

Análise multivariada aplicada a zoneamento para a predição de desmatamento: o método AMAZON-PD

Darcton Policarpo Damião ¹
Newton Moreira de Sousa ²

¹Instituto de Estudos Avançados – IEAv
Rod. Tamoios, km 5,5 – 12.228-840 – São José dos Campos – SP, Brasil
darcton@ieav.cta.br

²Universidade de Brasília – UnB
Campus Darcy Ribeiro – Brasília – DF, Brasil
nmsousa@unb.br

Abstract. This paper, as well as its predecessor – “Análise multivariada aplicada a zoneamento: o método AMAZON”, presents the main results of a doctorate thesis (University of Brasilia) on how predictable is the Amazon deforestation throughout multivariate analysis techniques, more specifically a logistic regression model. Likewise it was done within the AMAZON method, the data available for AMAZON-PD were divided in four categories – forest, non-forest, hydrograph, and deforestation – taking into account the methodological procedures developed to PRODES DIGITAL by the Brazilian National Institute for Space Research (INPE). This time the AMAZON method is pushed forward and becomes AMAZON-PD (“Multivariate Analysis Applied to Zoning for Deforestation Prediction”), going from simply understanding the independent variables effects on the response-variable (environmental impact = deforestation), in a static fashion, to a dynamic way of treating the deforestation phenomenon. The attained results, which accuracy was verified upon classification tables, allowed inferring the future location of deforestation sites.

Palavras-chave: deforestation, remote sensing, geographical information systems, multivariate analysis, logistic regression, land cover change, desmatamento, sensoriamento remoto, sistemas de informação geográfica, análise multivariada, regressão logística, mudanças na cobertura do solo.

1. Introdução

No artigo intitulado “Análise multivariada aplicada a zoneamento: o método AMAZON” (DAMIÃO e SOUZA, *no prelo*), foi possível concluir que, 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.

Ainda que as conclusões obtidas a partir dos diversos modelos gerados ano a ano tenham servido de base de entendimento e de validação das variáveis escolhidas para a formulação do AMAZON, a abordagem estática inerente ao método não torna possível atender a proposta original que motivou o trabalho – predição da atividade de desmatamento para datas futuras. A razão para tal está no fato de que ‘IPT’, obtida pela classificação ‘floresta x desmatamento’ da data subsequente à do ano-base ao qual se aplica o modelo, não está disponível nesses casos.

Assim, a forma idealizada para solucionar esse problema foi a análise do comportamento dos coeficientes (β) 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 (DAMIÃO, 2007).

2. Procedimentos Metodológicos

2.1 AMAZON – o modelo estático

Em sua concepção, a eficiência do AMAZON é medida por meio de 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 modelo¹ (AM), de verdadeiros positivos² (VP) e de falsos positivos³ (FP), que permitem escolher o modelo preservado para o ano-base em questão, com base no desempenho do modelo⁴ (DM), cujos critérios são tais que: $AM \geq 70\%$; $VP \geq 70\%$; e $FP < 30\%$.

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’).

Para a formulação do AMAZON, são aplicadas as conclusões obtidas nos diversos modelos gerados ano a ano, que servem de base de entendimento e de validação com relação às variáveis escolhidas. Dessa forma, obtém-se a variável dependente ‘IPT’ por meio da classificação ‘floresta x desmatamento’ na data subsequente à do ano-base ao qual se aplica o modelo.

Embora possa permitir um melhor entendimento do papel de cada uma das variáveis preservadas do modelo, tal abordagem não torna possível prever a 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, o que permitiu uma evolução conceitual para o desafio de se prever áreas de desmatamento futuro.

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. Em função da clara tendência observada na variação dos valores de β_0 e β_1 , foi possível estimar esses coeficientes para 2000, por meio de uma regressão linear destes coeficientes (β_0 e β_1) nos anos-base de 1985, 1988, 1992 e 1996.

O modelo resultante, denominado de ‘Análise Multivariada Aplicada a Zoneamento para Predição de Desmatamento (AMAZON-PD⁵)’, gerou o mapa de tendências apresentado na Figura 1 (AMAZON-PD_{P00} - o índice ‘_{P00}’ significa que o modelo faz uso somente da variável ‘PAD’ e é referente ao ano-base ‘2000’), cuja tabela de classificação apresenta resultados ainda melhores do que aqueles alcançados para o mesmo ano-base com o AMAZON_{P00}. Destaca-se a diferença de se utilizar o conhecimento prévio da variável dependente ‘IPT₀₄’, uma vez que o modelo passou a contar com o “aprendizado” das cenas de datas anteriores.

¹ - Percentual resultante do somatório das predições corretas para as classes ‘floresta’ e ‘desmatamento’, dividido pelo total de células da máscara analisada.

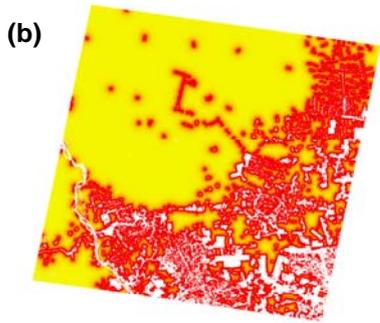
² - Percentual resultante da divisão do número de células corretamente preditas para a classe ‘desmatamento’ pelo número total de células dessa classe.

³ - Percentual resultante da divisão do número de células erroneamente preditas para a classe ‘desmatamento’ pelo número total de células da classe ‘floresta’.

⁴ - $DM = AM \times (VP - FP)$

⁵ - 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).

$$(a) \quad \text{Logito}(IPT_{04}) = -0,602 - 0,879PAD_{00}$$



(f) **AMAZON-PD_{P00}** **DM = 0,348**

2387747	DTO obs	FLT obs
DTO esp	170700	581872
FLT esp	58252	1576923
73,19%	74,56%	26,95%
ajuste do modelo	verdadeiros positivos	falsos positivos

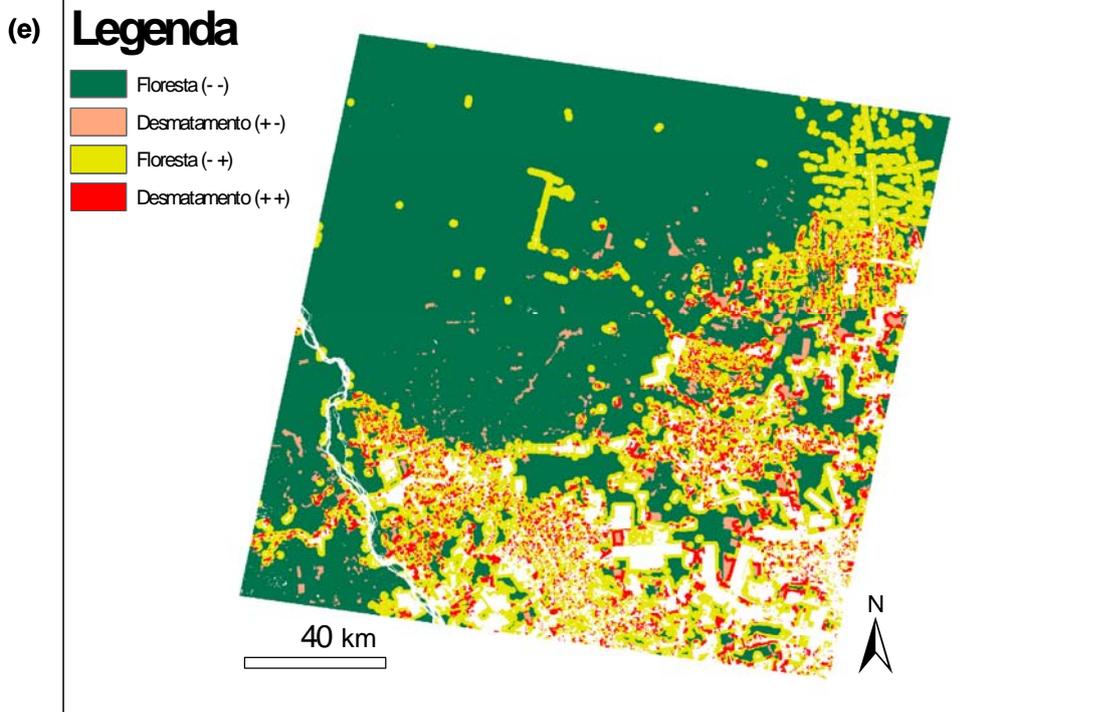
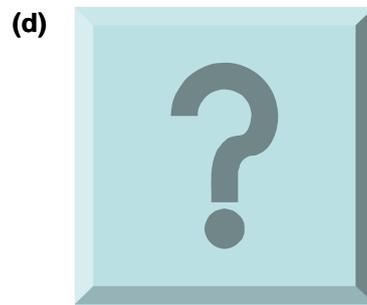
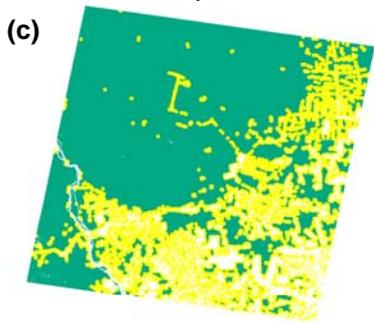


Figura 1. Desempenho do AMAZON no ano-base de 2000, levando-se em conta somente a variável 'PAD' e baseando-se nos coeficientes ' β_0 ' e ' β_1 ' de anos-base anteriores (1985, 1988, 1992 e 1996).

Vale ressaltar, ainda, que o desempenho do AMAZON-PD_{P00} apresentou um valor de DM igual a 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 estabelecidos para o desempenho do modelo (DM).

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

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

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.

3. Resultados e Discussão

3.1 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 mais próximas as primeiras candidatas 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 1), proveniente da interpretação, da segmentação e da classificação da imagem de satélite (em outras palavras, informação primária), cujos erros podem ser considerados desprezíveis na resolução adotada.

Essas razões, aliadas ao comportamento consistente do seu coeficiente (β_i) em todos os anos-base, fazem com que PAD seja mantida em cada um dos modelos preservados.

TABELA 1 – 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

3.2 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. Essa falta de sincronismo entre a construção de rodovias e sua catalogação provoca oscilações no grau de contribuição da variável.

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 (Tabela 2). 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 2 – 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

3.3 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.

A partir de 1996, foram homologados o Parque Nacional da Serra do Pardo e a FLONA do Itacaí, bem como as terras indígenas de Trincheira/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 3, 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 3 – 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.

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 2, nas cores vermelho e laranja, áreas onde UC/terras indígenas seriam homologadas mais tarde.

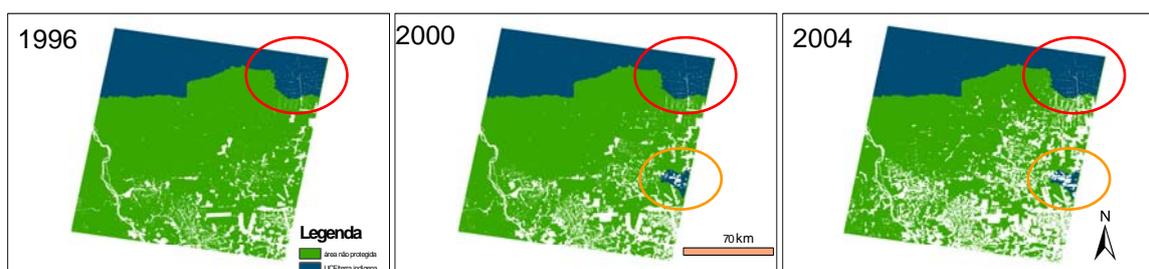


Fig. 2 – Evolução da variável EPA de 1996 a 2004.

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.

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

3.4 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. 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”.

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.

TABELA 4 – 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”.

3.5 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 5, 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 5 – 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

3.6 Comportamento da variável IFR

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.

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, que 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.

3.7 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.

3.8 Comportamento das variáveis W1 (PAD x ALT) E W2 (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 W1 quanto W2 tiveram bons desempenhos na composição dos modelos testados. No processo de escolha dos modelos preservados, entretanto, somente W1 se tornou parte de um modelo preservado, o que ocorreu no ano-base de 1996.

4 Conclusões

A agregação da dimensão temporal à regressão logística permitiu um melhor balanceamento do AMAZON, com a conseqüente melhoria de um desempenho que já era bastante satisfatório. Como resultado, houve uma redução das ocorrências de ‘falsos positivos’ e uma melhoria no ‘ajuste do modelo’, “em troca” de uma redução de ‘verdadeiros positivos’, trazendo o conjunto para os valores especificados como critério de desempenho.

Assim, implantado como AMAZON-PD, o método apresenta uma espacialização das probabilidades de impacto ambiental para uma data futura. 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 conceituais, que o AMAZON-PD é diretamente aplicável a outras regiões da Amazônia. Mais do que isso, são inúmeras as possibilidades de aplicação do método a qualquer outro bioma ou em um contexto de mudança global, onde outras formas de cobertura do solo venham a ser de interesse.

A razão para tal está no fato de que o AMAZON-PD se utiliza de variáveis biofísicas que são observáveis 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 biofísicas que venham a ser compreendidas como relevantes ao processo sob modelagem.

Referências Bibliográficas

Damião, D. P. **Uso de técnicas de análise multivariada para a predição de desmatamento na Amazônia: O modelo ‘AMAZON-PD’**. 2007. 125 f. Doutorado (Tese em Desenvolvimento Sustentável) – Centro de Desenvolvimento Sustentável – CDS, Universidade de Brasília – UnB, Brasília – DF

Damião, D. P.; Souza, N. M. **Análise multivariada aplicada a zoneamento: o método AMAZON**. *No prelo*.