

TÉCNICAS DE REGRESSÃO LOGÍSTICA APLICADA À ANÁLISE AMBIENTAL

TECHNIQUES OF LOGISTIC REGRESSION APPLIED TO ENVIRONMENTAL ANALYSIS

Carlos André Bulhões Mendes

Engenheiro Civil. Doutorado (Ph.D) em Planejamento Ambiental pela University of Bristol (Inglaterra). Professor Associado do Instituto de Pesquisas Hidráulicas (IPH) da Universidade Federal do Rio Grande do Sul (UFRGS).

Fausto Alfredo Canales Vega

Engenheiro Civil. M.Sc., Doutorando em Recursos Hídricos e Saneamento Ambiental pelo Instituto de Pesquisas Hidráulicas (IPH) da Universidade Federal do Rio Grande do Sul (UFRGS).

RESUMO: Este estudo tenta identificar quanto entendimento das forças motrizes de alterações no padrão do uso do solo, pode ser ganho com uma análise geográfica e estatística. Tal abordagem é aplicada no ecossistema Pantanal no Brasil e leva em conta a variabilidade espacial dos processos de mudanças da terra. Os modelos aqui desenvolvidos sugerem que alguns fatores principais dirigem os processos de mudanças, relacionando-se a algumas fundamentações teóricas bem estabelecidas. Os resultados sugerem que a conversão para uso agrícola é controlada por custos de transporte e fatores agroclimáticos. Heterogeneidade espacial como também a variabilidade temporal afeta a habilidade para utilizar os modelos de regressão para grandes extrapolações. Os modelos permitem avaliar o impacto de mudanças derivados das forças indutoras que são representadas através de critérios geográficos. Os modelos poderiam ser incorporados em um arcabouço institucional de planejamento, análise de cenários e avaliação de impactos a nível governamental.

Palavras-chave: regressão logística, Pantanal, Sistema de informações geográficas.

ABSTRACT: *This study attempts to identify how much understanding of the driving forces of land-use changes can be gained through a spatial, statistical analysis. Such approach was developed to Pantanal ecosystem, taking into account the spatial variability of the land-use change processes. The spatial models developed here suggest some important factors driving the land-use changes that can be related to some well-established theoretical frameworks. The results, suggest that the conversion to agriculture land-use is controlled by the transportation costs, and agro-climatic potential factors. Spatial heterogeneity as well as the variability in time of land-use change processes affects the ability to use regression models for wide ranging extrapolations. The models allow evaluating the impact of changes in driving forces that are well represented by proximate causes of land-use change. The models could be incorporated into a land use framework for planning purposes, scenario analysis or impact assessment at the government level.*

Keywords: *logistic regression, Pantanal, Geographical information system.*

INTRODUÇÃO

Para Verburg *et al.* (2004), os modelos de dinâmicas de alteração de uso do solo podem ser definidos como ferramentas de apoio à análise das causas e conseqüências de mudanças no tipo de uso dos solos, a fim de compreender melhor o funcionamento do sistema de uso da terra e auxiliar no planejamento e formulação de políticas do uso da terra. Mudanças de uso do solo referem-se a alterações no padrão de ocupação e/ou cobertura de parcelas do espaço geográfico. A cobertura do solo é descrita pelo estado biofísico da superfície. Floresta, mangue e mar são exemplos de coberturas do solo. A ocupação do solo, por sua vez, envolve a maneira como os atributos biofísicos do solo são manipulados e

utilizados por seres humanos. Assentamentos humanos e áreas de agricultura são exemplos de ocupações do solo. Neste trabalho, o termo “mudanças de uso do solo” é utilizado abrangendo tanto as mudanças de cobertura do solo quanto às mudanças de ocupação do solo.

O potencial do impacto adverso dessas transformações depende de sua escala, extensão, padrão espacial e das dinâmicas relações entre os usos do solo. Segundo Meyer & Turner (1994, *apud* KOOMEN & STILLWELL, 2007), a mudança de tipo de uso do solo é um processo dinâmico e complexo, que liga sistemas naturais e humanos. Tem impactos diretos no solo, água e atmosfera e, portanto, está diretamente relacionado com muitas questões ambientais de importância global. Lambin *et al.* (2003) citam como exemplo de dinâmicas de alteração de uso do solo, os desmatamentos em grande escala e posterior transformação em terras agrícolas na região dos trópicos, e destacam os seus prováveis impactos sobre a biodiversidade, a degradação do solo e a capacidade produtiva da terra para satisfazer as necessidades humanas. Além disso, Koomen & Stillwell (2007), com base nos trabalhos de Dale (1997) e Watson *et al.* (2000), mencionam que as dinâmicas de alteração do uso do solo são também um fator importante no ciclo das mudanças climáticas, e que a relação entre os dois é interdependente. As modificações no uso da terra podem afetar o clima, e por sua vez, as mudanças climáticas também influenciarão o futuro uso da terra.

Neste contexto, modelos geográficos que analisam as dinâmicas de uso do solo podem ser entendidos como um conjunto de técnicas utilizadas para descrever os processos locais das mudanças em condições quantitativas e para testar o entendimento destes processos. Atualmente, existem vários esquemas de modelagem da dinâmica de alteração do uso e cobertura do solo. No geral, estes modelos podem ser classificados de duas maneiras:

- Previsão do padrão geográfico de mudanças, formulando a pergunta “onde as modificações do uso do solo ocorrerão?”
- Previsão das taxas de mudança, redigindo a questão “a que taxa as mudanças ocorrerão?”

Enfatize-se que a projeção de taxas futuras de modificação do uso do solo é uma tarefa mais difícil, pois requer um profundo entendimento das dinâmicas das forças motrizes das mudanças (RIEBSAME *et al.*, 1994). As forças motrizes podem, às vezes, estar distantes em espaço e/ou tempo das observações das mudanças, e freqüentemente envolvem condições macroeconômicas e mudanças de política que são difíceis se

antecipar. Neste quadro os objetivos gerais deste artigo, através de técnicas de regressão logística, são:

1. Desenvolver um modelo geográfico e estatístico das relações causais de modificações do uso do solo, levando em conta a variabilidade espacial dos processos de mudança,
2. Identificar quanto entendimento das forças indutoras destas mudanças pode ser ganho com esta análise estatística.

A metodologia proposta é testada no Pantanal. Observa-se na **figura 1** a divisão política dos países (Brasil, Bolívia e Paraguai) a bacia hidrográfica do Pantanal, também denominada Alto Paraguai (BAP), tendo como limite de jusante a drenagem da sub-bacia do Rio Apa, na fronteira entre Brasil e Paraguai. O Pantanal insere-se neste contexto.

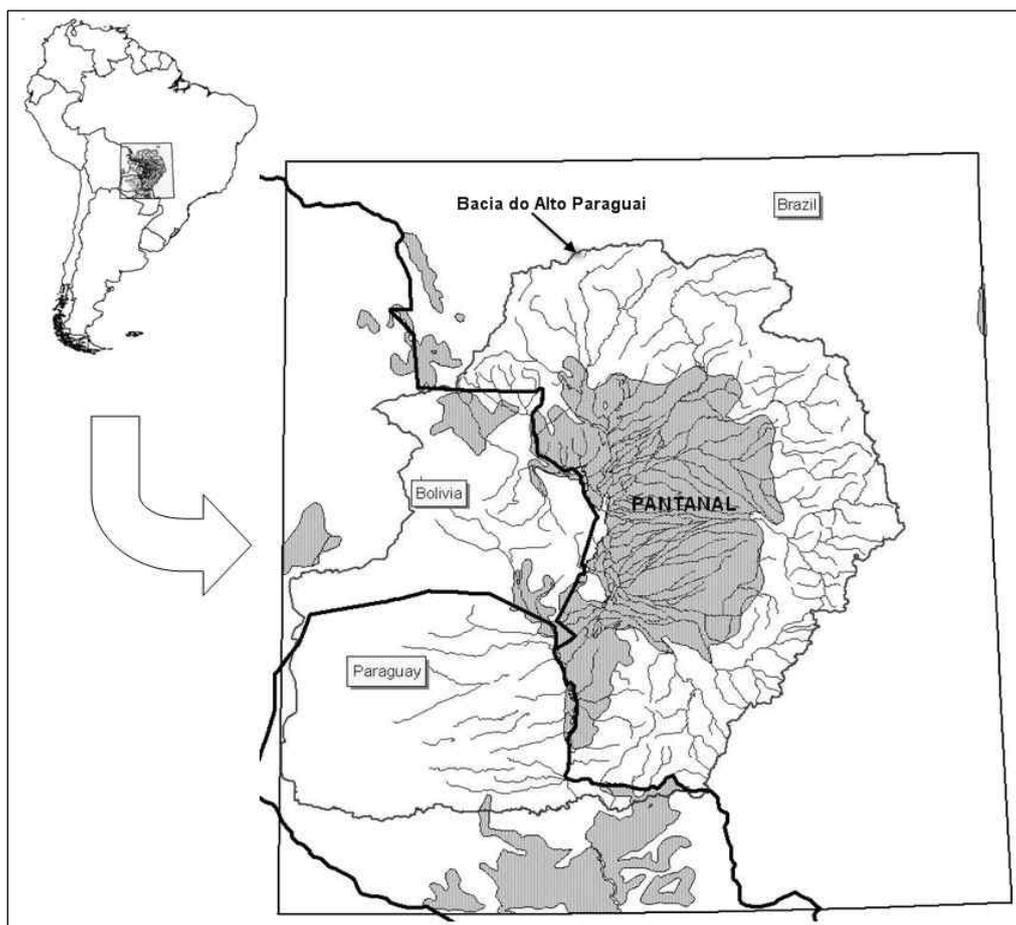


Figura 1 - Localização do Pantanal, inserido na Bacia do Alto Paraguai, no contexto da América do Sul.

A **figura 2 (a)** apresenta o resultado da classificação digital, com várias classes de uso e cobertura dos solos. A **figura 2 (b)** ilustra a área de estudo em formato matricial,

representado a projeção da Bacia hidrográfica do Alto Paraguai (BAP) em território brasileiro. A interseção dos mapas (a) e (b) é mostrada na **figura 2 (c)**. Esta última imagem (**figura 2c**) é o alvo da investigação deste estudo, ou seja, a projeção da bacia em território brasileiro.

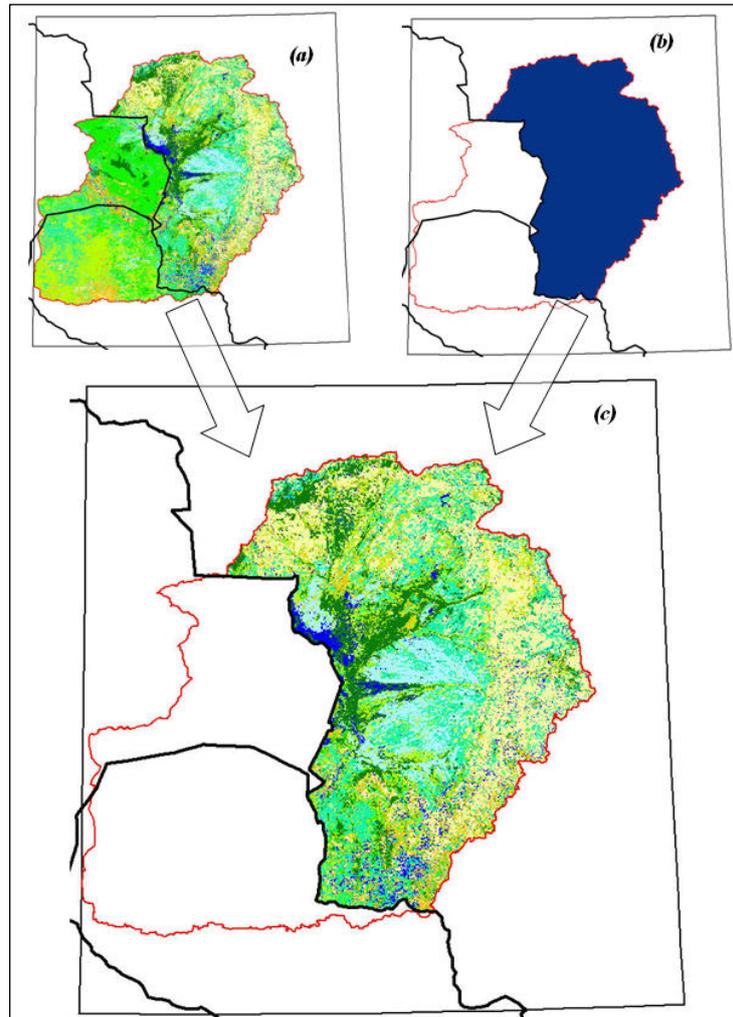


Figura 2 - Uso e cobertura do solo, para o ano 2000, na Bacia do Pantanal, onde: a) Imagem classificada para toda a bacia; b) Máscara utilizada para definição da área de estudo; c) Uso e cobertura do solo para a Bacia do Pantanal em território brasileiro

A **figura 3** apresenta apenas a área agrícola. Reclassificou-se a imagem original (**figura 2c**) de modo a gerar uma imagem binária, isto é, com duas classes (0 ou 1, pertence ou não pertence a classe em questão). Assim, o que se deseja é a identificação das variáveis explicativas da ocorrência de áreas agrícolas na Bacia do Alto Paraguai (BAP) em território brasileiro (**figura 3**). A seguir são apresentados os detalhes metodológicos da regressão

logística, estudo de caso aplicado na região do Pantanal no Brasil com a descrição das etapas de obtenção dos dados, extração dos parâmetros do modelo em programa estatístico e avaliação da qualidade dos resultados. Com o modelo construído é testado um cenário hipotético de construção de uma ferrovia, sendo seu impacto avaliado.

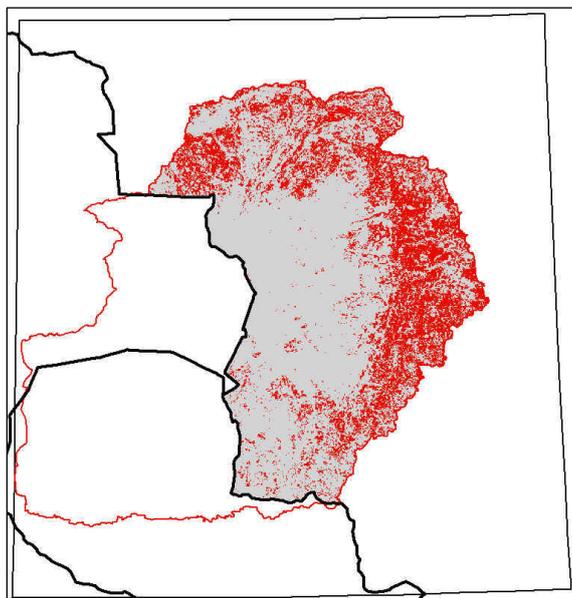


Figura 3 - Agricultura intensa, na Bacia do Alto Paraguai pelo lado brasileiro no ano 2000.

MODELO LOGÍSTICO

A análise aqui apresentada baseia-se no uso de técnicas de regressão logística. A regressão logística é uma ferramenta estatística que permite o ajuste de um conjunto de variáveis independentes a uma variável de resposta categórica. Ao contrário das variáveis contínuas, as variáveis categóricas podem assumir apenas alguns valores particulares de resposta, podendo estes ser binários (dicotômicos) cuja resposta possui apenas dois níveis (falha ou sucesso) ou politômicos (mais de três classes), uma extensão do anterior, no qual a resposta pode assumir múltiplos níveis de saída (HOSMER & LEMESHOW, 2000). A regressão logística aplicada na modelagem de mudanças de uso do solo tem sido utilizada por Serneels & Lambin (2001), Aspinall (2004) e Bakker *et al.* (2005), entre outros autores.

O método de Regressão Logística é utilizado neste estudo para avaliar quais os fatores afetam, e em que extensão, a localização de usos do solo na região de análise. Para ilustrar estes conceitos considere a variável dependente e dicotômica “Uso Agrícola ($A_{0, 1}$)” onde o valor **0** representa as áreas sem uso agrícola e **1** com o uso agrícola, conforme ilustrado na **figura 2**. Supondo-se que uma variável independente **z** (por exemplo, cota

topográfica) consegue explicar ou prever o comportamento da variável dependente **A**, de tal forma que:

$$\begin{array}{ccc} \mathbf{z} & \xrightarrow{\text{explica}} & \mathbf{A}_{0,1} \\ \text{Var. Independente} & & \text{Var. Dependente} \end{array} \quad (1)$$

Onde **z** poder ser uma única variável ou um conjunto de variáveis independentes. A regressão logística é o procedimento matemático que pode ser utilizado para descrever a relação de $\mathbf{z}(z_1, z_2, \dots, z_n)$ com uma variável dicotômica, como **A**. A função logística $f(z)$ descreve a forma matemática na qual o modelo logístico se baseia, sendo dado por (STRAUSS, 1992):

$$f(z) = \frac{e^z}{1 + e^z} \quad \text{ou} \quad f(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

A **figura 4** ilustra o comportamento do modelo de regressão logística versus um modelo linear. Observa-se que, independente do valor de **z**, a amplitude de resultados de $f(z)$ está entre 0 e 1, ou seja, $0 \leq f(z) \leq 1 \quad \forall \quad -\infty \leq z \leq +\infty$. Comparando-se com o modelo linear observa-se que os mesmos escapam dos limites entre 0 e 1. Esta forma de abordagem também é utilizada para descrever probabilidades (que também é um número entre 0 e 1).

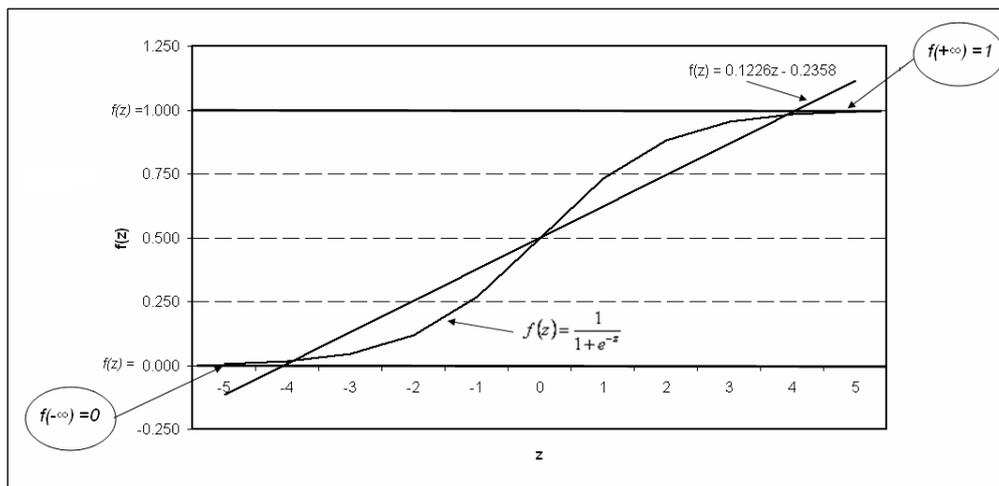


Figura 4 - Comparação entre a curva logística e linear.

Aplicando tais conceitos ao exemplo “Uso Agrícola (**A**)”, tem-se a probabilidade de uma área ser convertida em uso agrícola como função dos valores de **z** (cota topográfica,

por exemplo). Saliente-se, entretanto, que o procedimento normalmente utilizado faz uso de n variáveis independentes (ou explicativas) de tal forma que:

$$\begin{array}{ccc} X_1, X_2, \dots, X_n & \xrightarrow{\text{explica}} & A_{0,1} \\ \text{Var. Independentes} & & \text{Var. Dependente} \end{array} \quad (3)$$

Face ao exposto, esta técnica se mostra bastante adequada para a modelagem da transição entre tipos de uso do solo, devido que estas passagens serão tratadas neste estudo como estados individuais ou categorias particulares (mata → área rural → zona urbana) e estarão relacionadas a uma série de variáveis independentes ou explanatórias, extraídas a partir de imagens de sensoriamento remoto orbital. Neste contexto o papel da regressão logística é semelhante ao dos modelos de regressões tradicionais, ou seja, encontrar o melhor modelo que relacione variáveis dependentes e independentes. Neste quadro, a variável independente z (equação 2) torna-se um índice (combinação linear) que agrega as n variáveis independentes (\mathbf{X} 's):

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (4)$$

Substituindo-se a equação 4 em 2, resulta em:

$$f(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}} \quad (5)$$

A expressão anterior ($f(z)$) pode ser entendida em um conceito probabilístico como:

$$P(D = 1 | X_1, X_2, \dots, X_n) = P(\mathbf{X}) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}} \quad (6)$$

Onde:

- $P(D = 1 | X_1, X_2, \dots, X_n)$: probabilidade condicional de ocorrer o evento $D = 1$ (variável dicotômica $D_{0,1}$), conhecido como modelo logístico, dado o vetor \mathbf{X} .
- X_1, X_2, \dots, X_n ou o vetor \mathbf{X} : variáveis independentes.
- α, β_i : parâmetros desconhecidos a serem estimados por métodos matemáticos.

Saliente-se que o valor do parâmetro α controla $P(\mathbf{X})$ quando \mathbf{X} 's são zero, e os valores de β_i ajustam a taxa de modificação de $P(\mathbf{X})$ com relação às variáveis independentes \mathbf{X} 's. Devido a que a relação entre X e $P(X)$ é não linear, β não tem uma interpretação direta neste modelo como em uma regressão linear. Isto implica que a taxa de modificação da curva logística modifica-se com diferentes valores de X_0 . A declividade mais elevada é quando $P(X_0) = 0,5$; enquanto que as declividades mais baixas são com $P(X)$ aproximando-se de 0 ou 1.

Uma propriedade interessante é que o modelo logístico pode ser linearizado. Este procedimento visa facilitar a obtenção dos parâmetros (α, β_i) tornando-se necessária a apresentação de alguns conceitos adicionais, bem como efetuar alguns procedimentos algébricos para simplificar a equação 6. Saliente-se que enquanto $P(\mathbf{X})$ está restrita ao intervalo $0 \leq P(\mathbf{X}) \leq 1$, os parâmetros (α, β_i) não possuem tais restrições. Uma forma de resolução do problema é transformar $P(\mathbf{X})$ para eliminar as restrições. Para o caso de variáveis dicotômicas, pode-se eliminar a fronteira superior ($P(\mathbf{X})= 1$), utilizando a razão $P(\mathbf{X})/(1- P(\mathbf{X}))$. Esta razão sempre é positiva, pois $0 \leq P(\mathbf{X}) \leq 1$, mas não há limite superior. Quando $P(\mathbf{X}) \rightarrow 1$, a razão $P(\mathbf{X})/(1- P(\mathbf{X})) \rightarrow \infty$. O limite inferior, com o valor 0, pode ser eliminado aplicando-se o logaritmo natural da razão, ou seja, $\ln(P(\mathbf{X})/(1- P(\mathbf{X})))$. O resultado desta transformação são números reais no intervalo $-\infty \leq \ln(P(\mathbf{X})/(1- P(\mathbf{X}))) \leq +\infty$, de forma semelhante aos parâmetros (α, β_i) . Em síntese, efetua-se a transformação da probabilidade $P(\mathbf{X})$:

$$\text{logit } P(\mathbf{X}) = \ln_e \left[\frac{P(\mathbf{X})}{1 - P(\mathbf{X})} \right] \quad (7)$$

Cramer (2003) explica que esta transformação da probabilidade $P(\mathbf{X})$ é denominada *logit (logistic probability unit)*, termo criado por Berkson (1944). A razão $P(\mathbf{X})/(1- P(\mathbf{X}))$ na transformação *logit* é chamada de *Odds (Chance)*. Substituindo-se

$$P(\mathbf{X}) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}} \text{ em } 1 - P(\mathbf{X}), \text{ tem-se:}$$

$$1 - P(\mathbf{X}) = 1 - \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}} = \frac{e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}}{1 + e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}} \quad (8)$$

A razão de *Odds* (chance) resulta em:

$$\frac{P(\mathbf{X})}{1 - P(\mathbf{X})} = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}} = e^{\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)} \quad (9)$$

Substituindo-se os termos anteriores na equação 7, conclui-se que:

$$\text{logit } P(\mathbf{X}) = \ln_e \left[\frac{P(\mathbf{X})}{1 - P(\mathbf{X})} \right] = \ln_e \left[e^{\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)} \right] = \alpha + \sum_{i=1}^n \beta_i X_i, \quad (10)$$

Em síntese, a transformação *logit* da probabilidade $P(\mathbf{X})$ simplifica o modelo logístico para uma combinação linear das variáveis independentes. Outro conceito importante é a razão de *Odds* (Chance) na transformação *logit*. Numa forma simples, uma *Chance* é razão entre a probabilidade que um evento irá ocorrer ($P(\mathbf{X})$) sobre a probabilidade que o mesmo evento não irá ocorrer ($1 - P(\mathbf{X})$). Por exemplo, considere-se a probabilidade de ocorra uma modificação de uso do solo da classe floresta → agricultura, em uma determinada posição do terreno, com $P = 0,25$. Então $1 - P = 0,75$. A Razão ($P / (1 - P)$) = $0,25/0,75 = 1 / 3$, ou seja, as chances são 3 vezes maiores do evento não ocorrer do que ocorrer. Neste contexto, Hosmer & Lemeshow (2000) citam que a importância desta transformação reside no fato de que *logit* $P(\mathbf{X})$ assume muitas propriedades do modelo de regressão linear, com parâmetros contínuos e com uma faixa de valores possíveis expandida para o intervalo $[-\infty, +\infty]$.

APLICAÇÃO METODOLÓGICA NA BACIA DO ALTO PARAGUAI (BAP) NO BRASIL

Determinação da variável dependente

Para o início do processo é selecionada a categoria de estudo, ou seja, reclassifica-se a imagem original de modo a gerar uma imagem binária (0 ou 1, pertence ou não pertence a classe em questão). Assim, para o caso da categoria Agricultura obtêm-se as imagens conforme ilustrado na **figura 5**, correspondente a dois períodos, 1992 e 2000.

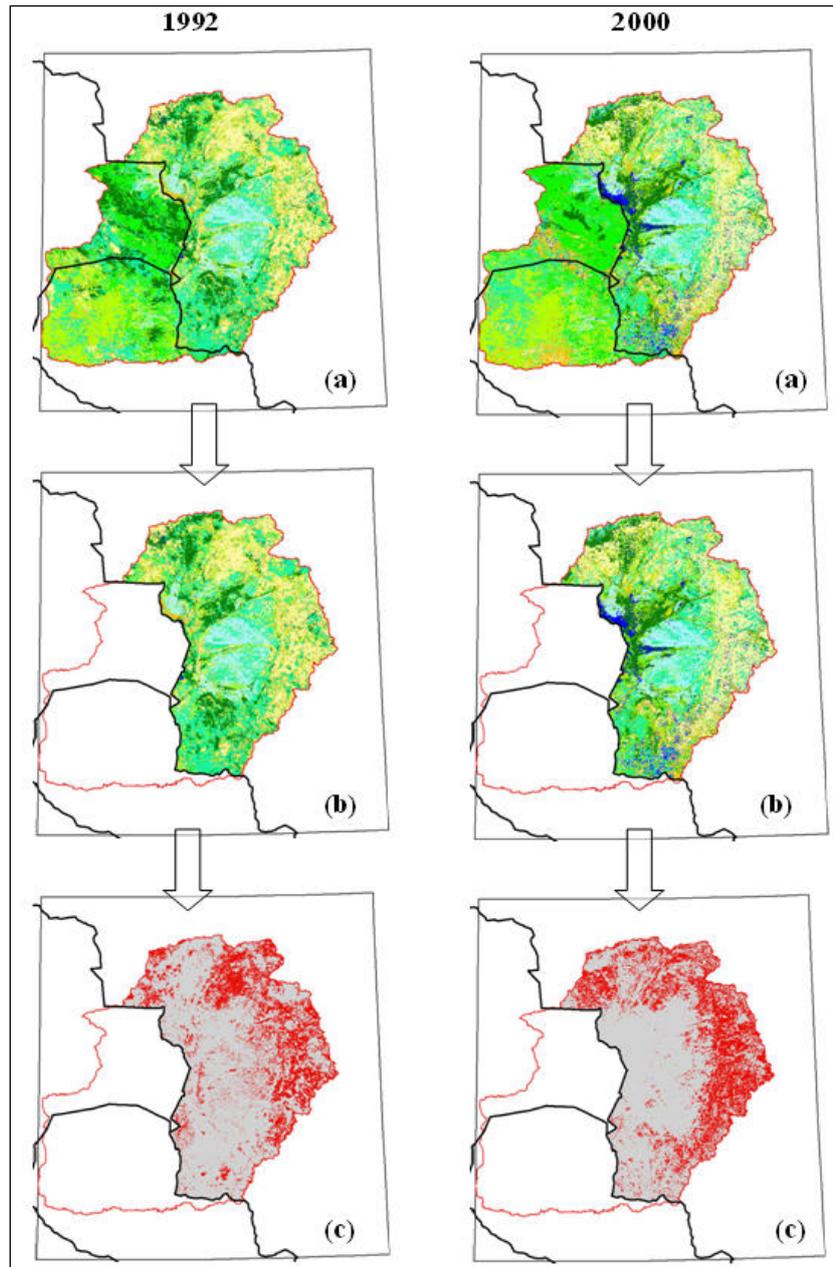


Figura 5 - Determinação de áreas de uso agrícola na BAP, onde: (a) Uso do solo; (b) uso do solo no território brasileiro da BAP; e (c) uso do solo com áreas agrícolas e “não-agrícola”.

O próximo passo é a determinação de uma imagem de probabilidade de transição (“não-agrícola” → Área agrícola) no período de tempo considerado (1992 → 2000). Para fins da presente análise na categoria “não-agrícola” inclui-se toda e qualquer forma de ocupação do solo desde que não agrícola, ou seja, cobertura vegetal nativa, áreas de assentamentos humanos, cursos de água etc. Saliente-se que procedimento similar pode ser adotado quando na determinação de outras classes, como florestas e cerrados. As duas imagens das áreas agrícolas e “não-agrícola” (1992 e 2000) da **figura 5 (c)** são sobrepostas

e o resultado obtido corresponde a novas imagens onde se observam as transições ocorridas na área agrícola para o período 1992-2000.

O **quadro 1** ilustra as possíveis transições que podem resultar do cruzamento das imagens binárias da classe agricultura relativas ao período 1992-2000, onde as colunas correspondem aos dados do ano 2000 e as linhas a 1992: os estados 1 e 0 indicam a classe agrícola e “não-agrícola” do solo, sucessivamente. A posição “a” do **quadro 1** (estados 1-1) indica que 13,29% da área da bacia era agricultura em 1992 permaneceu agricultura no ano 2000. Já a posição “c” (estado 0-1) mostra que 15,13% da área da bacia era “não agrícola” em 1992 e foi convertida em agricultura no ano 2000. A **figura 6** ilustra todos os casos indicados no **quadro 1**.

Agricultura		2000		
1992	Estados	1	0	Total
	1	a = 13,29%, permanência	b = 10,37%, substituição	23,66%
	0	c = 15,13%, expansão	d = 61,21%, permanência	76,34%
	Total	28,41%	71,59%	100,00%

Quadro 1 - Transições na classe agricultura para o intervalo 1992-2000.

Onde:

- a = em 1992 e em 2000 são classe agrícola, ou seja em ambas as imagens da **figura 5 (c)** os valores são 1. Logo tem-se uma situação de 1 – 1 → permanência.
- b = em 1992 era agricultura e no ano 2000 transformou-se em outro uso, indicando uma situação de 1 – 0 → substituição nas áreas agrícolas.
- c = em 1992 apresentava uso “não-agrícola” do solo e no ano 2000 converteu-se em área agrícola, indicando uma situação de 0 – 1 → expansão; nas áreas agrícolas.
- d = em 1992 e em 2000 apresenta uso “não-agrícola” do solo, ou seja em ambas as imagens da **figura 5 (c)** os valores são 0. Logo tem-se uma situação de 0 – 0 → permanência.

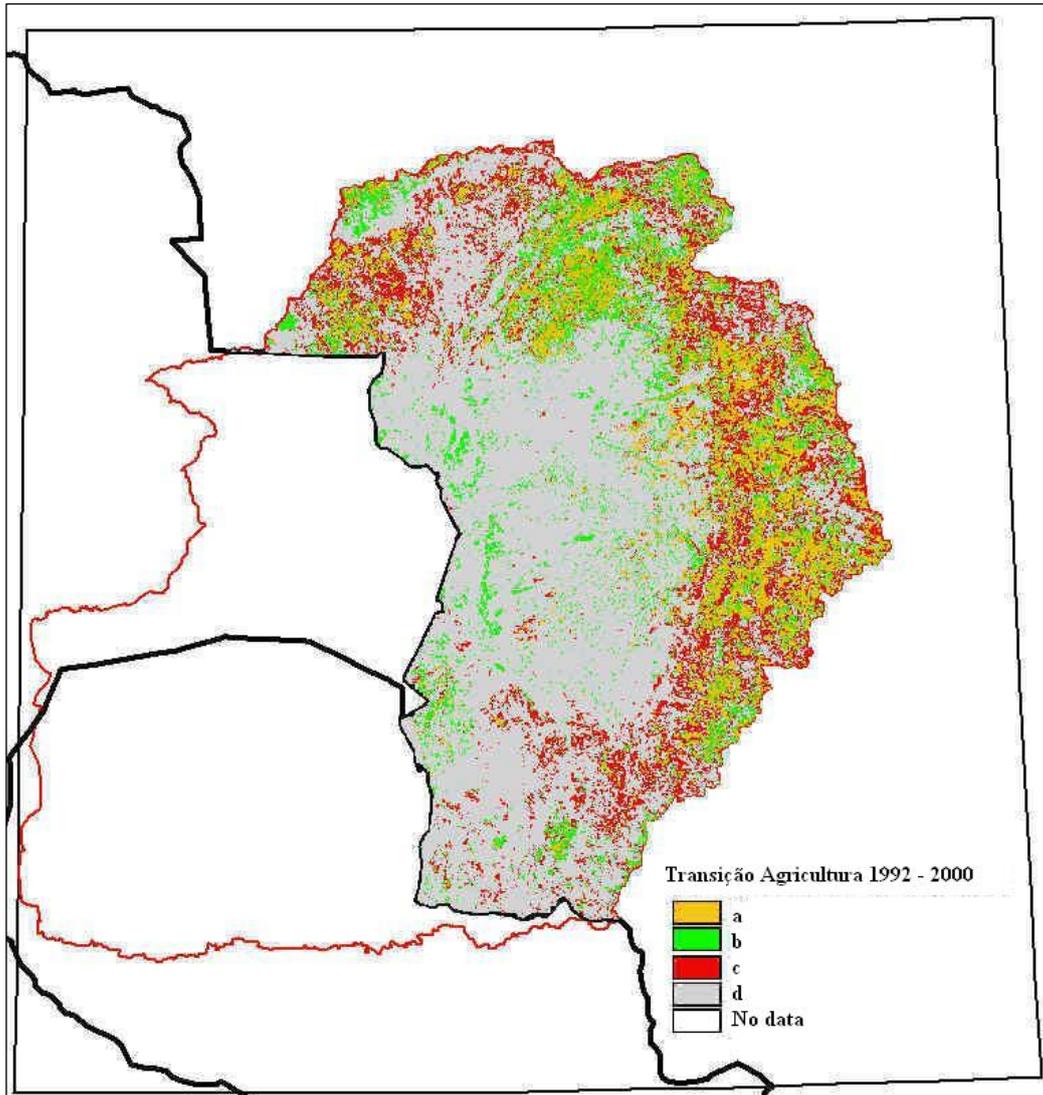


Figura 6 - Transições na classe agricultura para o intervalo 1992-2000.

Enfatize-se que a localização espacial do caso “c” (0-1, “não-agrícola” → agricultura) ocorre na sua íntegra na região de planalto da Bacia Hidrográfica do Alto Paraguai (BAP), no Brasil, onde também se localizam a maioria das cabeceiras dos rios afluentes ao Pantanal, torna-se urgente o controle destas conversões, pois as mesmas causam um impacto direto no Pantanal. Das mudanças indicadas no **quadro 1**, apenas as que se referem aos eventos de expansão (0-1, não-agrícola → agricultura) serão consideradas para a aplicação do modelo logístico. A retração da classe no período não será considerada. A transição 1 – 1 refere-se à permanência da classe agricultura não será considerada, pois se deseja explicar o processo de expansão da agricultura na área. Neste contexto a imagem da **figura 6**, que indica todas as transições de uso de solo, relacionadas à área agrícola, entre o período de 1992-2000 é reclassificada para uma imagem binária, indicando apenas o caso os eventos de expansão (0-1, não-agrícola → agricultura, caso “c”

do **quadro 1**). A **figura 7** indica a localização espacial das áreas convertidas em agricultura entre o período de 1992-2000, sendo esta a variável dependente.

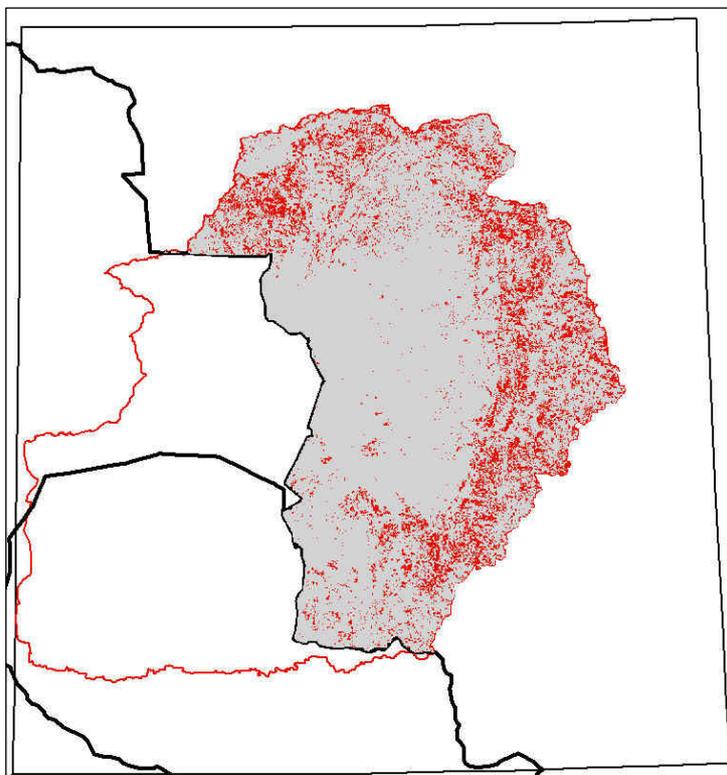


Figura 7 - Eventos de expansão (0-1, “não-agrícola” → agricultura).

Enfatize-se que a **figura 7** representa uma imagem de transição (dicotômica) na qual estão representados os locais onde houve permanência de outras classes e os locais onde houve a transição “não-agrícola” → agricultura, no intervalo de tempo em análise. Ou seja, a **figura 7** corresponde ao item (c) definido pelo **quadro 1**. Para o cálculo das probabilidades de transição (“não-agrícola” → agricultura) definiu-se um filtro, que consiste em uma matriz unitária ($N \times N$). O tamanho da amostra utilizado no cálculo dos ponderadores W_i será justamente o produto do número de linhas e colunas ($N \times N$) da matriz de filtragem, conforme ilustrado na **figura 8**. A imagem de transição (**figura 7**) é então percorrida por esta matriz e para cada pixel central é calculada a probabilidade de ocorrência do evento somando-se o número de pixels da janela de filtragem que correspondam ao valor **1** da imagem booleana e dividindo-se esta soma pelo número de pixels da matriz filtro, isto é, efetua-se uma média simples. Esta probabilidade é calculada em função da ocorrência do evento nos pixels vizinhos assumindo desta forma um caráter de dependência espacial. Desta forma, valores elevados (próximos a **1**) são indicativos de uma maior tendência ou maior inércia para ocorrência do evento em questão (transição

“não-agrícola” → agricultura). O resultado da operação de filtragem espacial é apresentado na **figura 9**, sendo esta correspondente à variável dependente do modelo logístico.

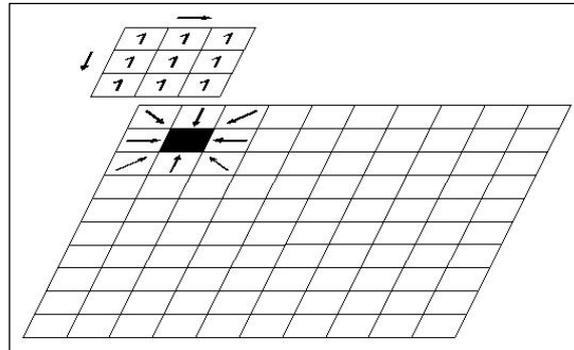


Figura 8 - Filtro aplicado à imagem de transição.

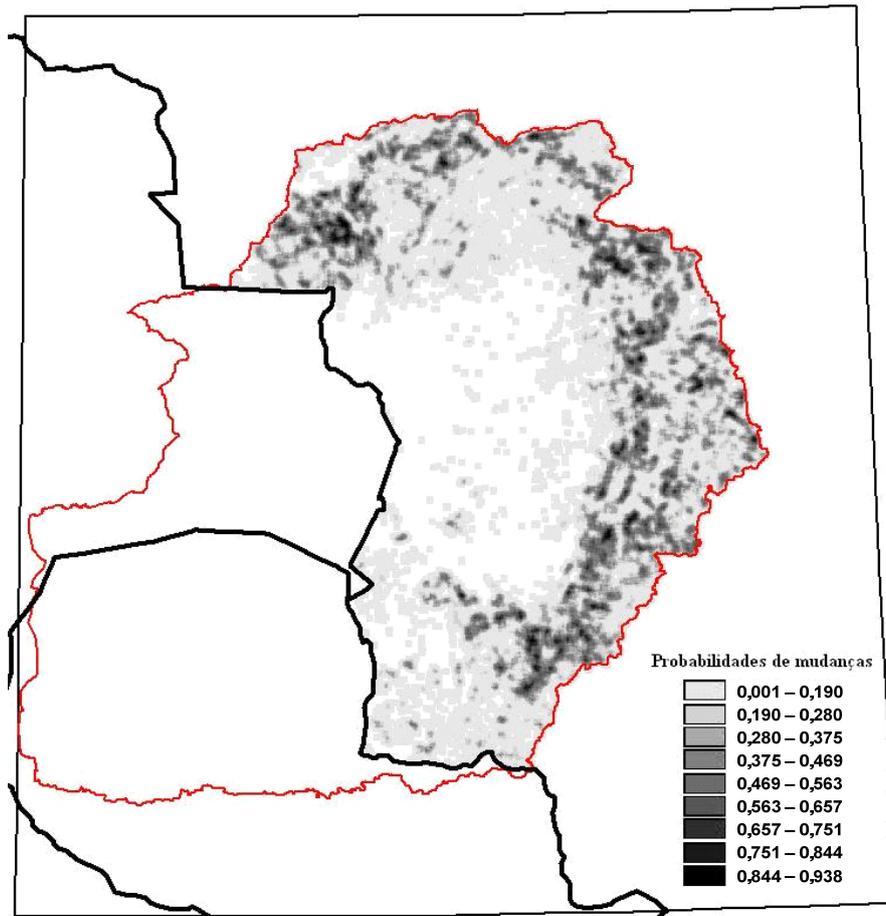
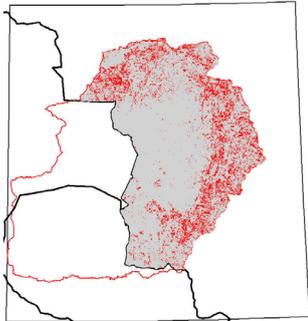
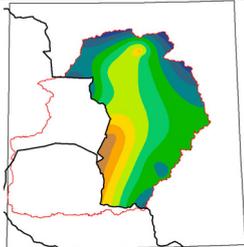


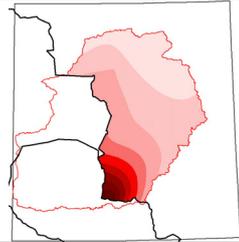
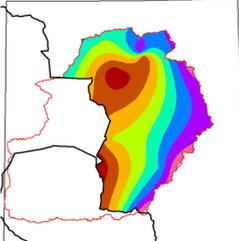
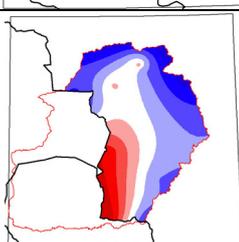
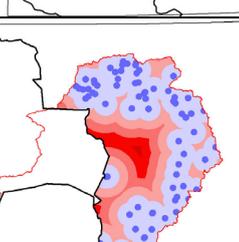
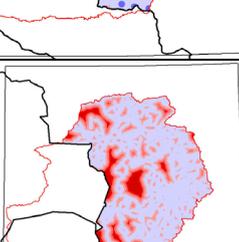
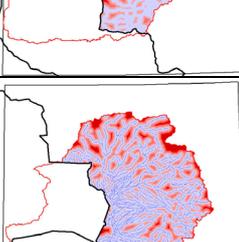
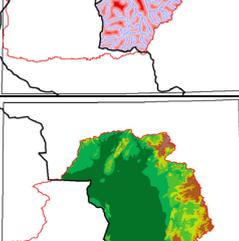
Figura 9 - Probabilidade da ocorrência de transição “não-agrícola” → agricultura.

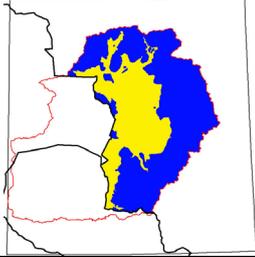
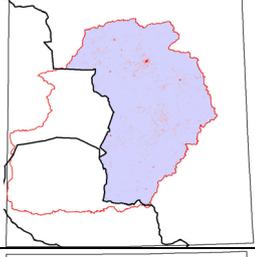
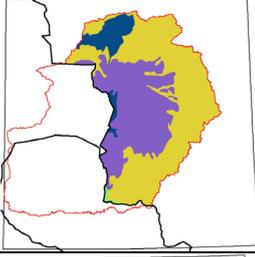
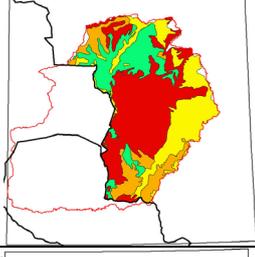
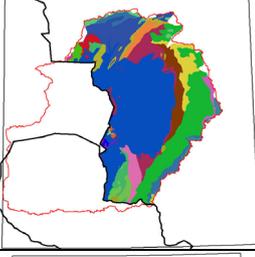
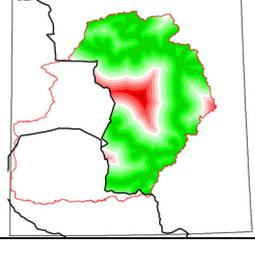
Realização de análises estatísticas

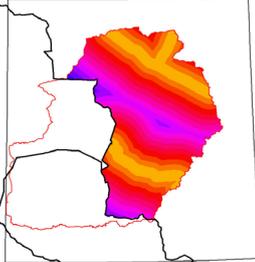
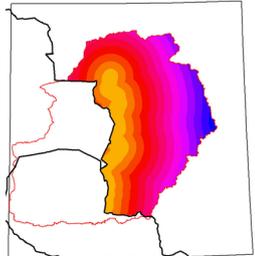
Assume-se que a ocorrência de eventos que modifiquem o uso do solo se deve a proximidade geográfica de agentes presentes na área em questão e que são considerados motivadores para as modificações. Desta forma é razoável supor que a distância à rodovias, cursos d'água, núcleos urbanos, limites florestais, condições climáticas, geologia, etc. constituem variáveis de interesse (variáveis independentes) e devem estar presentes na elaboração do modelo.

A probabilidade da ocorrência de transição “não-agrícola” → agricultura (**figura 9**) corresponde à variável dependente do modelo, ou seja, a expansão da agricultura na Bacia Hidrográfica do Alto Paraguai - BAP, pelo lado brasileiro. As variáveis independentes, que induzem ou explicam as mudanças ocorridas entre as classes de uso e cobertura do solo (não-agrícola → agricultura), encontram-se listadas no **quadro 2**. Estes fatores indutores de mudanças podem ser agrupados em duas classes: (a) fatores estáveis, isto é, que permanecem constantes durante o período de simulação, por exemplo, a topografia e (b) fatores dinâmicos no caso contrário, tendo como exemplo a densidade populacional.

VARIÁVEIS	TIPO	UNIDADES	FIGURAS
Variável dependente			
Áreas de transição de classes de uso do solo: “não-agrícola” → Agricultura.	Binária	0-1	
VARIÁVEIS	TIPO	UNIDADES	FIGURAS
Variáveis independentes			
Precipitação média anual (PREC)	Contínua	mm/ano	

<p>Evapotranspiração (ETP) média anual</p> <p>(ETP)</p>	Contínua	mm/ano	
<p>Temperatura média anual</p> <p>(TMA)</p>	Contínua	°C	
<p>Balanço Hídrico Climático (PREC - ETP) médio anual</p> <p>(BHC)</p>	Contínua	mm/ano	
<p>Distâncias para as sedes urbanas de municípios brasileiros inseridos na bacia</p> <p>(D_URB)</p>	Contínua	Múltiplos de graus	
<p>Distâncias para as rodovias brasileiras inseridas na bacia</p> <p>(D_ROD)</p>	Contínua	Múltiplos de graus	
<p>Distâncias para os rios brasileiros inseridos na bacia</p> <p>(D_RIOS)</p>	Contínua	Múltiplos de graus	
<p>Cotas topográficas na Bacia pelo lado Brasileiro</p> <p>(TOPO)</p>	Contínua	m	

<p>Cotas topográficas reclassificadas (1 - ≤ 150 m., 2 - > 150 m.)</p> <p>(TOPO_REC)</p>	<p>Categórica</p>	<p>1-2</p>	
<p>Densidade Populacional em 2000</p> <p>(DENS_POP)</p>	<p>Contínua</p>	<p>Habitantes/km²</p>	
<p>Ecorregiões na Bacia pelo lado Brasileiro</p> <p>(ECOR)</p>	<p>Categórica</p>	<p>1-6</p>	
<p>Mapa de solos reclassificado para os procedimentos da Curva- Número (CN) do N.R.C.S.</p> <p>(SOL_REC)</p>	<p>Categórica</p>	<p>1-4</p>	
<p>Mapa geológico representado pela litologia</p> <p>(LITO)</p>	<p>Categórica</p>	<p>1-32</p>	
<p>Distâncias para Estradas Principais (Pavimentadas e Federais) brasileiras inseridas na bacia</p> <p>(D_ESTRADA)</p>	<p>Contínua</p>	<p>Múltiplos de graus</p>	

<p>Distâncias para as Ferrovias brasileiras inseridas na bacia</p> <p>(D_FERRO)</p>	<p>Contínua</p>	<p>múltiplos de graus</p>	
<p>Distâncias para a hidrovía do rio Paraguai (Cáceres → Foz do Apa) inseridas na bacia</p> <p>(D_RPAR)</p>	<p>Contínua</p>	<p>múltiplos de graus</p>	

Quadro 2 - Variáveis, simbologia, tipos e unidades utilizadas para a modelagem logística.

De posse destes dados, aplica-se a equação 10 que descreve a relação matemática entre a variável dependente e as variáveis independentes consideradas para cada uso do solo. Inicialmente identificam-se as áreas cobertas pelo uso do solo específico em análise (variáveis dependentes). Posteriormente, é selecionado o conjunto de fatores indutores/condicionantes potenciais para este uso do solo (variáveis independentes). Em seguida, são analisadas estatisticamente as características locais das células ocupadas por um determinado uso do solo descritas pelos mapas de fatores locais. Este processo produz dois resultados:

- A equação de regressão que descreve a relação matemática entre a variável dependente e as variáveis independentes consideradas para cada uso do solo;
- Um mapa para cada uso do solo descrevendo a probabilidade de transição de cada célula deste uso do solo considerando as características locais.

Esses mapas de probabilidade de transição, para cada classe de uso, baseiam-se estatisticamente nas características de diversos fatores locais, sendo comparados entre si, atribuindo-se a cada classe de uso do solo aquela de maior probabilidade. Os dados empregados neste estudo encontram-se listados no **quadro 2** (variáveis dependentes e independentes). Foram identificadas uma variável dependente (probabilidade de expansão da área agrícola, **figura 9**) e 16 variáveis independentes. Todos os dados são representados em formato matricial, com o mesmo número de linhas e colunas, e georreferenciado. Saliente-se que as análises aqui apresentadas foram efetuadas com os softwares: *Microsoft Office EXCEL*, *ArcGIS* (Geoprocessamento) e *SPSS* (Estatístico). No nível operacional, a seqüência de operações compreende:

1. Transformar os dados de formato *GRID* (formato matricial do *ArcView*, *ArcGIS*) em ASCII (texto). Isto é feito para todos os arquivos utilizados. O resultado de tal operação pode ser lido em qualquer editor de texto.
2. Cada um dos arquivos em formato ASCII foi convertido em formato XYZ (texto) onde cada coluna representa X = longitude, Y = latitude e Z = atributo na célula da matriz. Posteriormente estes dados foram inseridos no EXCEL, sendo apagadas as colunas X e Y. Também no EXCEL, todos os dados foram concatenados. O resultado final é um arquivo texto contendo os atributos de todos os dados analisados. Cada linha representa uma célula da estrutura *raster*.
3. A execução da análise estatística, isto é, a determinação dos parâmetros desconhecidos α , β_i da equação 10, pode ser efetuada em qualquer pacote estatístico tradicional (do tipo *SPSS*, *SAS*, *Statistica*, etc.). Por facilidade de uso, optou-se neste estudo pela utilização do *SPSS* (*Statistical Package for Social Sciences*) para realização dos procedimentos de regressão logística com variáveis dicotômicas. Foi realizada uma regressão passo a passo à frente (*Forward Stepwise Regression*) onde é feita uma seqüência de modelos de regressão, em cada passo adicionando ou retirando uma variável X preditora do modelo.
4. O exame dos vários elementos de listagens de saída do *software SPSS* permite a determinação dos parâmetros α e β_i da função *logit*.

Os resultados obtidos são inerentes da classe “Transição para agricultura” (0-1). Foram efetuados sete passos, onde neste passo foi encontrado o maior nível de explicação do modelo com sete variáveis independentes mais a constante. As demais variáveis (são 16 variáveis independentes ao total, apresentadas no **quadro 2**) foram descartadas, devido à baixa contribuição destas ao ajuste do modelo com os dados observados. O resultado é:

$$\begin{aligned} \text{logit } P\left(\frac{P(X)}{1-P(X)}\right) = & 6.133 + 0.002 * BHC - 0.335 * TMA \\ & - 0.813 * D_URB - 0.002 * TOPO \\ & - 0.700 * D_ESTRADA + 0.315 * D_FERRO \\ & + 0.708 * TOPO_REC \end{aligned}$$

Resultando em:

$$P(X) = \frac{1}{1 + e^{\left(\begin{array}{l} 6.133 + 0.002 * BHC - 0.335 * TMA - 0.813 * D_URB - 0.002 * TOPO \\ - 0.700 * D_ESTRADA + 0.315 * D_FERRO + 0.708 * TOPO_REC \end{array} \right)}} \quad (11)$$

Enfatize-se que para o usuário final, a equação 11 é o modelo de alteração de uso do solo, onde $P(X)$ representa a probabilidade de transição de classes de uso do solo, neste caso “não-agrícola” → agricultura. Saliente-se, mais uma vez, a presença de (a) fatores estáveis, e (b) fatores dinâmicos como variáveis independentes na equação 11. No

caso de análise de um cenário hipotético, como uma nova estrada, esta é introduzida no processo através de um arquivo vetorial (linha) no ambiente do *ArcGIS* e estimam-se as distâncias desta linha para as demais células. Este arquivo de distâncias entra na equação 11 no termo “D_ESTRADA” sendo recalculada uma nova imagem com as probabilidades de transição para área agrícola.

RESULTADOS E DISCUSSÃO

Para explicar melhor os resultados, um exemplo de aplicação pode ser observado na **figura 10** e na **figura 11**. Na **figura 10** apresentam-se as linhas ferroviárias da região, constituídas de dois troncos principais. A Ferronorte na porção superior e o eixo que liga São Paulo, Brasil → *Santa Cruz de La Sierra*, Bolívia passando pelo meio da **figura 10**. Observa-se também a variação de tonalidades na **figura 10**, indicando a distância de cada posição no terreno para a ferrovia mais próxima. Na **figura 11** explora-se um cenário hipotético de construção de um trecho ferroviário no sentido, Norte → Sul, passando pelo divisor da bacia na porção leste, unindo os dois troncos principais.

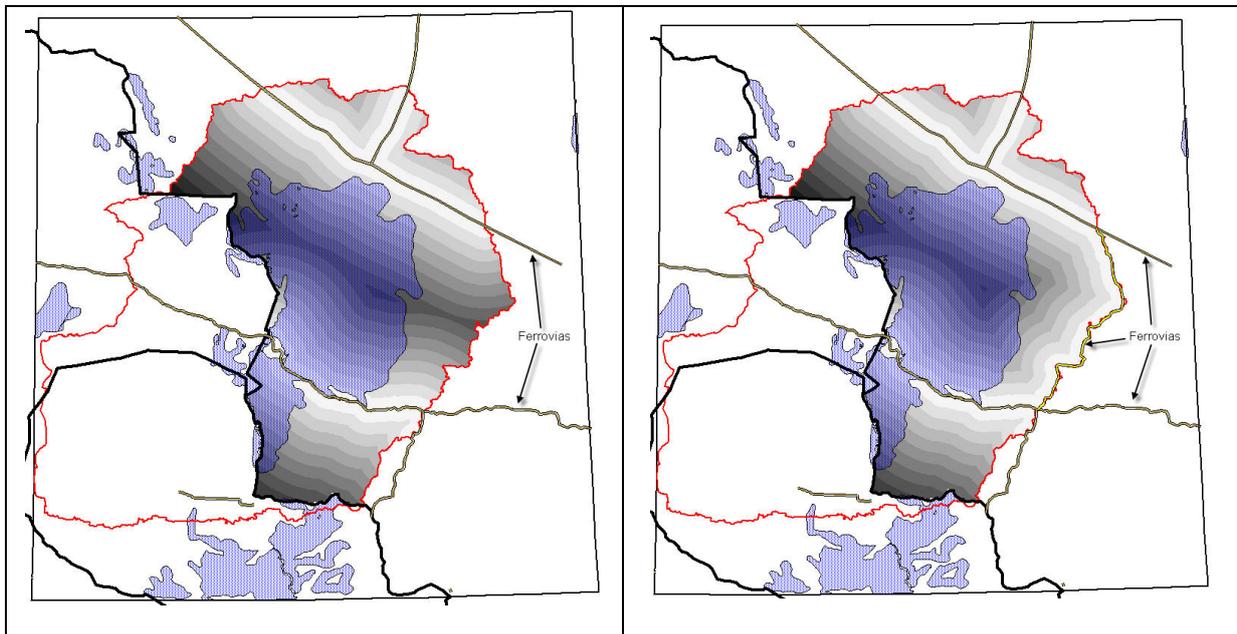


Figura 10 - Linhas ferroviárias e mapa de distâncias associadas.

Figura 11 - Cenário hipotético de construção de um trecho ferroviário.

Na **figura 12** e na **figura 13** observa-se a estimativa da probabilidade de transição das classes de uso do solo “não-agrícola” → área agrícola, utilizando-se a equação 11. A **figura 12** ilustra as probabilidades de transição no cenário atual, correspondendo à situação ferroviária indicada na **figura 10**. Já as probabilidades de

transição da **figura 13**, relacionam-se ao cenário hipotético de construção de um trecho ferroviário no divisor da bacia. Saliente-se que todas as demais variáveis da equação 11 foram mantidas constantes, modificando-se apenas a variável “distância para ferrovias” em função da proposta de construção do novo trecho ferroviário. Na **figura 10** à **figura 13** o pantanal se indica na área de cor azul.

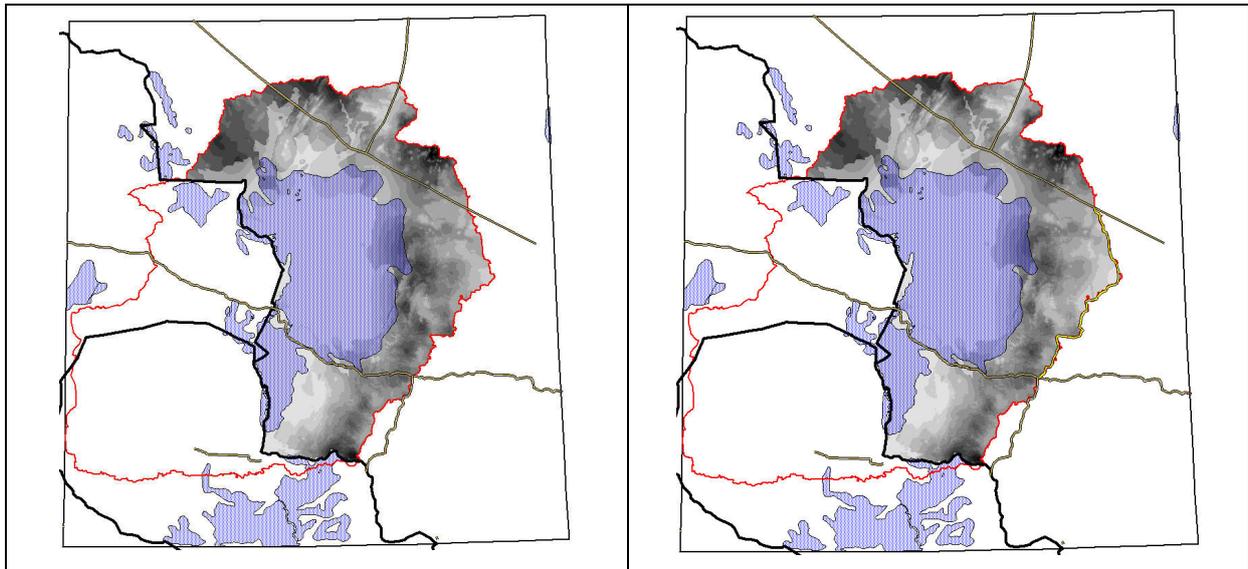


Figura 12 - Probabilidades de transição no cenário atual.

Figura 13 - Probabilidades de transição para cenário hipotético de construção de um trecho ferroviário no divisor da bacia.

Mais interessante do que as probabilidades de transição da situação atual e do cenário hipotético (**figuras 12 e 13**), são as diferenças entre os dois cenários, isto é, cenário atual (**figura 12**) – cenário hipotético (**figura 13**). A **figura 14** apresenta estes resultados onde o produto é a área de influência direta, devido à construção do trecho ferroviário com a condição *ceteris paribus*. Dito de outra forma, a **figura 14** identifica e avalia o efeito da construção do trecho ferroviário, nas probabilidades de transição do uso do solo, desconsiderando-se mudanças nas demais variáveis envolvidas. Enfatize-se que geralmente a área de influência direta de infraestrutura de transportes (rodovias, hidrovias e ferrovias) constitui-se em faixas de alguns quilômetros do eixo principal do empreendimento proposto, por exemplo: 5 km para cada lado do eixo da ferrovia. Porém o padrão espacial das modificações apresentados na **figura 14** identifica a área de influência direta de forma mais realista, constituindo-se numa das principais inovações aqui apresentadas.

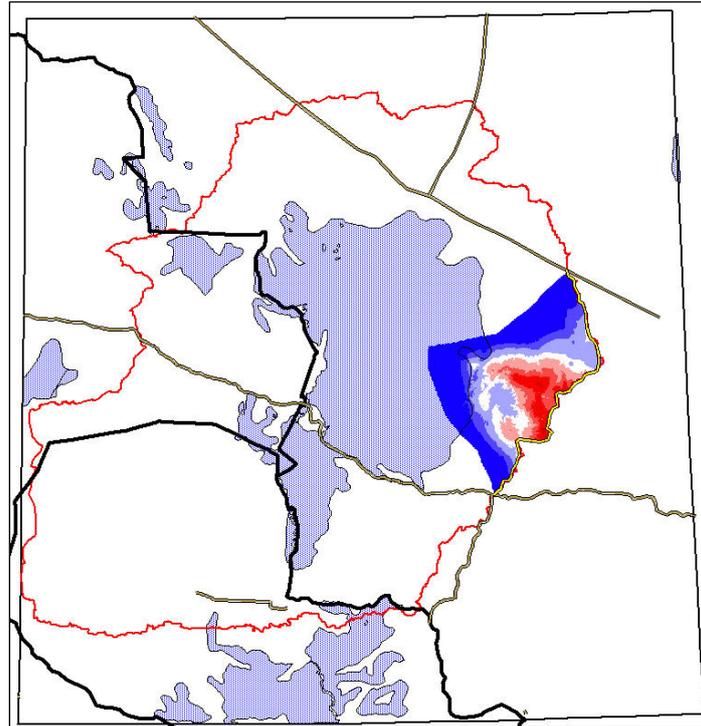


Figura 14 - Diferenças nas probabilidades de transição entre o cenário atual e hipotético, resultando na área de influência direta do empreendimento.

Apesar do caráter estatístico do método apresentado, enfatiza-se que é possível estabelecer relações causais mais determinísticas. No contexto anterior, é assumido que o desenvolvimento agrícola ocorre em áreas que são atrativas devido à sua proximidade com solos agricultáveis de boa qualidade existentes e facilidades de transporte, tais como acessos a rodovias, linhas ferroviárias e/ou rios navegáveis, que facilitam o escoamento de produção agrícola. Seguindo esses pressupostos, dessa oferta de renda bruta, cada atração para o desenvolvimento agrícola é cercada por obstáculos (uma espécie de “atrito” ao desenvolvimento). Os modelos de Von Thünen (1826) e Ricardo (1817, *apud* SOUZA, 1990, páginas 195-199) é uma das formas mais conhecidas de representação destes processos, sendo dado por:

$$L = \eta * (p - c) - \eta * t * d \quad (12)$$

Onde:

L : Lucratividade agrícola, por unidade de área produzida. (R\$/ha);

η : Produção bruta, por unidade de área produzida. (Ton./ha);

p : preço de mercado, por unidade produzida (em massa ou volume). (R\$/Ton.);

c : custo de produção, por unidade produzida (em massa ou volume). (R\$/Ton.);

t: custo de transporte por unidade distância, por unidade produzida (em massa ou volume). (R\$/km/Ton.);

d: distância percorrida em km.

Para ilustrar os conceitos da equação de Von Thünen exemplifique-se a ocorrência de uma fazenda localizada a certa distância da ferrovia. Em função da construção da ferrovia (hipotética) altera-se a estrutura de custos (e lucros) desta fazenda. A região apresenta um constante custo fixo de produção, envolvendo a custo da terra, capital e trabalho. Existe um custo variável de transporte dependente da distância a ferrovia. Quanto mais distante maior o custo de transporte. O lucro líquido [(preço de mercado – custos) * quantidade produzida] é maior quanto mais próximo da área de produção estiver localizada da ferrovia. Neste exemplo, as áreas de produção mais próximas da ferrovia lucram mais. Vale lembrar que para fins da presente análise considera-se que o embarque da produção agrícola poderia ser feito em qualquer ponto ao longo da linha férrea.

Entretanto na ótica ambiental deve-se avaliar se a área em que ocorrerão as mudanças decorrentes da construção da ferrovia, resultando em maior lucratividade para os fazendeiros da região; ocorrerão conseqüências negativas. Pode-se pensar em efeitos negativos como:

- Supressão da cobertura vegetal original,
- Uso abusivo de defensivos agrícolas,
- Surgimento de intensos processos erosivos,
- Poluição ou contaminação das águas da região,
- Manejo inadequado das áreas de pastagens,
- Queimadas,
- Práticas inadequadas de conservação do solo,
- Perda de biodiversidade,
- Introdução de espécies invasoras.

Tais efeitos deletérios podem ainda ser amplificados quando se considera a bacia hidrográfica como um mecanismo de transporte de problemas de montante para jusante, especificamente quando as mudanças de uso e cobertura dos solos e seus impactos associados ocorrem em partes altas de bacias. Enfatize-se, que a premissa básica utilizada é que a bacia hidrográfica surge como unidade ambiental de planejamento mais apropriada, pois recebe diretamente todos os impactos (positivos e negativos) decorrentes de alterações de uso do solo (que por sua vez são derivados de políticas públicas) sendo expresso na seguinte relação de causa-efeito:

1. Políticas públicas, como construção de estradas, ferrovias etc.



2. Alterações do uso do solo, como os modelos de regressão logística, por exemplo.



3. Efeitos (positivos e negativos) em bacias, analisados através de modelos ambientais.

Saliente-se que o presente artigo tratou apenas da fase 2 acima. Como consequência, a abordagem aqui utilizada integra conceitos hidrológicos, ambientais e socioeconômicos dentro da mesma estrutura de modelagem. Aqui é advogado que, a compreensão das tendências de alteração do uso do solo, e as complexas consequências decorrentes são essenciais para a garantia do desenvolvimento sustentável do Pantanal. Neste contexto a estratégia utilizada é composta de Forças Causais – C, Pressão – P, Estado – E, Impacto – I e Resposta - R, semelhante ao modelo apresentado pela Agência Européia de Meio Ambiente (EEA, 2000). As forças causais (C) são representadas por processos sociais e naturais que podem acarretar problemas ambientais como exemplo cita-se a aplicação de políticas de energia, agricultura e indústria com estudos fragmentados e inadequadas avaliações de impactos ambientais. Os indicadores de pressão (P) são os resultados da ação das forças causais que influenciam o estado ambiental atual. Os indicadores de Estado (E) descrevem os fenômenos físico, químico e/ou biológico em uma determinada área de referência. Os indicadores de Impacto (I) se referem às consequências de uma mudança de estado no ambiente. Os resultados (R) de um impacto, como poluição de solo, ar, e/ou água é seguido por muitos efeitos (alteração do regime hidrológico, perda de biodiversidade, etc.) em várias escalas temporais e espaciais. Em síntese, conhecendo-se previamente os impactos pode-se alterar os projetos executivos, minimizar e compensar de forma mais adequada à sociedade.

CONSIDERAÇÕES FINAIS

Os procedimentos de regressão logística apresentados neste artigo constituem-se em uma técnica de elaboração de cenários futuros para o ordenamento de mudanças do uso do solo. Saliente-se que tais procedimentos apenas são uma quantificação dos processos de mudanças de uso do solo e identificação dos seus principais fatores determinantes, não se constituindo em um modelo de previsão futura do uso do solo. Tal abordagem pode ser utilizada na elaboração de cenários futuros hipotéticos (do tipo “*what if?*”), estabelecendo de forma clara as relações causais entre as muitas variáveis.

Os cenários são hipóteses, coerentes e consistentes, de futuro. Saliente-se que não se deseja adivinhar o futuro, mas o que se pretende é reduzir suas incertezas para possibilitar, entre outros, um planejamento estratégico flexível para contemplar a constante mudança que marca a sociedade atual. O objetivo final deste esforço é definir o que deve ser feito agora para alcançar um futuro desejável.

Saliente-se, também, que o presente estudo não pretende fazer uma análise aprofundada dos padrões e tendências do uso do solo na Bacia do Alto Paraguai no Brasil e das medidas necessárias para promover sua sustentabilidade ambiental. Seu objetivo é apenas demonstrar aplicabilidade de técnicas de regressão logística na análise ambiental de determinado espaço geográfico.

REFERÊNCIAS

- ASPINALL, Richard. Modelling land use change with generalized linear models--a multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. **Journal of Environmental Management**. v. 72, n. 1-2, p. 91-103, ago./2004.
- BAKKER, Martha M.; GOVERS, Gerard; KOSMAS, Costas; VANACKER, Veerle; VAN OOST, Kristof; ROUNSEVELL, Mark. Soil erosion as a driver of land-use change. **Agriculture, Ecosystems & Environment**. v. 105, n. 3, p. 467-481, fev./2005.
- BERKSON, Joseph. Application of the logistic function to bio-assay. **Journal of the American statistical association**. v. 39, n. 227, p. 357-365, set./1944.
- CRAMER, Jan S. **The origins and development of the logit model**. 2003. University of Amsterdam and Tinbergen Institute. Disponível em: <http://www.cambridge.org/resources/0521815886/1208_default.pdf> Acesso em: 29 junho 2011.
- DALE, Virginia H. The relationship between land-use change and climate change. **Ecological Applications**. v. 7, n. 3, p. 753-769, ago./1997.
- EUROPEAN ENVIRONMENTAL AGENCY (EEA) **Sustainable use of Europe's water? state prospects and issues**: environmental assessment report n. 7. Copenhagen: European Environmental Agency, 2000, 36 p.
- HOSMER David W.; LEMESHOW Stanley. **Applied logistic regression**. New York: Wiley & Sons, 2000, 347 p.
- LAMBIN, Eric F.; GEIST, Helmut J.; LEPERS, Erika. Dynamics of land-use and land-cover change in tropical regions. **Annual Review of Environment and Resources**. v. 28, p. 205-241, nov./2003.
- MEYER, William B.; TURNER II, Billie Lee. **Changes in land use and land cover: a global perspective**. Cambridge: Cambridge University Press, 1994, 537 p.
- RIEBSAME, William E.; MEYER, William B.; TURNER II, Billie Lee. Modeling land use and cover as part of global environmental change. **Climatic Change**. v. 28, n. 1, p. 45-64, out./1994.
- SERNEELS, Suzanne; LAMBIN Eric F. Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. **Agriculture, Ecosystems & Environment**. v. 85, n. 1-3, p. 65-81, Jun./2001.

SOUZA, Anthony. R. **A Geography of world economy**. New York: Macmillan Publisher Company. 1990, 497 p.

STRAUSS, David. The many faces of logistic regression. **The American Statistician**. v. 46, n. 4, p. 321-326, nov./1992.

VERBURG, Peter H.; SCHOT Paul P.; DIJST Martin J.; VELDKAMP A. Land use change modelling: current practice and research priorities. **GeoJournal** v. 61, n. 4, p. 309-324, dez./2004.

WATSON, Robert T.; NOBLE, Ian R.; BOLIN, Bert; RAVINDRANATH, N. H.; VERARDO, David J.; DOKKEN, David J. (ed.) **Land use, land-use change, and forestry: a special report of the intergovernmental panel on climatic change**. Cambridge: Cambridge University Press, 2000. 375 p.

COMO CITAR ESTE ARTIGO:

MENDES, Carlos André Bulhões; CANALES VEGA, Fausto Alfredo. Técnicas de regressão logística aplicada à análise ambiental. **Geografia (Londrina)**, Londrina, v. 20, n. 1, p. 5-30, jan./abr. 2011. URL: <<http://www.uel.br/revistas/uel/index.php/geografia>>

EDITOR DE SEÇÃO:

Edison Archela

TRAMITAÇÃO DO ARTIGO:

✓ Recebido em 26/10/2010

✓ Aceito para publicação em 20/07/2011