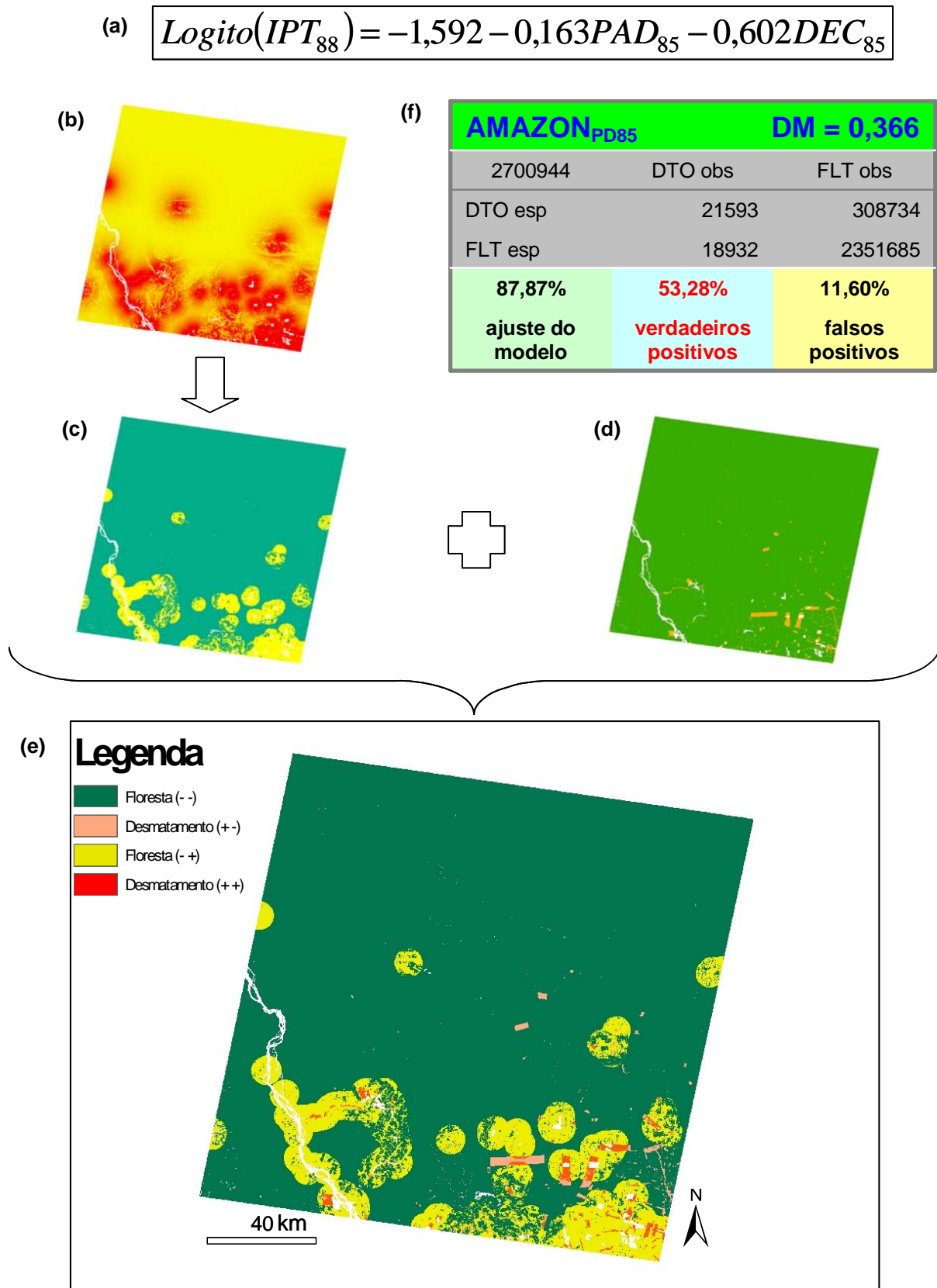
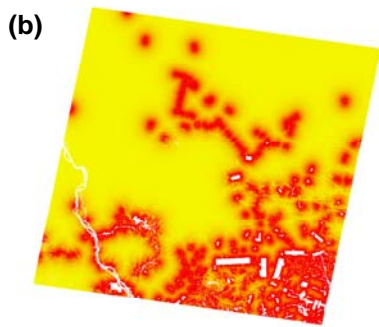


Fig. 4.9 – Melhor desempenho do AMAZON no ano-base de 1985.

(a)
$$\text{Logito}(IPT_{92}) = -1,903 - 0,392PAD_{88} - 0,349DEC_{88}$$



(f)

AMAZON _{PD88}		DM = 0,460
2660445	DTO obs	FLT obs
DTO esp	28439	410554
FLT esp	11995	2209457
84,12%	70,33%	15,67%
ajuste do modelo	verdadeiros positivos	falsos positivos

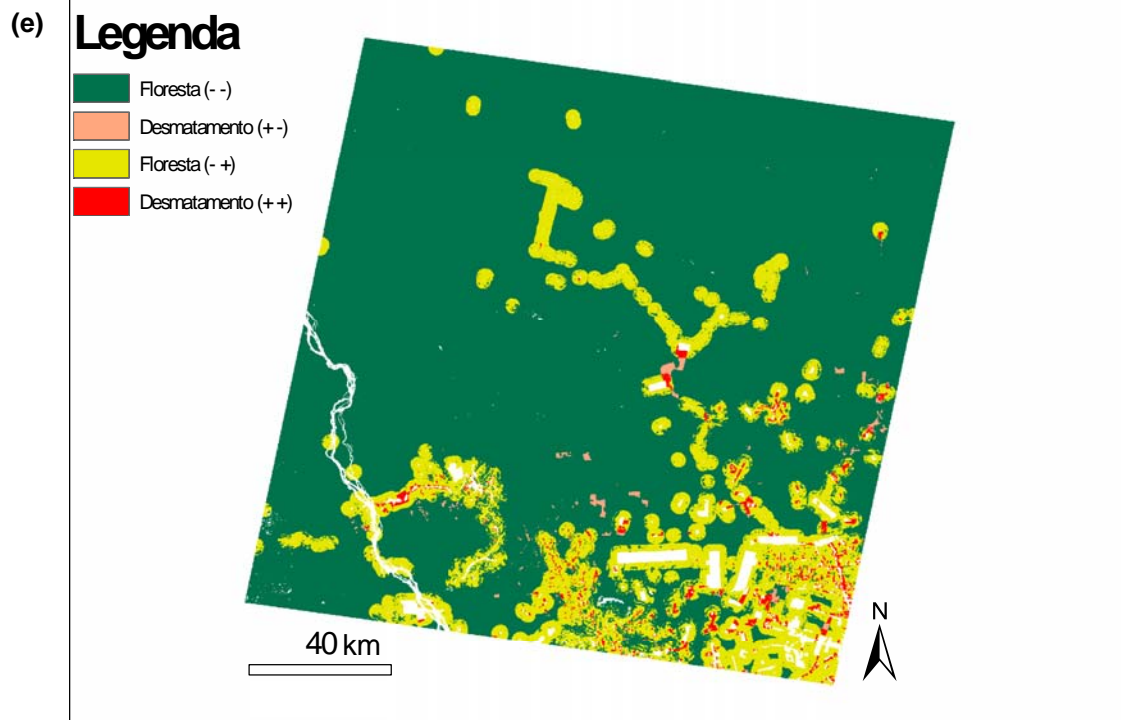
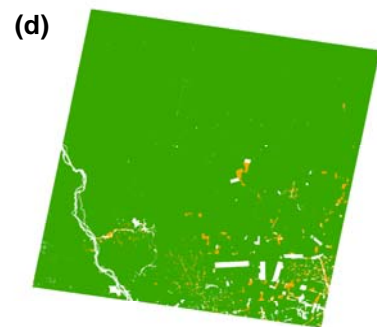
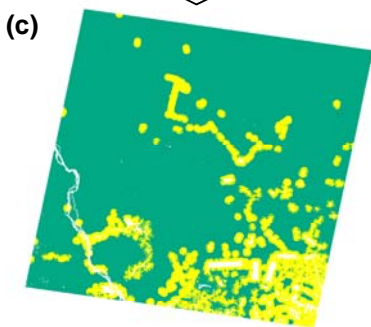
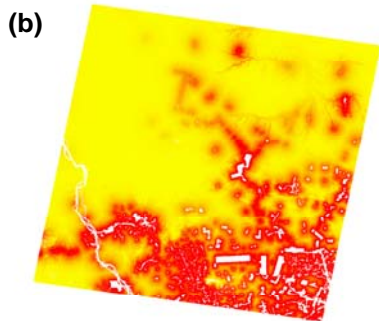


Fig. 4.10 – Melhor desempenho do AMAZON no ano-base de 1988.

(a)
$$\text{Logito}(IPT_{96}) = -0,897 - 0,445PAD_{92} - 0,139ALT_{92}$$



(f)

AMAZON_{PA92}		DM = 0,410
2620011	DTO obs	FLT obs
DTO esp	90351	523715
FLT esp	33365	1972580
78,74%	73,03%	20,98%
ajuste do modelo	verdadeiros positivos	falsos positivos

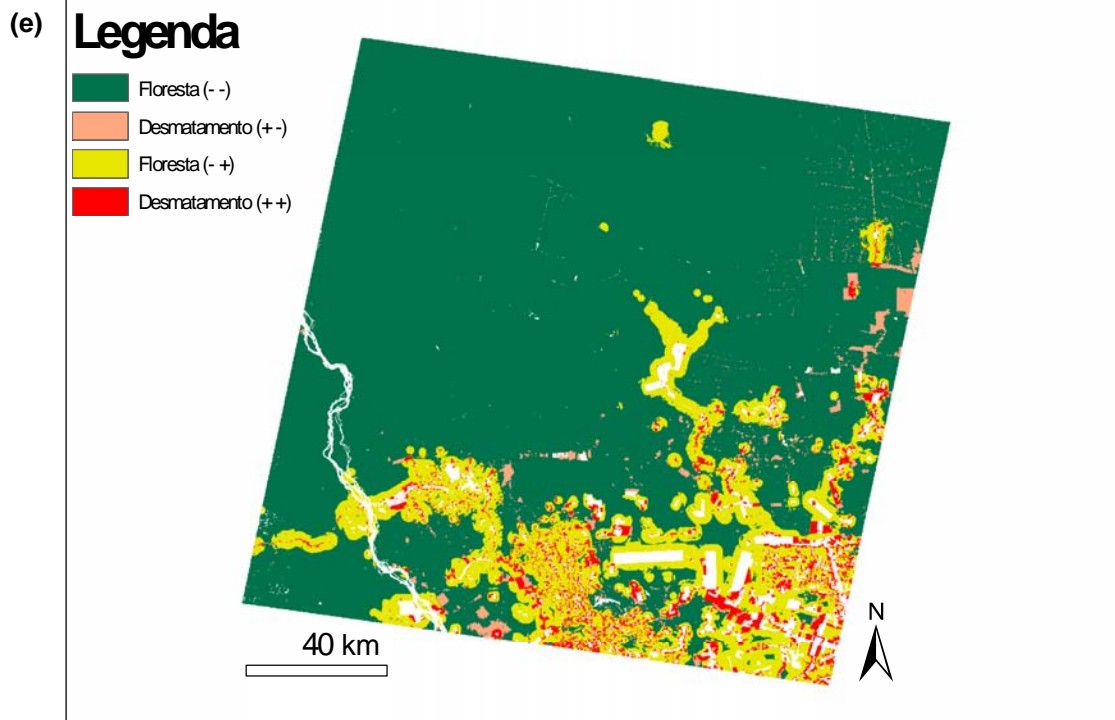
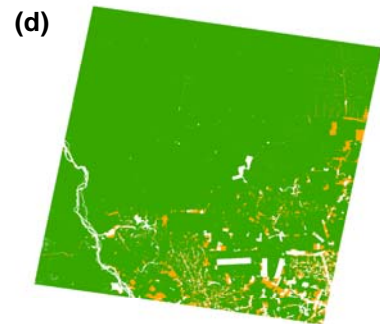
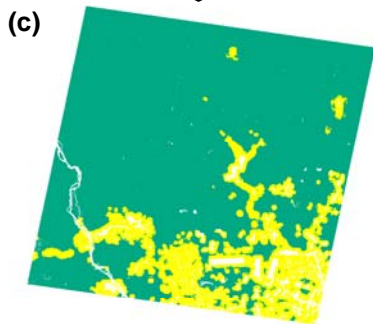
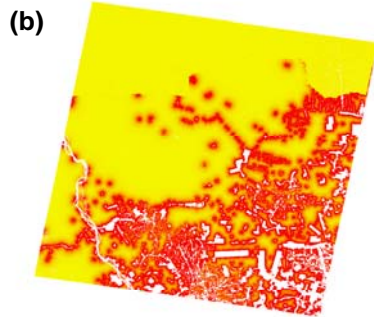


Fig. 4.11 – Melhor desempenho do AMAZON no ano-base de 1992.

(a)

$$\text{Logito}(IPT_{00}) = -0,697 - 0,570PAD_{96} - 0,037ROD_{96} - 3,903EPA_{96} - 0,202DEC_{96} - 0,026W_{196}$$



(f)

AMAZON_{PREDW1}		DM = 0,393
2496295	DTO obs	FLT obs
DTO esp	75523	492518
FLT esp	31723	1896531
79,00%	70,42%	20,62%
ajuste do modelo	verdadeiros positivos	falsos positivos

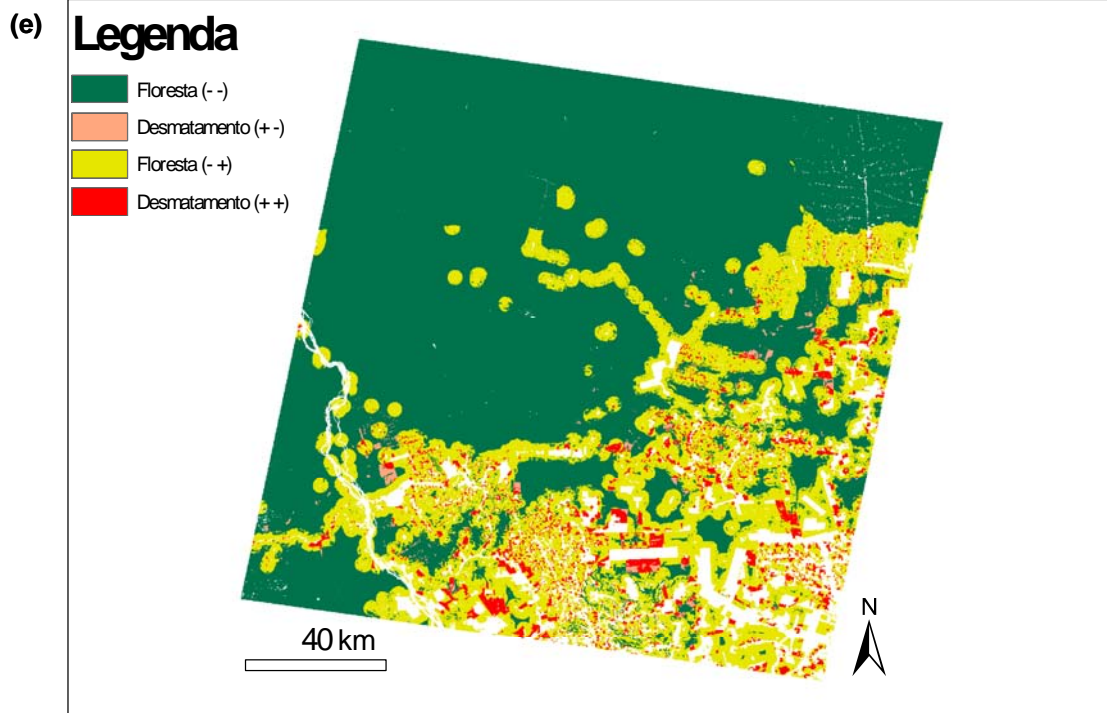
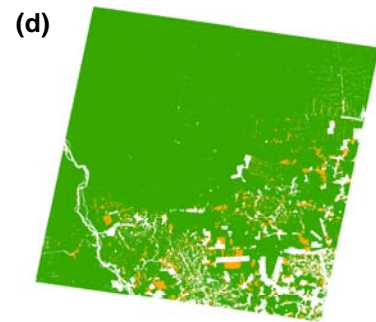
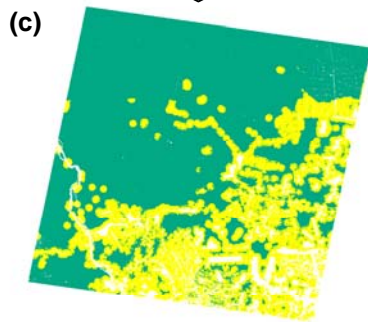
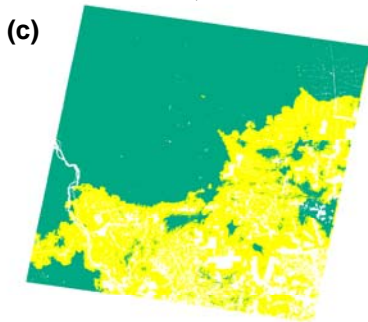
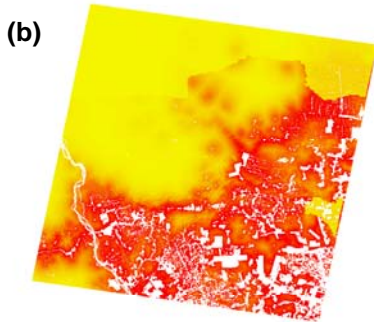


Fig. 4.12 – Melhor desempenho do AMAZON no ano-base de 1996.

(a)

$$\text{Logito}(IPT_{04}) = -0,703 - 0,208PAD_{00} - 0,067ROD_{00} - 2,145EPA_{00} - 0,210DEC_{00} + 0,087ALT_{00}$$



(f)

AMAZON_{PRED}A00		DM = 0,370
2387747	DTO obs	FLT obs
DTO esp	170889	543933
FLT esp	58063	1614862
74,79%	74,64%	25,20%
ajuste do modelo	verdadeiros positivos	falsos positivos

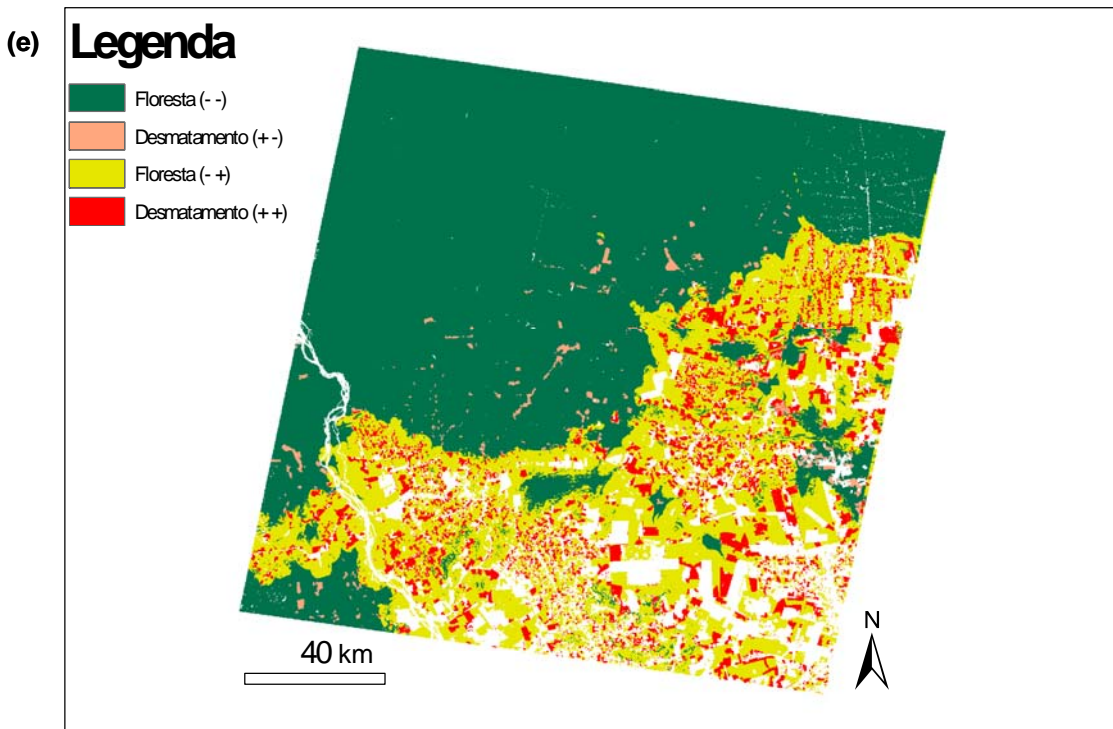
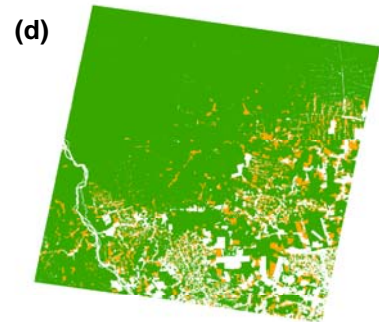


Fig. 4.13 – Melhor desempenho do AMAZON no ano-base de 2000.

4.2.1 - AMAZON – o modelo estático

De maneira geral, nos primeiros estágios do desmatamento (1985, 1988 e 1992), o AMAZON apresenta melhor desempenho somente com as variáveis ‘PAD’, ‘DEC’ e ‘ALT’.

À medida que a atividade de desmatamento é intensificada, o desempenho do AMAZON melhora com o incremento de novas variáveis, como ‘EPA’ e ‘ROD’, e até mesmo com as variáveis ‘IFR’, ‘APE’ e as multiplicativas (‘W1’ e ‘W2’).

A partir de 2000, as variáveis ‘IFR’ e ‘APE’ passaram a ter influência no modelo, que o Capítulo 3 evidencia graficamente por intermédio das figuras 3.16 e 3.17, respectivamente.

As conclusões obtidas a partir dos diversos modelos gerados ano a ano serviram de base de entendimento e de validação das variáveis escolhidas para a formulação do AMAZON. Para tanto, a variável dependente (‘IPT’) é obtida pela classificação ‘floresta x desmatamento’ da data subsequente à do ano-base ao qual se aplica o modelo.

Como foi dito acima, isso permite um melhor entendimento do papel de cada uma das variáveis preservadas do modelo. Por outro lado, essa abordagem não torna possível a proposta original de predição⁴⁷ da atividade de desmatamento para datas futuras, uma vez que ‘IPT’ não está disponível nesses casos.

A forma idealizada para solucionar esse problema foi a análise do comportamento dos coeficientes (β_i) de cada variável ao longo do período 1985-1996, com vistas a aplicar o conhecimento daí resultante no ano-base de 2000 e validá-lo com os dados de 2004.

4.2.2 - AMAZON-PD – o modelo dinâmico

Como nem todas as variáveis utilizadas permitem sua observação ao longo de todo o período estudado, ‘PAD’ foi escolhida para, isoladamente, testar o conceito de agregação da dimensão temporal ao estudo. A Tabela 4.7 permite inferir uma clara tendência na variação dos valores de β_0 e β_1 . Com isso, por meio de uma regressão linear de β_0 e β_1 dos anos-base de 1985, 1988, 1992 e 1996, foi possível estimar esses coeficientes para 2000.

A agregação da dimensão temporal à regressão logística permitiu um melhor balanceamento do modelo, com a conseqüente melhoria de um desempenho que já era bastante satisfatório. Como resultado, houve uma redução das ocorrências de ‘falsos

⁴⁷ - Nunca é demais enfatizar que a predição estatística não diz respeito a tempos futuros, mas sim a novos eventos.

positivos’ e uma melhoria no ‘ajuste do modelo’, em troca de uma redução de ‘verdadeiros positivos’, trazendo o conjunto para os valores especificados para este trabalho de pesquisa.

O modelo resultante, denominado de ‘Análise Multivariada Aplicada a Zoneamento para Predição de Desmatamento (AMAZON-PD⁴⁸)’, gerou o mapa de tendências apresentado na Figura 4.14, cuja tabela de classificação apresenta um desempenho ainda melhor do que aquele alcançado para o mesmo ano-base com o conhecimento prévio da variável dependente ‘IPT₀₄’ (Tabela 4.7), uma vez que o modelo passou a contar com o “aprendizado” das cenas de datas anteriores.

Vale destacar que o desempenho do AMAZON-PD_{P00}, detalhado na Figura 4.14, apresentou um valor de DM = 0,348, bastante superior àquele encontrado para o AMAZON_{P00} (DM = 0,254). Este, além de apresentar um desempenho inferior, também o faz de maneira desbalanceada, de acordo com os critérios explicitados na Equação 4.2⁴⁹ (AM < 70% e FP > 30%).

Após essas considerações, o AMAZON-PD ficou assim definido:

$$\text{Logito}[P(IPT_{pd})] = \ln \left[\frac{P(IPT_{pd})}{1 - P(IPT_{pd})} \right] = \beta_{0_{reg}} + \beta_{1_{reg}} PAD_{ab} + \beta_{2_{reg}} ROD_{ab} + \beta_{3_{reg}} HID_{ab} + \beta_{4_{reg}} EPA_{ab} + \beta_{5_{reg}} DEC_{ab} + \beta_{6_{reg}} ALT_{ab} + \beta_{7_{reg}} IFR_{ab} + \beta_{8_{reg}} APE_{ab} + \beta_{9_{reg}} W_{1ab} + \beta_{10_{reg}} W_{2ab} \quad (4.3)$$

onde $\beta_{i_{reg}}$ são os coeficientes das variáveis independentes, obtidos via regressão a partir dos coeficientes dessas variáveis até a data anterior ao ano-base.

TABELA 4.7 – Tabela de classificação do AMAZON_{P00}

AMAZON _{P00}		DM = 0,254
2387747	DTO obs	FLT obs
DTO esp	210723	1018541
FLT esp	18229	1140254
56,58%	92,04%	47,18%
ajuste do modelo	verdadeiros positivos	falsos positivos

⁴⁸ - Pela forma como foi concebido, o AMAZON pode vir a adotar diferentes técnicas de análise multivariada à parte da regressão logística e, em função disso, possibilitar outros tipos de aplicação, que extrapolam o objetivo desta tese. Neste estudo, foi adotada uma extensão como forma de caracterizar a aplicação específica do modelo (‘AMAZON-PD’ para ‘Predição do Desmatamento’). Na eventualidade de uma aplicação distinta à desta tese, o modelo passaria a adotar diferentes extensões (‘AMAZON-CD’ para ‘Canais de Distribuição’, por exemplo).

⁴⁹ - DM = AM x (VP – FP)

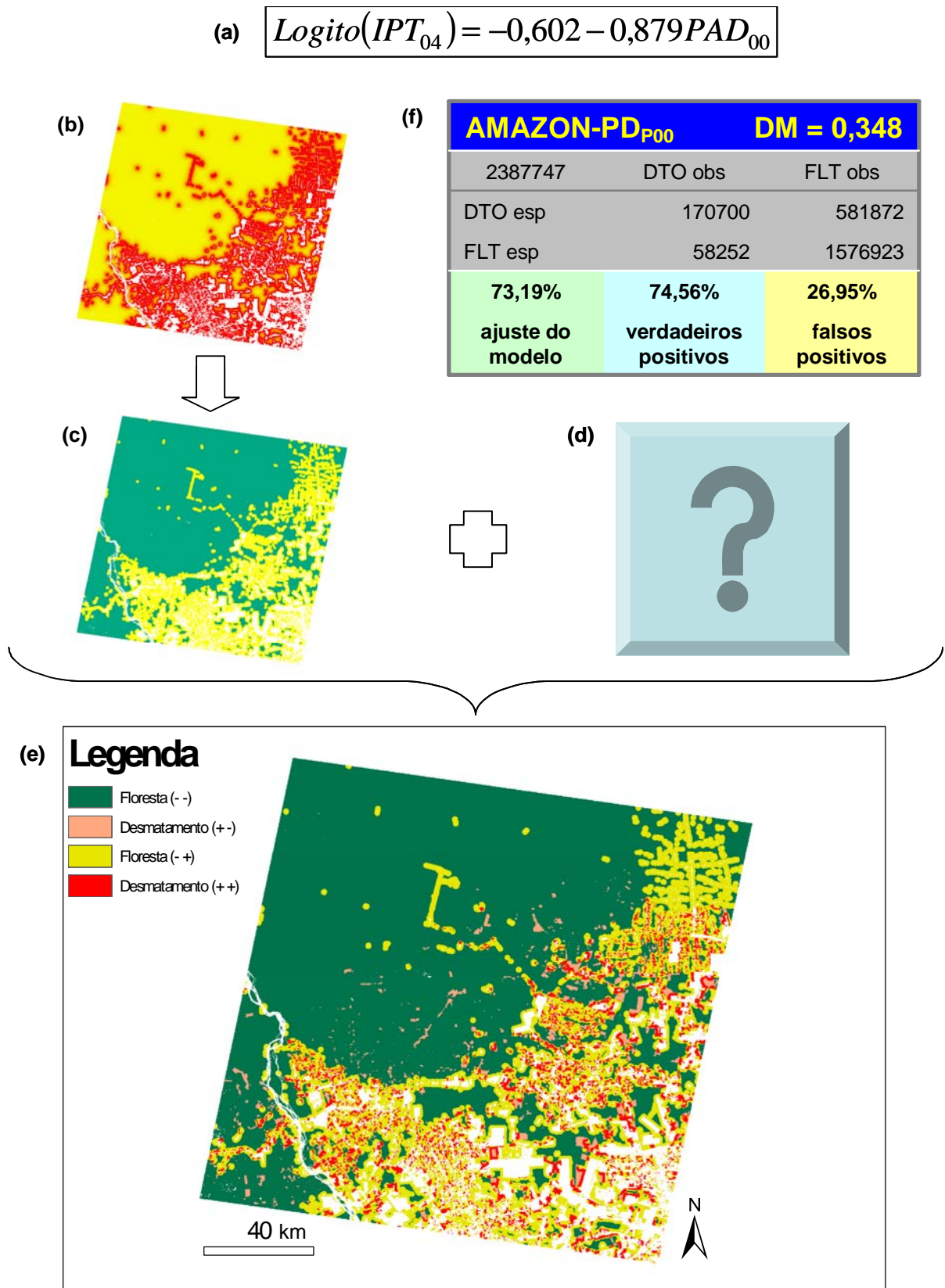


Fig. 4.14 – Modelo para o ano-base de 2000, que utiliza somente a variável PAD (P00), mas com a incorporação da dimensão temporal (AMAZON-PD).

Utilizando-se do mesmo expediente aplicado na validação do modelo AMAZON-PD em 2000, é possível prever a variável IPT para 2008, a partir do conjunto de imagens anteriores a 2004. Resta, entretanto, especular sobre qual configuração de modelo possui o melhor desempenho, dado que somente em 2008 será possível demonstrar esse resultado. Assim, três possibilidades de AMAZON-PD₀₄ foram criadas e são apresentadas a seguir.

A primeira delas, definida como AMAZON-PD_{P04}, levou em conta todas as datas anteriores (1985-2000) e somente a variável PAD, presente em todas essas datas, foi usada no modelo. O resultado é apresentado na Figura 4.15(a).

A Figura 4.15(b) apresenta o modelo AMAZON-PD_{PA04}, que emprega as variáveis PAD e ALT referentes aos anos-base de 1988, 1996 e 2000, datas nas quais essas variáveis foram mais consistentes em termos de desempenho. Em relação ao modelo anterior, pode-se observar uma predição mais “enxuta” com a adição da variável ALT.

Finalmente, a Figura 4.15(c) apresenta o modelo AMAZON-PD_{PRED04}. Este modelo faz uso das variáveis PAD, ROD, EPA, DEC e ALT, mas somente para os anos-base de 1996 e 2000, uma vez que não havia a variável EPA em datas anteriores a 1996. Diferentemente das duas anteriores, observa-se que o modelo reflete o efeito da variável EPA, que pode ser notado na parte superior da imagem.

Levando-se em conta o que foi apresentado até aqui, é de se esperar que todos os três modelos atendam os requisitos de desempenho pré-estabelecidos ($DM \geq 0,28$; AM e $VP > 70\%$; e $FP < 30\%$), a despeito da impossibilidade de verificação imediata de seus desempenhos (praticável após a geração e a classificação da imagem em 2008).

Além disso, também é de se esperar que, a despeito de ser construído a partir de uma amostragem menor – utiliza somente duas datas – o modelo AMAZON-PD_{PRED04} apresente o melhor desempenho entre os três, em função do uso de uma quantidade maior de variáveis.

Tendo isso em mente, pode-se dizer que o AMAZON-PD apresenta melhor desempenho à medida que as demais variáveis sejam agregadas e que uma maior frequência seja utilizada, como uma amostragem anual, por exemplo. Em caso de se utilizar somente datas posteriores a 1996 (início da variável EPA), ainda assim é possível aplicar o modelo com 12 datas (de 1996 a 2007) para a predição do desmatamento para 2008. Essa configuração deve proporcionar um desempenho ainda melhor para o AMAZON-PD, ano-base 2007.

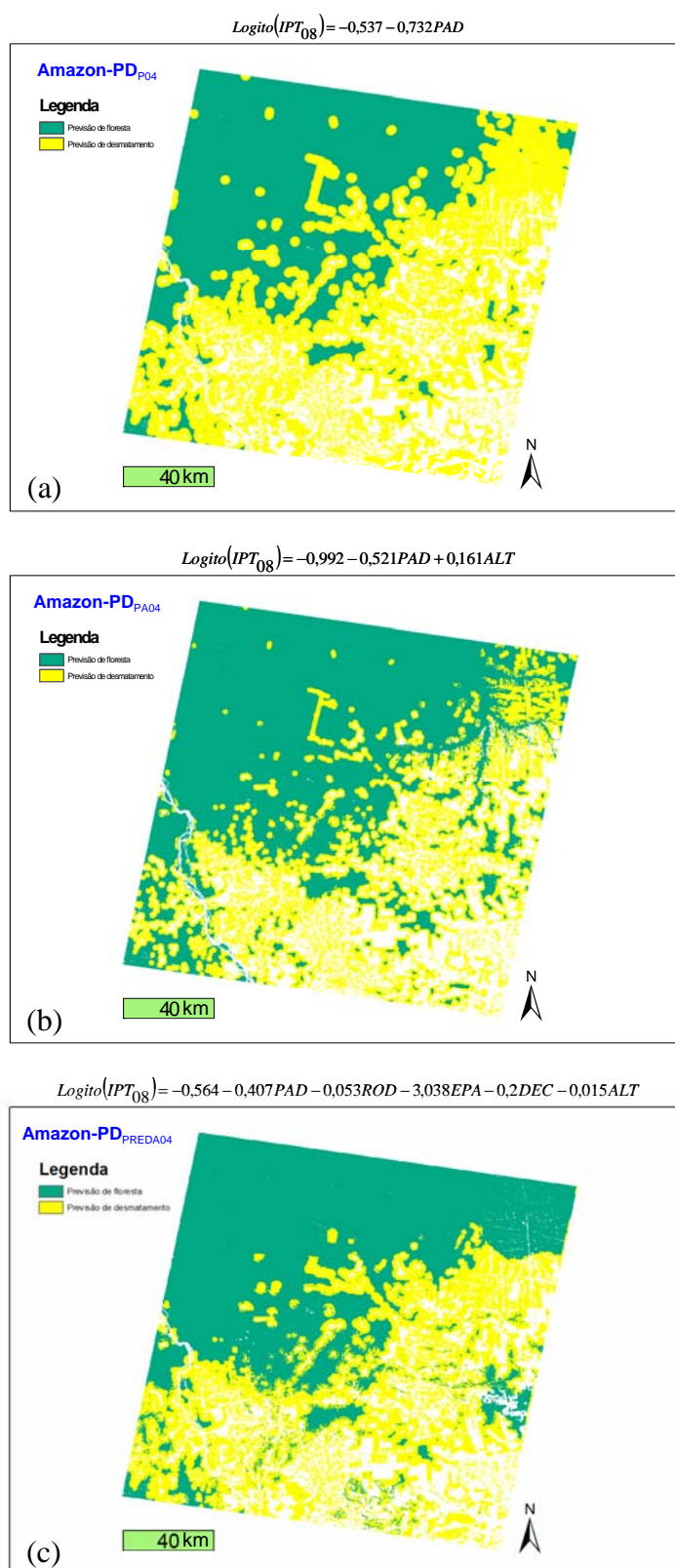


Fig. 4.15 – AMAZON-PD para o ano-base de 2004 (predição de IPT para 2008) com três diferentes configurações: (a) somente com a variável PAD, consideradas todas as datas anteriores – 1985 a 2000; (b) com as variáveis PAD e ALT, considerados os anos-base de melhor desempenho para essas variáveis – 1988, 1996 e 2000; e (c) com as variáveis PAD, ROD, EPA, DEC e ALT, considerado o “aprendizado” proporcionado pelos anos-base de 1996 e 2000.

CONCLUSÕES E RECOMENDAÇÕES

O Brasil possui cerca de 3,6 milhões de km² de florestas e ocupa o terceiro lugar na lista dos países com maior área florestal densa do mundo, fato que o deixa atrás apenas da Rússia e do Canadá. Em termos de florestas tropicais, no entanto, o Brasil é o país com maior cobertura, possuindo uma área três vezes superior à existente na República Democrática do Congo, segundo colocado no *ranking* mundial para esses tipos de florestas (UNEP, 2001).

O aumento dos preços da madeira nos mercados nacional e internacional tornou a exploração madeireira uma atividade atraente. Isso aumentou a quantidade de espécies exploradas. Somem-se a isso as previsões e projeções de exaustão dos recursos florestais do sudeste da Ásia, que hoje ainda suprem em elevada proporção os mercados mundiais de madeiras tropicais, e pode-se ter uma boa idéia da importância que a região amazônica detém, somente nesse contexto.

Outros fatores, além da atração econômica dos preços alcançados pela madeira de lei, provocaram o recrudescimento da exploração madeireira na Amazônia. O governo brasileiro, por exemplo, desempenhou um papel preponderante nesse panorama, tanto em função de iniciativas tomadas, como pela falta delas.

Ainda que as exportações de madeira da Amazônia representem uma parcela modesta do mercado internacional de madeiras tropicais, há uma tendência natural de crescimento desse mercado, em vista da expectativa de escassez de madeiras nobres no futuro próximo.

A se confirmar essa tendência, a exploração madeireira será a principal atividade econômica de uso do solo em toda a Amazônia. Portanto, adequar práticas de manejo à dinâmica dos mercados nacional e internacional será o próximo grande desafio para a Amazônia. Além disso, há que se considerar que a expansão da atividade madeireira já atinge várias áreas de uso restrito, como terras indígenas e unidades de conservação, em particular.

Tão logo perdem seu interesse vinculado ao valor intrínseco das madeiras de lei, as terras amazônicas assumem o interesse provocado pela atividade pecuária extensiva, o que provoca o efeito mais visível de impacto ambiental, que consiste do avanço da frente de desmatamento sobre a floresta.

O controle sobre o meio ambiente em uma região extensa e de difícil acesso como a Amazônia ainda está muito longe de um padrão minimamente aceitável. Ainda que haja denúncias e processos, a falta de meios e de métodos, por parte dos órgãos competentes, induz

as indústrias a permanecerem na ilegalidade, dilapidando os patrimônios natural e nacional existentes na região.

Existem preocupações por parte da sociedade brasileira e isso se verifica na elaboração cuidadosa de leis ambientais. No entanto, a legislação pertinente não passa de uma carta de intenções, com pouco ou nenhum resultado de ordem prática, dada a limitação do estado brasileiro para implementá-la a contento.

Nesse sentido, foi idealizado um modelo capaz de prover uma ferramenta voltada a políticas ambientais de amplo alcance. Em especial, busca-se uma orientação ao emprego de meios de controle de impactos ambientais com base na predição estatística desses impactos.

A opção pelo estudo da cobertura do solo (desmatamento) e não de seu uso, que pode ter diferentes vertentes (extração madeireira, pecuária extensiva e agricultura, por exemplo) evita uma forçosa adoção de variáveis de cunho socioeconômico, o que foge ao escopo do estudo. Não obstante essa busca por uma maior simplicidade em termos de modelagem, deve-se ressaltar que a compreensão do efeito (cobertura do solo = desmatamento), a partir de uma análise de fatores intervenientes (variáveis independentes) que possibilite sua predição em datas futuras, permite lançar um renovado olhar sobre suas causas (uso do solo = extração madeireira ou pecuária ou agricultura).

Em outras palavras, a compreensão dessas causas (mudanças no uso do solo), adquirida pelo vasto conhecimento já acumulado a partir de fatores de ordem socioeconômica e extensamente tratados na literatura, pode ser bastante enriquecida pelo aporte da abordagem aqui proposta.

O uso de um modelo restrito a “apenas” uma cena Landsat (aproximadamente 34000km²) deve-se ao simples fato de que o custo computacional de estendê-lo a toda a região amazônica, que requer um total de 229 imagens para ser coberta pelo satélite Landsat, tornaria o presente estudo inviável, a se considerar a sua concepção pixel-a-pixel aqui adotada.

Ainda assim, foi necessária a reamostragem da cena em células de 4 x 4 pixels, de maneira a permitir o processamento dos mais de 126 modelos com mais de 2,5 milhões de amostras cada. Essa condição ressalta ainda mais a importância do estudo em função da impossibilidade de se realizar um estudo com tantos dados num pacote estatístico convencional. Não menos trabalhosa seria a necessidade de formatação dos dados para a inserção nesses aplicativos.

A agilidade no manuseio dos dados, a compatibilidade das extensões dos arquivos – que tornou desnecessárias transformações nos dados para adequá-los a outros aplicativos – e a otimização do tempo de processamento – com menos tempo perdido na formatação de informações – tornam o AMAZON uma “solução elegante”, na medida em que ele permite usar um único *software* em sua construção.

Por outro lado, são diversas as razões que conferem ao AMAZON uma abordagem inédita na modelagem de cobertura do solo. Uma delas está no fato de que, ao invés de extrapolar o conhecimento obtido de certo número de amostras para todo o universo amostral, por meio de inferências estatísticas, o AMAZON é descritivo estatisticamente, ou seja, reflete o comportamento de todas as amostras, sem exceção.

A consistência dos resultados alcançados com o AMAZON se fundamenta na observação sistemática e de longa duração de uma mesma área por cerca de duas décadas. Essa característica, aliada ao uso conjunto de variáveis categorizadas de exposição, de controle e de interação (multiplicativas), permitiu a obtenção de resultados concretos sobre a dinâmica do desmatamento no município de São Félix do Xingu.

É bastante razoável presumir um comportamento similar para essas variáveis em outros pontos da Amazônia. Mesmo que não venha a ser extrapolável nos parâmetros, ainda assim o AMAZON certamente o será nos procedimentos metodológicos de modelagem.

O AMAZON também ratificou algumas observações do trabalho de campo, no qual se percebeu uma inversão de expectativas quanto ao desmatamento, especialmente quando analisadas as variáveis ‘DEC’ e ‘ALT’. Na medida em que regiões “ideais para desmatamento” escasseiam, o uso de fogo como instrumento de “limpeza” do terreno passa a ser mais freqüente.

A variável ‘ROD’, por sua vez, poderia apresentar uma melhor contribuição, caso as datas da atualização das rodovias e das imagens consideradas fossem as mesmas. Isso foi observado durante a análise desta variável nos modelos quando, em alguns modelos, o ‘ β ’ foi nulo. Assim, recomenda-se que, em estudos futuros da mesma natureza, as rodovias sejam extraídas das próprias imagens usadas para o mapeamento do uso e da cobertura do solo, por meio de técnicas de interpretação de imagens.

A variável ‘EPA’ também não possui necessariamente as mesmas datas das imagens. Além disso, não se sabe a partir de quando as terras indígenas e as UC começaram a ser respeitadas efetivamente. Ainda assim, ‘EPA’ mostrou ser uma variável muito importante, até

mesmo na diferenciação do efeito provocado por diferentes políticas ambientais, dado que as áreas indígenas ratificaram sua maior eficiência como fator de proteção ambiental, em relação aos parques nacionais e às florestas nacionais.

Ao abstrair questões socioeconômicas – importantes, mas de difícil caracterização em uma escala regional – o AMAZON provê palpáveis contribuições científicas à área de política e gestão ambiental. Como o modelo se concentra em variáveis biofísicas, os mapas de tendência gerados a partir do AMAZON permitem uma alocação otimizada dos recursos sempre finitos em relação à área virtualmente infinita que constitui a Amazônia brasileira e cuja conservação se almeja.

Além disso, a aplicação direta de uma técnica de análise multivariada a uma quantidade tão elevada de dados espacialmente explícitos é suficiente para atestar o feito inédito alcançado pelo AMAZON para esse tipo de modelo.

Contudo, é a sua capacidade de aplicar essa técnica ao binômio ‘espaço-tempo’ que lhe emprestam um caráter ainda mais inovador, na medida em que, implementado como AMAZON-PD, o modelo torna efetivamente possível localizar, com razoável precisão, desmatamentos futuros a partir de variáveis biofísicas. Esta característica permite a interpretação dos resultados de maneira direta e inequívoca por tomadores de decisão, sem a necessidade de conhecimento especializado em sensoriamento remoto, estatística ou geoprocessamento.

Embora tenha sido aplicado a uma única cena Landsat (185km x 185km), situada na Amazônia Oriental (São Félix do Xingu-PA), pode-se dizer, com base em suas características conceptuais, que o AMAZON-PD é diretamente aplicável a outras regiões da Amazônia. Mais do que isso, são inúmeras as possibilidades de aplicação do modelo a qualquer outro bioma ou em um contexto de mudança global, aonde outras formas de cobertura do solo venham a ser de interesse.

A razão para isso está no fato de que o AMAZON-PD se utiliza de variáveis existentes em quaisquer das situações mencionadas, como declividade, altura relativa do terreno, índice de fragmentação florestal, área do polígono envolvente, posicionamento relativo a áreas de proteção ambiental conhecidas, proximidade de áreas desmatadas, de rodovias, de hidrovias etc. Além destas, o AMAZON-PD pode incorporar outras variáveis venham a ser compreendidas como relevantes ao processo sob modelagem.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGUIAR, A. P. D. Modelagem de mudanças de uso e cobertura do solo na Amazônia: Questões Gerais. In: **Tutorial de Introdução à Modelagem Dinâmica Espacial**. São José dos Campos: MCT/INPE. 2003.
- ALENCAR, A. A. C.; SOLÓRZANO, L. A.; NEPSTAD, D. C. **Modeling Forest Understory Fires in an Eastern Amazonian Landscape**. *Ecological Applications*, 14(4) Supplement, 2004, pp. S139-S149. Ecological Society of America. 2004.
- ANAS, A. **Residential Location Markets and Urban Transportation: Economic Theory, Econometrics and Policy Analysis with Discrete Choice Models**. Oxford: Academic Press. 1982.
- ARIMA, E. Desmatamento e Economia Local e Políticas Públicas. In: **BRASIL. Ministério do Meio Ambiente. Causas e Dinâmica do Desmatamento na Amazônia**. Brasília: MMA. 2001.
- BATISTA, G. T.; MEDEIROS, J. S.; MELLO, E. M. K.; MOREIRA, J. C.; BINS, L. S. **New Approach for Deforestation Assessment**. *Int. Symp. On Resource and Environmental Monitoring (ISPRS, Rio de Janeiro), Proceedings*, 30, Part 7a, 170-174. 1994.
- BATTEN, D. F.; BOYCE, D. E. Spatial Interaction, Transportation and Interregional Commodity Flow Models. In: **Handbook of Regional and Urban Economics**. Vol. 1. ed. P. Nijkamp, 357-406. Amsterdam: North-Holland. 1986.
- BECKER, B. K. (1999). Cenários de Curto Prazo para o Desenvolvimento da Amazônia. In: PASQUIS, R.; MACHADO, L.; GUERRA, R. **Diagnóstico dos Formatos de Ocupação do Espaço Amazônico**. Brasília: CIRAD/MMA. 2001.
- BECKER, B. K.; MIRANDA, M. H. P.; MACHADO, L. O. **Fronteira Amazônica: Questões Sobre a Gestão do Território**. Brasília: UnB. 1990.
- BOYCE, D.E. Renaissance of Large-Scale Models. **Papers of the Regional Science Association** 65: 1-10. 1988.
- BRIASSOULIS, H. Analysis of Land Use Change: Theoretical and Modeling Approaches. **The Web Book of Regional Science**. 2000. Disponível em: <http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm>. Acesso em: 13 set 2004.
- BÜCKNER, J.; JUNG, S.; PAKZAD, K. **Image Interpretation and GIS analysis as an Approach for Moor Monitoring**. In: International Conference on Dynamic and Multi-Dimensional and GIS DMGIS 99, Pequim, China. **Proceedings** p. 341-349. Pequim, China. 1999.

- CAMARGO, A.; CAPOBIANCO, J. P. R.; OLIVEIRA, J. A. P. (Org.) **Meio ambiente Brasil: Avanços e Obstáculos Pós-Rio-92**. Rio de Janeiro: FGV. 2002.
- CAMPOS, V. **Um Método de Modelagem do Conhecimento Multitemporal para a Interpretação Automática de Imagens de Sensores Remotos**. (PUC-CDD:621.3)172 p. Dissertação (Mestrado em Métodos de Apoio à Decisão) – Departamento de Engenharia Elétrica, PUC-RIO, Rio de Janeiro. 2005.
- CHRISTALLER, W. **Central Places in Southern Germany**. Englewood Cliffs, NJ: Prentice-Hall (original German edition, 1933). 1966.
- COCCOSSIS, H. Historical Land Use Changes: Mediterranean Regions of Europe. In: **Land Use Changes in Europe: Processes of Change, Environmental Transformations and Future Patterns**. eds. BROUWER, F. M.; THOMAS, A. J.; CHADWICK, M. J. 441-462. Dordrecht: Kluwer Academic Publishers. 1991.
- COOKE, P. **Theories of Planning and Spatial Development**. London: Hutchinson. 1983.
- CREPANI, E.; DUARTE, V.; SHIMABUKURO, Y. E. **Sensoriamento Remoto e Geoprocessamento no Mapeamento Regional da Cobertura e Uso Atual da Terra**. São José dos Campos, INPE, 36p. (INPE -8478-NTC/346). 2001.
- DAVIS, L. **Handbook of Genetic Algorithms**, VNR Comp. Library. 1990.
- ESRI. http://webhelp.esri.com/arcgisdesktop/9.2/printBooks_topics.cfm?pid=3042. Acesso em: 1 de agosto de 2005.
- FISCHER, M.; SCHOLTEN, H. J.; UNWIN, D. Geographic Information Systems, Spatial Data Analysis, and Spatial Modeling: An Introduction. In: **Spatial Analytical Perspectives on GIS**, 3-19. London: Taylor and Francis. 1996.
- Food and Agriculture Organization (FAO). **Planning for Sustainable Use of Land Resources**. FAO Land and Water Bulletin 2. Rome: Food and Agriculture Organization of the United Nations. 1995.
- GROWE, S.; SHRÖDER, T.; LIEDTKE, C. E. **Use of Bayesian Networks as Judgement Calculus in a Knowledge Based Image Interpretation System**. In: 19th Congress of the International Society of Photogrammetry and Remote Sensing, julho de 2000, Amsterdam, Holanda. **Proceedings**, v. XXXIII, Amsterdam. 2000.
- HAYNES, K.E.; FOTHERINGHAM, A.S. **Gravity and Spatial Interaction Models**. Beverly Hills Sage. 1984.

- HOOVER, E. M.; GIARRATANI, F. **An Introduction to Regional Economics**, 3rd edition. New York: Alfred Knopf. 1984. In: LOVERIDGE, S. ed. **The Web Book of Regional Science**. 1999. Disponível em: <http://www.rri.wvu.edu/WebBook/Loveridge/contents.htm>. Acesso em: 22 novembro 2004.
- HOSHINO, S. **Statistical Analysis of Land Use Change and Driving Forces in the Kansai District, Japan**. WP-96-120, IIASA. 1996.
- HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**, 2nd Edition, Wiley, New York. 2000.
- JOHNSTON, R. J.; GREGORY, D.; SMITH, D. M. **The Dictionary of Human Geography**, 3rd edition. Oxford: Blackwell. 1994.
- KATES, R. W.; TURNER II, B.L.; CLARK, W. C. The Great Transformation. In: **The Earth as Transformed by Human Action**. eds. TURNER II, B. L.; CLARK, W. C.; KATES, R.W.; RICHARDS, J.E.; MATHEWS, J. T.; MEYER, W. B. 1-17. Cambridge: Cambridge University Press. 1990.
- KAUFMAN, R. L. Comparing Effects in Dichotomous Logistic Regression: A Variety of Standardized Coefficients. **Social Science Quarterly**, 77, 90-109. 1996.
- KLEINBAUM, D. G.; KLEIN, M. **Logistic Regression: A Self-Learning Text**. Springer-Verlag, 2nd edição. New York, NY. 2002.
- LeSAGE, J. Spatial Econometrics. In: **The Web Book of Regional Science**, ed S. Loveridge West Virginia University, Regional Research Institute, Morgantown, WV. 1999.
- LESSCHEN, J. P.; VERBURG, P. H.; STAAL, S. J. **Statistical Methods For analyzing The Spatial Dimension of Changes in Land Use and Farming Systems - LUCC Report Series N° 7**, Wageningen, The Netherlands: Wageningen University. 2005.
- LIVERMAN, D.; MORAN, E. F.; RINDFUSS, R. R.; STEIN, P. C. **People and the Pixels: Linking Remote Sensing and Social Science**. Washington, D. C: National Academy Press. 1998.
- LONG, J. S. **Regression Models for Categorical and Limited Dependent Variables: Analysis and Interpretation**. Thousand Oaks, CA: Sage. 1997
- LONERGAN, S.; PRUDHAM, S. Modeling global Change in an Integrated Framework: A View from the Social Sciences. In: **Changes in Land Use and Land Cover: A Global Perspective**, eds. MEYER, W. B.; TURNER II, B. L., 411-435. New York: John Wiley. 1994.

- LUCC. **Land Use and Land Cover: Implementation Strategy**. IGBP Report 48, IHDP Report 10, Stockholm: IGBP Secretariat. 1999.
- LUCC. **Proceedings**, 1st Data Requirements Workshop Survey of Needs, Gaps and Priorities on Data for Land Use/Land Cover Research. LUCC Report Series No 3. International Biosphere Geosphere Programme/International Human Development Programme, Land Use and Cover Change. Barcelona, 11-14 November 1997. 1997.
- LUTZ, W. The Population Module. In: **Population-Development-Environment: Understanding their Interactions in Mauritius**. Berlin: Springer-Verlag, ed. 1994a.
- MANNING, E. W. Analysis of Land Use Determinants in Support of Sustainable Development. In: **Land Use Changes in Europe**. eds. BROUWER, F. M.; THOMAS, A. J.; CHADWICK, M. J. 485-514. Dordrecht: Kluwer Academic Publishers. 1991.
- MENDEL, J. M. **Fuzzy Logic Systems for Engineering: A Tutorial**, Proceedings of the IEEE, vol. 83, no 3, março. 1995.
- METZGER, J. P. O que é Ecologia de Paisagens? **Biota Neotropica**, volumes 1 e 2, <http://www.biotaneotropica.org.br/v1n12/pt/abstract?thematic-review+BN00701122001>, 2001.
- MEYER, W. B.; TURNER II, B. L. Land-Use/Land-Cover Change: Challenges for Geographers. **Geojournal** 39(3): 237-240. 1996.
- NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; WASSERMAN, W. **Applied Linear Statistical Models**. McGraw-Hill, 3rd Ed. Boston, Massachusetts. 1996.
- PACHECO, M. A. C. **Algoritmos Genéticos: Princípios e aplicações**. ICA: Laboratório de Inteligência Computacional Aplicada. Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro. http://www.ica.ele.puc-rio.br/inteligencia_computacional/index.html Acesso em 22-4-06.
- PAMPEL, F. C. Logistic Regression: A Primer. **Sage University Papers Series on Quantitative Applications in the Social Sciences**. 07-132. Thousand Oaks, CA: Sage. 2000.
- PASQUIS, R.; MACHADO, L.; GUERRA, R. Diagnóstico dos Formatos de Ocupação do Espaço Amazônico. Brasília: CIRAD/MMA, 2001.
- PONTIUS Jr, R. G.; SCHNEIDER, L. C. **Land-cover Change Model Validation by an ROC Method for the Ipswich Watershed, Massachusetts, USA**. Disponível em www.elsevier.com/locate/agee.

- ROBERTS, D. A.; NUMATA, I.; HOLMES, K.; BATISTA, G.; KRUG, T.; MONTEIRO, A.; POWELL, B.; CHADWICK, O. A Large Area Mapping of Land-Cover Change in Rondônia Using Multitemporal Spectral Mixture Analysis and Decision Tree Classifiers. **Journal of Geophysical Research**, vol. 107, No D20, 8073, doi:10.1029/2001JD000374, 2002.
- SANTOS, J. R. Métodos de tratamento de dados orbitais para estimativa do desmatamento da Amazônia brasileira: a experiência do INPE. **Aplicações do sensoriamento remoto e sistemas de informação geográfica no monitoramento e controle do desmatamento na Amazônia brasileira**. Brasília: MMA/SCA/PPG7, 2002, 17-21.
- SAWYER, D. Evolução Demográfica, Qualidade de Vida e Desmatamento na Amazônia. In: **BRASIL. Ministério do Meio Ambiente. Causas e Dinâmica do Desmatamento na Amazônia**. Brasília: MMA. 2001.
- SAYER, A. A Critique of Urban and Regional Modeling. **Progress in Planning** 6 (3): 191-254. 1976.
- SAYER, A. Understanding Urban Models Versus Understanding Cities. **Environment and Planning A** 11:853-862. 1979a.
- SAYER, A. Philosophical Bases of the Critique of Urban Modeling: A Reply to Wilson. **Environment and Planning A** 11: 1055-1069. 1979 b.
- SAYER, A. **Method in Social Science: A Realist Approach**. London: Hutchinson. 1984.
- SHIMABUKURO, Y. E.; SMITH, J. A. The Least-Squares Mixing Models to Generate Fraction Images Derived From Remote Sensing Multispectral Data. **IEEE Transactions on Geoscience and Remote Sensing**, vol. 29. No 1. January. 1991.
- SHIMABUKURO, Y. E.; BATISTA, G. T.; MELLO, E. M. K.; MOREIRA, J. C.; DUARTE, V. Using Shade Fraction Image Segmentation to Evaluate Deforestation in Landsat Thematic Mapper Images of the Amazon Region. **International Journal of Remote Sensing**, vol. 21. No 3. 535-541. January. 1998.
- SILVA, A. N. **Análise das Relações entre o Desflorestamento e o Potencial Agropecuário das Terras da Amazônia**. Dissertação de Mestrado em Sensoriamento Remoto, INPE, São José dos Campos, SP. 2006.
- SILVA, J. C. Re: Origin of term “logistic regression”. Disponível em <http://mathforum.org/kb/thread.jspa?forumID=193&threadID=438337&messageID=1383550#1383550> (2003) Acesso em 20 de maio de 2005.

- SILVA, J. R. P.; TOZZI, C. L. **Um Algoritmo Geométrico para a Geração de Funções Discriminantes.** In: IX Simpósio Brasileiro de Computação Gráfica e Processamento De Imagens. Anais IX Simpósio Brasileiro de Computação Gráfica e Processamento De Imagens. Caxambú, MG p.31-38. 1996.
- SIMON, H. **Models of Bounded Rationality.** Cambridge, MA.: The MIT Press. 1982.
- SLOCOMBE, D. S. Environmental Planning, Ecosystem Science, and Ecosystem Approaches for Integrating Environment and Development. **Environmental Management** 17: 288-390. 1993.
- SOARES-FILHO, B. S.; CERQUEIRA, G. C.; PENNACHIN, C. L. **DINAMICA – A Stochastic Cellular Automata Model Designed to Stimulate the Landscape Dynamics in an Amazonian Colonization Frontier.** Disponível em www.elsevier.com/locate/ecolmodel. Acesso em 24-10-06./
- TOU, J.T.; GONZALEZ, R. C. **Pattern Recognition Principles.** Addison-Wesley Publishing Company, Reading, Massachussets. 1974.
- TURNER, B. L. II; MEYER, B. L.; Global Land Use and Land Cover Change: An Overview. In: **Changes in Land Use and Land Cover: A Global Perspective,** eds MEYER, W.B.; TURNER II, B. L., 3-10. Cambridge: Cambridge University Press. 1994.
- TURNER, B. L. II; SKOLE, D.; SANDERSON, S.; FISCHER, G.; FRESCO, L.; LEEMANS, R. **Land Use and Land Cover; Science/Research Plan.** IGBP report no.35, HDP Report No 7. IGBP e HDP, Stockholm e Geneva. 1995.
- WEGENER, M. Integrated Forecasting Models of Urban and Regional Systems. In: **Integrated Analysis of Regional Systems,** eds. BATEY, P. W. J; MADDEN, M. 9-24. London: Pion. 1986b.
- WEGENER, M. Operational Urban Models: State-of-the-Art. **Journal of the American Planning Association** 60(1): 17-29. 1994.
- WOLMAN, M. G. Criteria for Land Use. In: **Resources and World Development,** eds. MCLAREN, D. J.; SKINNER, B.J., 643-657. New York: John Wiley. 1987.
- ZADEH, L. A. **Fuzzy Sets as a Basis for a Theory of Possibility.** Fuzzy Sets and Systems, 1:3-28. 1978.

ANEXOS

ANEXO I – MODELOS PRESERVADOS PARA O ANO-BASE DE 1985

ANEXO II – MODELOS PRESERVADOS PARA O ANO-BASE DE 1988

ANEXO III – MODELOS PRESERVADOS PARA O ANO-BASE DE 1992

ANEXO IV – MODELOS PRESERVADOS PARA O ANO-BASE DE 1996

ANEXO V – MODELOS PRESERVADOS PARA O ANO-BASE DE 2000

Observações:

Embora tenham sido gerados vários modelos para cada ano-base (14 de 1985 a 1992 e 28 de 1996 a 2000), somente foram preservados aqueles que apresentaram valores de DM superiores a 0,28⁵⁰. Entretanto, em alguns casos, esse valor foi alcançado sem que o critério fosse atingido (modelos cujos títulos estão marcados na cor verde-claro). Os demais têm seus cabeçalhos marcados na cor azul, sendo que o de melhor desempenho entre eles – realçado por um tom de azul-escuro – foi o modelo utilizado na implementação dos modelos AMAZON de cada ano-base, apresentados no corpo da tese, Capítulo 4.

As letras utilizadas para identificar os modelos preservados em seus respectivos anos-base nas diversas tabelas de classificação seguem a seguinte convenção:

P = Proximidade da Área Desmatada (PAD)

R = Proximidade de Rodovias (ROD)

E = Existência de roteção Ambiental (EPA)

D = Declividade (DEC)

A = Altura (ALT)

W1 = Primeira Variável Interativa (PAD x DEC)

W2 = Segunda Variável Interativa (PAD x ALT)

⁵⁰ - Valor mínimo necessário para satisfazer o critério pré-estabelecido de AM e VP > 70% e FP < 30%, uma vez que $DM = AM * (VP - FP)$.

ANEXO I – MODELOS PRESERVADOS PARA O ANO-BASE DE 1985

P85			DM = 0,363
2700944	DTO obs	FLT obs	
DTO esp	23719	410972	
FLT esp	16806	2249447	
ajuste do modelo	84,16%	58,53%	15,45%
	verdadeiros positivos	falsos positivos	

PD85			DM = 0,366
2700944	DTO obs	FLT obs	
DTO esp	21593	308734	
FLT esp	18932	2351685	
ajuste do modelo	87,87%	53,28%	11,60%
	verdadeiros positivos	falsos positivos	

PA85			DM = 0,320
2700076	DTO obs	FLT obs	
DTO esp	18302	254463	
FLT esp	22204	2405107	
ajuste do modelo	89,75%	45,18%	9,57%
	verdadeiros positivos	falsos positivos	

PDA85			DM = 0,323
2700944	DTO obs	FLT obs	
DTO esp	18021	234273	
FLT esp	22504	2426146	
ajuste do modelo	90,49%	44,47%	8,81%
	verdadeiros positivos	falsos positivos	

PDW185			DM = 0,366
2700944	DTO obs	FLT obs	
DTO esp	21593	308734	
FLT esp	18932	2351685	
ajuste do modelo	87,87%	53,28%	11,60%
	verdadeiros positivos	falsos positivos	

PAW285			DM = 0,333
2700944	DTO obs	FLT obs	
DTO esp	19402	278554	
FLT esp	21123	2381865	
ajuste do modelo	88,90%	47,88%	10,47%
	verdadeiros positivos	falsos positivos	

PDAW1W285			DM = 0,286
2700944	DTO obs	FLT obs	
DTO esp	14036	117483	
FLT esp	26489	2542936	
ajuste do modelo	94,67%	34,64%	4,42%
	verdadeiros positivos	falsos positivos	

Como nenhum dos modelos deste ano-base atingiu o critérios pré-estabelecido de VP > 70%, o modelo escolhido foi o PD85, que obteve o maior valor de DM para esta data.

ANEXO II – MODELOS PRESERVADOS PARA O ANO-BASE DE 1988

P88			DM = 0,451
2660445	DTO obs	FLT obs	
DTO esp	29875	483302	
FLT esp	10559	2136709	
ajuste do modelo	81,44%	73,89%	18,45%
	verdadeiros positivos	falsos positivos	

PR88			DM = 0,451
2660445	DTO obs	FLT obs	
DTO esp	29875	483302	
FLT esp	10559	2136709	
ajuste do modelo	81,44%	73,89%	18,45%
	verdadeiros positivos	falsos positivos	

PD88			DM = 0,460
2660445	DTO obs	FLT obs	
DTO esp	28439	410554	
FLT esp	11995	2209457	
ajuste do modelo	84,12%	70,33%	15,67%
	verdadeiros positivos	falsos positivos	

PA88			DM = 0,450
2660445	DTO obs	FLT obs	
DTO esp	27718	400293	
FLT esp	12716	2219718	
ajuste do modelo	84,48%	68,55%	15,28%
	verdadeiros positivos	falsos positivos	

PRA88			DM = 0,451
2660445	DTO obs	FLT obs	
DTO esp	27713	399011	
FLT esp	12721	2221000	
ajuste do modelo	84,52%	68,54%	15,23%
	verdadeiros positivos	falsos positivos	

PDA88			DM = 0,453
2660445	DTO obs	FLT obs	
DTO esp	27730	395536	
FLT esp	12704	2224475	
ajuste do modelo	84,66%	68,58%	15,10%
	verdadeiros positivos	falsos positivos	

PDW188			DM = 0,453
2660445	DTO obs	FLT obs	
DTO esp	27728	395477	
FLT esp	12706	2224534	
ajuste do modelo	84,66%	68,58%	15,09%
	verdadeiros positivos	falsos positivos	

PAW288			DM = 0,453
2660445	DTO obs	FLT obs	
DTO esp	29889	481150	
FLT esp	10545	2138861	
ajuste do modelo	81,52%	73,92%	18,36%
	verdadeiros positivos	falsos positivos	

PRAW288			DM = 0,456
2660445	DTO obs	FLT obs	
DTO esp	29853	473213	
FLT esp	10581	2146798	
ajuste do modelo	81,82%	73,83%	18,06%
	verdadeiros positivos	falsos positivos	

ANEXO III – MODELOS PRESERVADOS PARA O ANO-BASE DE 1992

P92		DM = 0,405	
2620011	DTO obs	FLT obs	
DTO esp	91487	547118	
FLT esp	32229	1949177	
ajuste do modelo	77,89%	73,95%	21,92%
	verdadeiros positivos	falsos positivos	

PR92		DM = 0,415	
2620011	DTO obs	FLT obs	
DTO esp	85570	452139	
FLT esp	38146	2044156	
ajuste do modelo	81,29%	69,17%	18,11%
	verdadeiros positivos	falsos positivos	

PD92		DM = 0,406	
2620011	DTO obs	FLT obs	
DTO esp	90492	532301	
FLT esp	33224	1963994	
ajuste do modelo	78,42%	73,14%	21,32%
	verdadeiros positivos	falsos positivos	

PA92		DM = 0,410	
2620011	DTO obs	FLT obs	
DTO esp	90351	523715	
FLT esp	33365	1972580	
ajuste do modelo	78,74%	73,03%	20,98%
	verdadeiros positivos	falsos positivos	

PRD92		DM = 0,410	
2620011	DTO obs	FLT obs	
DTO esp	81400	407350	
FLT esp	42316	2088945	
ajuste do modelo	82,84%	65,80%	16,32%
	verdadeiros positivos	falsos positivos	

PRA92		DM = 0,375	
2620011	DTO obs	FLT obs	
DTO esp	67617	286692	
FLT esp	56099	2209603	
ajuste do modelo	86,92%	54,66%	11,48%
	verdadeiros positivos	falsos positivos	

PDA92		DM = 0,408	
2620011	DTO obs	FLT obs	
DTO esp	89500	516676	
FLT esp	34216	1979619	
ajuste do modelo	78,97%	72,34%	20,70%
	verdadeiros positivos	falsos positivos	

PRDA92		DM = 0,383	
2620011	DTO obs	FLT obs	
DTO esp	69952	303499	
FLT esp	53764	2192796	
ajuste do modelo	86,36%	56,54%	12,16%
	verdadeiros positivos	falsos positivos	

PDW192		DM = 0,407	
2620011	DTO obs	FLT obs	
DTO esp	90549	532574	
FLT esp	33167	1963721	
ajuste do modelo	78,41%	73,19%	21,33%
	verdadeiros positivos	falsos positivos	

PAW292		DM = 0,405	
2620011	DTO obs	FLT obs	
DTO esp	72501	297749	
FLT esp	51215	2198546	
ajuste do modelo	86,68%	58,60%	11,93%
	verdadeiros positivos	falsos positivos	

PRDW192		DM = 0,395	
2573411	DTO obs	FLT obs	
DTO esp	69346	380927	
FLT esp	41229	2081909	
ajuste do modelo	83,60%	62,71%	15,47%
	verdadeiros positivos	falsos positivos	

PRAW292		DM = 0,418	
2620011	DTO obs	FLT obs	
DTO esp	80823	385037	
FLT esp	42893	2111258	
ajuste do modelo	83,67%	65,33%	15,42%
	verdadeiros positivos	falsos positivos	

PDAW1W292		DM = 0,361	
2620011	DTO obs	FLT obs	
DTO esp	62443	240733	
FLT esp	61273	2255562	
ajuste do modelo	88,47%	50,47%	9,64%
	verdadeiros positivos	falsos positivos	

PRDAW1W292		DM = 0,406	
2620011	DTO obs	FLT obs	
DTO esp	76195	345667	
FLT esp	47521	2150628	
ajuste do modelo	84,99%	61,59%	13,85%
	verdadeiros positivos	falsos positivos	

ANEXO IV – MODELOS PRESERVADOS PARA O ANO-BASE DE 1996

P96			DM = 0,321
2496295	DTO obs	FLT obs	
DTO esp	95960	917842	
FLT esp	11286	1471207	
ajuste do modelo	62,78%	89,48%	38,42%
	verdadeiros positivos		falsos positivos

PR96			DM = 0,358
2496295	DTO obs	FLT obs	
DTO esp	72187	513024	
FLT esp	35059	1876025	
ajuste do modelo	78,04%	67,31%	21,47%
	verdadeiros positivos		falsos positivos

PE96			DM = 0,381
2496295	DTO obs	FLT obs	
DTO esp	95742	790872	
FLT esp	11504	1598177	
ajuste do modelo	67,86%	89,27%	33,10%
	verdadeiros positivos		falsos positivos

PD96			DM = 0,329
2496295	DTO obs	FLT obs	
DTO esp	92012	847264	
FLT esp	15234	1541785	
ajuste do modelo	65,45%	85,80%	35,46%
	verdadeiros positivos		falsos positivos

PA96			DM = 0,321
2496295	DTO obs	FLT obs	
DTO esp	95960	917842	
FLT esp	11286	1471207	
ajuste do modelo	62,78%	89,48%	38,42%
	verdadeiros positivos		falsos positivos

PRE96			DM = 0,390
2496295	DTO obs	FLT obs	
DTO esp	79759	558317	
FLT esp	27487	1830732	
ajuste do modelo	76,53%	74,37%	23,37%
	verdadeiros positivos		falsos positivos

PRD96			DM = 0,357
2496295	DTO obs	FLT obs	
DTO esp	67871	451527	
FLT esp	39375	1937522	
ajuste do modelo	80,33%	63,29%	18,90%
	verdadeiros positivos		falsos positivos

PRA96			DM = 0,339
2496295	DTO obs	FLT obs	
DTO esp	60341	371552	
FLT esp	46905	2017497	
ajuste do modelo	83,24%	56,26%	15,55%
	verdadeiros positivos		falsos positivos

PED96			DM = 0,393
2496295	DTO obs	FLT obs	
DTO esp	91272	710237	
FLT esp	15974	1678812	
ajuste do modelo	70,91%	85,11%	29,73%
	verdadeiros positivos		falsos positivos

PEA96			DM = 0,389
2496295	DTO obs	FLT obs	
DTO esp	76863	520639	
FLT esp	30383	1868410	
ajuste do modelo	77,93%	71,67%	21,79%
	verdadeiros positivos		falsos positivos

PDA96			DM = 0,348
2496292	DTO obs	FLT obs	
DTO esp	72078	529914	
FLT esp	35168	1859132	
ajuste do modelo	77,36%	67,21%	22,18%
	verdadeiros positivos		falsos positivos

PRED96			DM = 0,393
2496295	DTO obs	FLT obs	
DTO esp	75230	489132	
FLT esp	32016	1899917	
ajuste do modelo	79,12%	70,15%	20,47%
	verdadeiros positivos		falsos positivos

PREA96			DM = 0,363
2496295	DTO obs	FLT obs	
DTO esp	64113	382703	
FLT esp	43133	2006346	
ajuste do modelo	82,94%	59,78%	16,02%
	verdadeiros positivos		falsos positivos

PEDA96			DM = 0,394
2496295	DTO obs	FLT obs	
DTO esp	91173	706930	
FLT esp	16073	1682119	
ajuste do modelo	71,04%	85,01%	29,59%
	verdadeiros positivos		falsos positivos

PRDA96			DM = 0,350
2496295	DTO obs	FLT obs	
DTO esp	64519	414404	
FLT esp	42727	1974645	
ajuste do modelo	81,69%	60,16%	17,35%
	verdadeiros positivos		falsos positivos

PREDA96			DM = 0,372
2495794	DTO obs	FLT obs	
DTO esp	66828	408080	
FLT esp	40259	1980627	
ajuste do modelo	82,04%	62,41%	17,08%
	verdadeiros positivos		falsos positivos

PDW196			DM = 0,329
2496295	DTO obs	FLT obs	
DTO esp	92012	847264	
FLT esp	15234	1541785	
ajuste do modelo	65,45%	85,80%	35,46%
	verdadeiros positivos		falsos positivos

PAW296			DM = 0,355
2496295	DTO obs	FLT obs	
DTO esp	76952	586389	
FLT esp	30294	1802660	
ajuste do modelo	75,30%	71,75%	24,54%
	verdadeiros positivos		falsos positivos

PRDW196			DM = 0,357
2496295	DTO obs	FLT obs	
DTO esp	67914	452096	
FLT esp	39332	1936953	
ajuste do modelo	80,31%	63,33%	18,92%
	verdadeiros positivos		falsos positivos

PRAW296			DM = 0,349
2496295	DTO obs	FLT obs	
DTO esp	64659	418229	
FLT esp	42587	1970820	
ajuste do modelo	81,54%	60,29%	17,51%
	verdadeiros positivos		falsos positivos

PEDW196			DM = 0,393
2496295	DTO obs	FLT obs	
DTO esp	91272	710237	
FLT esp	15974	1678812	
ajuste do modelo	70,91%	85,11%	29,73%
	verdadeiros positivos		falsos positivos

PEAW296			DM = 0,388
2496295	DTO obs	FLT obs	
DTO esp	76863	521515	
FLT esp	30383	1867534	
ajuste do modelo	77,89%	71,67%	21,83%
	verdadeiros positivos		falsos positivos

PDAW1W296			DM = 0,296
2496295	DTO obs	FLT obs	
DTO esp	57927	419013	
FLT esp	49319	1970036	
ajuste do modelo	81,24%	54,01%	17,54%
	verdadeiros positivos		falsos positivos

PREDW1			DM = 0,393
2496295	DTO obs	FLT obs	
DTO esp	75523	492518	
FLT esp	31723	1896531	
ajuste do modelo	79,00%	70,42%	20,62%
	verdadeiros positivos		falsos positivos

PREAW296			DM = 0,371
2496295	DTO obs	FLT obs	
DTO esp	67886	423650	
FLT esp	39360	1965399	
ajuste do modelo	81,45%	63,30%	17,73%
	verdadeiros positivos		falsos positivos

PRDAW1W296			DM = 0,358
2496295	DTO obs	FLT obs	
DTO esp	67511	443100	
FLT esp	39735	1945949	
ajuste do modelo	80,66%	62,95%	18,55%
	verdadeiros positivos		falsos positivos

PEDAW1W296			DM = 0,391
2496295	DTO obs	FLT obs	
DTO esp	76316	509099	
FLT esp	30930	1879950	
ajuste do modelo	78,37%	71,16%	21,31%
	verdadeiros positivos		falsos positivos

PREDAW1W296			DM = 0,382
2496295	DTO obs	FLT obs	
DTO esp	71008	448197	
FLT esp	36238	1940852	
ajuste do modelo	80,59%	66,21%	18,76%
	verdadeiros positivos		falsos positivos

ANEXO V – MODELOS PRESERVADOS PARA O ANO-BASE DE 2000

P00			DM = 0,254
2387747	DTO obs	FLT obs	
DTO esp	210723	1018541	
FLT esp	18229	1140254	
ajuste do modelo	56,58%	92,04%	47,18%
	verdadeiros positivos	falsos positivos	

PR00			DM = 0,326
2387747	DTO obs	FLT obs	
DTO esp	200182	805886	
FLT esp	28770	1352909	
ajuste do modelo	65,04%	87,43%	37,33%
	verdadeiros positivos	falsos positivos	

PE00			DM = 0,322
2387747	DTO obs	FLT obs	
DTO esp	206027	848250	
FLT esp	22925	1310545	
ajuste do modelo	63,51%	89,99%	39,29%
	verdadeiros positivos	falsos positivos	

PD00			DM = 0,265
2387747	DTO obs	FLT obs	
DTO esp	209265	985115	
FLT esp	19687	1173680	
ajuste do modelo	57,92%	91,40%	45,63%
	verdadeiros positivos	falsos positivos	

PA00			DM = 0,341
2387747	DTO obs	FLT obs	
DTO esp	171713	602025	
FLT esp	57239	1556770	
ajuste do modelo	72,39%	75,00%	27,89%
	verdadeiros positivos	falsos positivos	

PRE00			DM = 0,369
2387747	DTO obs	FLT obs	
DTO esp	200211	723758	
FLT esp	28741	1435037	
ajuste do modelo	68,48%	87,45%	33,53%
	verdadeiros positivos	falsos positivos	

PRD00			DM = 0,331
2387747	DTO obs	FLT obs	
DTO esp	197455	779221	
FLT esp	31497	1379574	
ajuste do modelo	66,05%	86,24%	36,10%
	verdadeiros positivos	falsos positivos	

PRA00			DM = 0,329
2387747	DTO obs	FLT obs	
DTO esp	195555	771156	
FLT esp	33397	1387639	
ajuste do modelo	66,30%	85,41%	35,72%
	verdadeiros positivos	falsos positivos	

PED00			DM = 0,335
2387747	DTO obs	FLT obs	
DTO esp	204349	812335	
FLT esp	24603	1346460	
ajuste do modelo	64,95%	89,25%	37,63%
	verdadeiros positivos	falsos positivos	

PEA00			DM = 0,352
2387747	DTO obs	FLT obs	
DTO esp	198590	746659	
FLT esp	30362	1412136	
ajuste do modelo	67,46%	86,74%	34,59%
	verdadeiros positivos	falsos positivos	

PRED00			DM = 0,377
2387747	DTO obs	FLT obs	
DTO esp	196555	687568	
FLT esp	32397	1471227	
ajuste do modelo	69,85%	85,85%	31,85%
	verdadeiros positivos	falsos positivos	

PREA00			DM = 0,371
2387747	DTO obs	FLT obs	
DTO esp	198889	711880	
FLT esp	30063	1446915	
ajuste do modelo	68,93%	86,87%	32,98%
	verdadeiros positivos	falsos positivos	

PEDA00			DM = 0,340
2387747	DTO obs	FLT obs	
DTO esp	138663	390000	
FLT esp	90289	11768795	
ajuste do modelo	79,89%	60,56%	18,07%
	verdadeiros positivos	falsos positivos	

PRDA00			DM = 0,311
2387747	DTO obs	FLT obs	
DTO esp	138169	440009	
FLT esp	90783	1718786	
ajuste do modelo	77,77%	60,35%	20,38%
	verdadeiros positivos	falsos positivos	

PREDA00			DM = 0,370
2387747	DTO obs	FLT obs	
DTO esp	170889	549393	
FLT esp	58063	1614862	
ajuste do modelo	74,79%	74,64%	25,20%
	verdadeiros positivos	falsos positivos	

PDW100			DM = 0,265
2387747	DTO obs	FLT obs	
DTO esp	209279	985260	
FLT esp	19673	1173535	
ajuste do modelo	57,91%	91,41%	45,64%
	verdadeiros positivos	falsos positivos	

PAW200			DM = 0,341
2387747	DTO obs	FLT obs	
DTO esp	171713	601983	
FLT esp	57239	1556812	
ajuste do modelo	72,39%	75,00%	27,89%
	verdadeiros positivos	falsos positivos	

PRDW100			DM = 0,331
2387747	DTO obs	FLT obs	
DTO esp	197526	780247	
FLT esp	31426	1378548	
ajuste do modelo	66,01%	86,27%	36,14%
	verdadeiros positivos	falsos positivos	

PRAW200			DM = 0,331
2387747	DTO obs	FLT obs	
DTO esp	190727	738826	
FLT esp	38225	1419969	
ajuste do modelo	67,46%	83,30%	34,22%
	verdadeiros positivos	falsos positivos	

PEDW100			DM = 0,335
2387747	DTO obs	FLT obs	
DTO esp	204349	812335	
FLT esp	24603	1346460	
ajuste do modelo	64,95%	89,25%	37,63%
	verdadeiros positivos	falsos positivos	

PEAW200			DM = 0,364
2387747	DTO obs	FLT obs	
DTO esp	197125	715378	
FLT esp	31827	1443417	
ajuste do modelo	68,71%	86,10%	33,14%
	verdadeiros positivos	falsos positivos	

PREDW100			DM = 0,378
2387747	DTO obs	FLT obs	
DTO esp	196788	688244	
FLT esp	32164	1470551	
ajuste do modelo	69,83%	85,95%	31,88%
	verdadeiros positivos	falsos positivos	

PREAW200			DM = 0,371
2387747	DTO obs	FLT obs	
DTO esp	199637	717608	
FLT esp	29315	1441187	
ajuste do modelo	68,72%	87,20%	33,24%
	verdadeiros positivos	falsos positivos	

PRDAW1W200			DM = 0,303
2387747	DTO obs	FLT obs	
DTO esp	133628	424474	
FLT esp	95324	1734321	
ajuste do modelo	78,23%	58,37%	19,66%
	verdadeiros positivos	falsos positivos	

PEDAW1W200			DM = 0,308
2387747	DTO obs	FLT obs	
DTO esp	124691	352364	
FLT esp	104261	1806431	
ajuste do modelo	80,88%	54,46%	16,32%
	verdadeiros positivos	falsos positivos	

PREDAW1W200			DM = 0,361
2387747	DTO obs	FLT obs	
DTO esp	160459	493440	
FLT esp	68493	1665355	
ajuste do modelo	76,47%	70,08%	22,86%
	verdadeiros positivos	falsos positivos	