Análise de regressão para o estudo da intensidade de degradação florestal no sudoeste do Pará

Danilo Avancini Rodrigues[1]

Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal nº 515 - 12227-010 - São José dos Campos - SP, Brasil
danilo.rodrigues@inpe.br

Abstract: Forest degradation is a long-term process that reduces forest's biodiversity and impoverishes the ecosystems. The study and modelling of forest degradation allows comprehension and inferences production about the factors that generate and intensifies forest degradation and its consequences. Through regression models it is possible to analyze certain variables behavior. This study performed a spatial regression analysis between forest degradation intensities (dependent variable) and environmental and territories variables The selection of variables to be considered on the regression model was made through the observation of R^2 and p-value of the relationship between the dependent variable and the independent variable isolated. Then, the VIF factor upon the multiple regression model demonstrated the presence of multicollinearity between the variables that were subsequently excluded from the model with the stepwise backward method. The Moran test (I = 0,2434; p-valor = 0 e z-score = 9,91) detected spatial dependency on the data with statistical significance, justifying the use of a spatial regression model. The Lagrange Multiplier test pointed out *Spatial Lag Model* as the best model adjusted to the data. The variables area of deforestation and total edge area were capable of explaining 40% (R^2-adj: 0,4092) of the forest degradation intensity.

Palavras-chave: degradação florestal, dependência espacial e análise espacial.

1. INTRODUÇÃO

Os impactos da antropização no bioma amazônico são amplamente discutidos no meio acadêmico, principalmente devido às alterações causadas à floresta, como a perda de biomassa e biodiversidade, e também a alteração nos ciclos biogeoquímicos (PEARSON, et al 2014). Estes impactos motivaram a Convenção-Quadro das Nações Unidas sobre mudanças climáticas a reconhecer, na 13º Conferência das Partes, em 2007, a exploração seletiva de madeira como um importante contribuinte para as emissões globais de carbono, passando a introduzi-la no mecanismo de Redução de Emissão por Desmatamento e Degradação Florestal (REDD+) (SHIMABUKURO, et al 2014). No mesmo ano o Instituto Nacional de Pesquisas Espaciais (INPE) passa a realizar o monitoramento de áreas em processo de degradação florestal na Floresta Amazônica brasileira através do sistema DEGRAD.

Similar ao sistema DEGRAD, o Sistema de Monitoramento da Floresta Amazônica Brasileira por Satélite (Sistema PRODES) monitora o desmatamento por corte raso na floresta amazônica, ou seja, áreas em que a fisionomia florestal primária tenha sofrido conversão completa por ações antropogênicas (INPE, 2008). Contudo, as regiões de floresta que sofrem distúrbio por fogo florestal e extração seletiva de madeira que não resultam na remoção completa do dossel florestal e, consequentemente, não são mapeadas pelo sistema PRODES, compreendem a definição de degradação florestal do sistema DEGRAD (INPE, 2008).

Fatores como o desmatamento e as dinâmicas de uso e cobertura da terra são processos conhecidamente influentes sobre a degradação, acelerando este processo e, consequentemente, todos os processos adjacentes à degradação (BALCH et al., 2011). Analisar os fatores e variáveis envolvidas no processo de alteração da cobertura florestal

a fim de compreender, qualificar e quantificar os atores envolvidos neste processo oferece subsídio à construção de modelos que permitam compreender os impactos gerados pela degradação florestal, bem como fazer predições para análise de cenários futuros.

Desta forma, o objetivo deste trabalho foi determinar a relação da intensidade de degradação florestal (PINHEIRO, 2015) com seus fatores determinantes, utilizando análise espacial de dados geográficos, para a região de Novo Progresso, no sudoeste do Pará.

2. MATERIAL E MÉTODOS

2.1. Conceito de degradação florestal

Na literatura são apresentados diversos conceitos de Degradação Florestal que variam de acordo com as causas da degradação florestal, tipo de vegetação ou objetivo do estudo (PINHEIRO, 2015), e ainda podem ser gerais, podendo ser aplicados em diferentes regiões, específicos ou operacionais. Contudo, neste estudo é abordado o conceito operacional utilizado pelo Sistema DEGRAD, que define a degradação florestal como o processo de perda da cobertura florestal pelo efeito da exploração de madeira e fogo florestal, de pelo menos 6,25ha da cobertura florestal, que não se qualifique como desmatamento pelo Sistema PRODES (INPE, 2008).

2.2. Área de estudo

A área de estudo compreende a órbita/ponto 227/65 do sensor TM/Landsat que envolve a região de Novo Progresso no estado do Pará (Figura 1), e ainda parte dos municípios de Itaituba e Altamira. Estes municípios tiveram seu surgimento associado a projetos de colonização do governo e à abertura do eixo rodoviário, especificamente a construção da rodovia BR-163. Contudo, o processo de ocupação desta região está associado à colonização espontânea (IPAM, 2006), caracterizado, neste caso, pela ocupação ilegal de terras por pequenos produtores agrícolas para a prática da agricultura de subsistência e garimpeiros dos aluviões de ouro. Mais recentemente, Novo Progresso é classificado como uma fronteira madeireira em expansão, com menos de 20 anos de história de exploração, e representa um importante polo madeireiro da Amazônia (PEREIRA et al., 2010).

2.3. Aquisição dos dados

Os dados utilizados neste trabalho foram originados de diferentes fontes: os dados de focos de calor foram produzidos e disponibilizados pelo Programa Queimadas do Instituto Nacional de Pesquisas Espaciais (INPE), os dados de desmatamento foram produzidos pelo sistema PRODES (INPE), os dados de unidade de conservação e polos madeireiros foram produzidos pelo Ministério do Meio Ambiente (MMA), os dados de terra indígena foram produzidos pela Fundação Nacional do Índio (FUNAI) e, por fim, os dados de malha viária foram disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Os dados de intensidade de degradação florestal foram gerados por Pinheiro (2015).

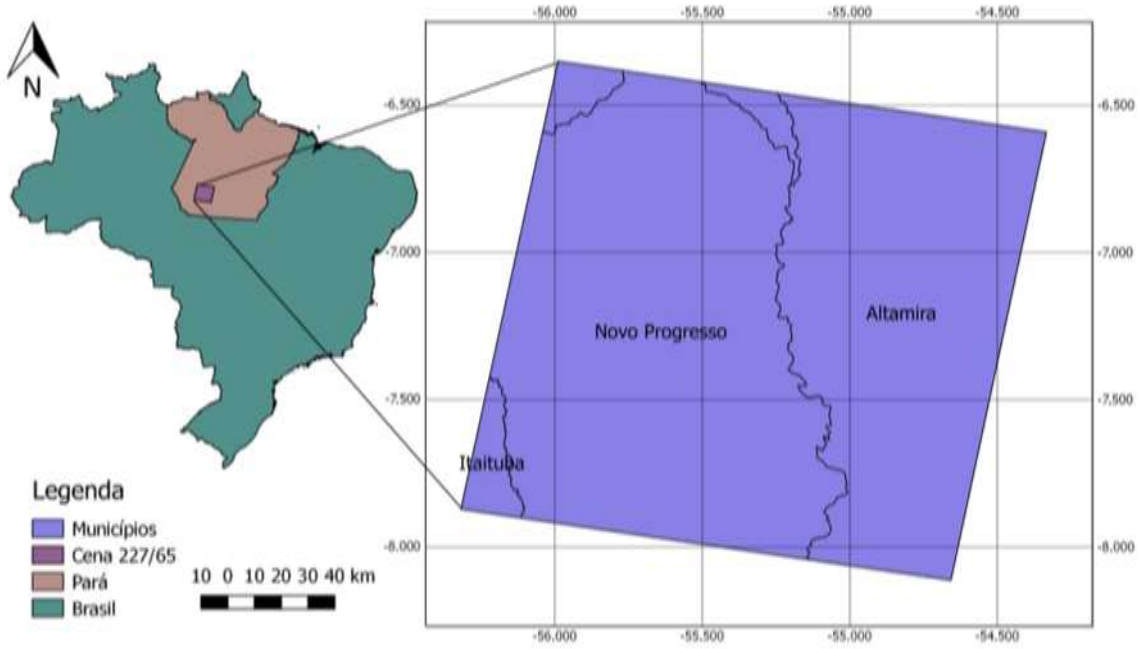

Figura 1. Localização da área de estudo (órbita/ponto 227/65 do sensor TM/Landsat) no sudoeste do Pará: Novo Progresso e região.

2.4. Preparo dos dados

2.4.1. Preparo da variável dependente

Para o preparo das variáveis (dependente e independentes) em um plano celular foi utilizado o sistema de informações geográficas TerraView v.5.2.2.

Os dados de intensidade de degradação florestal produzidos por Pinheiro (2015) compõem uma série temporal que concentra em um plano celular de 1x1 quilômetro a classificação de cada célula em intensidades que variam de 0 a 1, em que 0 representa a floresta não degradada, 1 representa corte raso, e os valores intermediários representam níveis de degradação. Uma vez que o presente trabalho busca explicar a intensidade de degradação florestal, os valores zero e 1, foram excluídos da análise. A construção da variável intensidade de degradação florestal se deu através da soma dos valores numéricos em cada célula do plano celular do ano de 2009 a 2011. A partir da somatória os valores passaram a variar de 0,1 a 2,7, representando um indicador de intensidade de degradação florestal para o período de análise.

Analisando as trajetórias da cobertura florestal em processo de degradação na região de Novo Progresso, Pinheiro (2015) constatou que aproximadamente 50% da floresta degradada é convertida em corte raso no prazo de 3 anos. Deste modo, o período de análise se restringiu ao intervalo de três anos com maior ocorrência de degradação florestal, de 2009 a 2011.

2.4.2. Preparo das variáveis independentes

A Tabela 1 apresenta a sumarização dos dados utilizados para a construção das variáveis independentes em um plano celular com as mesmas dimensões daquele utilizado para a construção da variável dependente intensidade de degradação florestal.

Através do preenchimento de células do software TerraView v.5.2.2 obteve-se um único plano celular que apresenta, em cada célula, todas as variáveis dependentes descritas na Tabela 1 e seu respectivo valor de intensidade de degradação florestal.

2.5. Modelos de regressão linear globais

2.5.1. Modelos de regressão linear simples

O modelo de regressão linear simples tem como objetivo prever a relação linear entre uma variável dependente e uma variável independente através do modelo apresentado pela equação 1 (NETER et al., 1996).

$$Yi = \beta 0 + \beta 1 \cdot Xi + \varepsilon \quad (1)$$

em que Y_i representa a variável dependente, X_i representa a i-ésima observação da variável independente, e ε uma constante que representa os erros ou resíduos.

Através dos valores de R^2 e significância estatística (p-valor) da relação entre cada variável independente com a variável independente intensidade de degradação florestal, foram selecionadas as variáveis que resultaram em maior R^2 e menor p-valor para comporem o modelo de regressão linear múltipla.

Os modelos múltiplos (ou multivariados) seguem o mesmo pressuposto do modelo simples, contudo possuem mais de uma variável independente. Desta forma, foi realizada uma regressão linear múltipla com todas as variáveis que apresentaram bom R^2 e p-valor para a verificação de multicolinearidade através do índice de variância inflacionado (VIF), significância estatística pela observação do p-valor e dependência espacial dos dados através do teste de Moran aplicado aos resíduos da regressão linear múltipla.

2.5.2 Modelo de regressão espacial

Os modelos de regressão espacial buscam resolver a limitação imposta pela dependência espacial sobre duas amostras da mesma variável aleatória que possui distribuição espaço-temporal. A autocorrelação espaço-temporal é uma característica comum em dados geográficos e leva a perda do poder explocativo da inferência estatística (ANSELIN, 2002).

Uma das maneiras de verificação da existência de dependência espacial entre as variáveis é a aplicação do teste de Moran I aos resíduos. O índice global de Moran é obtido através de um teste estatístico de normalidade em que a hipótese nula é de independência espacial ($I = 0$) contra a hipótese alternada de existência da dependência espacial ($I \neq 0$) (ANSELIN, 2002). Para o cômputo da matriz de vizinhança no teste de Moran foi utilizado o inverso da distância euclideana.

Para resolver as limitações impostas pela autocorrelação espaço-temporal, os modelos de regressão espacial incorporam os efeitos espaciais na modelagem da relação das variáveis. Na abordagem com efeitos espaciais globais, por exemplo, entende-se que é possível capturar a estrutura da autocorrelação espacial em um único parâmetro, adicionando-o ao modelo de regressão simples. Neste aspecto, o *Conditional AutoRegressive Model* (CAR, ou *Spatial Error Model*) (equação 2 e 3), que considera que os efeitos espaciais são ruídos e devem ser removidos, diferentemente, o *Spatial*

	Variável	Premissa	Fonte	Operação para preenchimento de células
Focos de calor	Número de focos de calor (**focos_n**)	Células com maior número de focos de calor possuem intensidade de degradação mais alta	Programa Queimadas (INPE)	Contagem dos focos de calor dentro da célula
	Distância dos focos de calor (**focos_dist**)	Células com menor distância aos focos de calor possuem intensidade de degradação mais alta		Menor distância euclidiana do foco à célula
	Densidade de focos de calor (**focos_dens**)	Células com maior densidade de focos de calor possuem intensidade de degradação mais alta		Moda da densidade kernel dentro da célula
Desmatamento	Área (**dsm_area**)	Células com maior área de desmatamento possuem intensidade de degradação mais alta	PRODES (INPE)	Somatória da área desmatada dentro da célula
	Distância (**dsm_dist**)	Células com menor distância às áreas desmatadas possuem intensidade de degradação mais alta		Menor distância euclidiana da borda da área desmatada à célula
	Númer de fragmentos (**dsm_n_frag**)	Células com maior número de fragmentos desmatados possuem intensidade de degradação mais alta		Somatória do número de fragmentos desmatados dentro da célula
	Total de borda (**dsm_te**)	Células com maior número de total de borda de áreas desmatadas possuem intensidade de degradação mais alta		Somatória do comprimento das bordas dos polígonos de desmatamento dentro da célula
	Densidade de borda (**dsm_ed**)	Células com maior número de densidade de borda de áreas desmatadas possuem intensidade de degradação mais alta		Densidade de borda dos polígonos de desmatamento dentro da célula
U. C. de uso sustentável	Distância (**uc_dist**)	Células com maior distância às unidades de conservação possuem intensidade de degradação mais alta	MMA	Menor distância euclidiana da borda da U.C. à célula
Terras indígenas	Distância (**ti_dist**)	Células com maior distância às terras indígenas possuem intensidade de degradação mais alta	FUNAI	Menor distância euclidiana da borda da T.I. à célula
Polos madeireiros	Distância (**po_dist**)	Células com menor distância aos polos madeireiros possuem intensidade de degradação mais alta	MMA	Menor distância euclidiana da borda do polo madeireiro à célula
Malha viária	Distância (**est_dist**)	Células com menor distância às estradas possuem intensidade de degradação mais alta	IBGE	Menor distância euclidiana da estrada à célula

Tabela 1. Resumo das variáveis independentes. Da esquerda para a direita: Variável utilizada, descrição da variável, fonte dos dados de origem à variável e operação utilizada para o preenchimento de células.

AutoRegressive Model (SAR ou *Spatial Lag Model*) (equação 4) atribui a autocorrelação espacial à variável dependente (ANSELIN, 2002).

$$Yi=\beta\cdot Xi+\varepsilon \quad (2)$$

$$\varepsilon=\lambda\cdot W\varepsilon+\xi \quad (3)$$

Na equação 3, $W\varepsilon$ é o componente do erro com efeito espacial, λ é o coeficiente autorregressivo e ξ é a componente do erro com variância constante e não correlacionada.

$$Yi = \beta \cdot Xi + \rho \cdot W \cdot Yi + \varepsilon \quad (4)$$

Na equação 4, W é a matriz de proximidade espacial e o produto $W \cdot Yi$ expressa a depend~encia espacial em Yi e ρ é o coeficiente espacial autorregressivo. Para a distinção de qual modelo utilizar, aplicou-se o teste de Multiplicadores de Lagrange (LM) (ANSELIN, 2002).

3. RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta o resultado da regressão linear simples entre a variável dependente intensidade de degradação florestal e as variáveis independentes individualmente. As variáveis independentes que apresentaram maior capacidade de explicar a variação da intensidade de degradação florestal neste modelo, representado por valores elevados de r^2, e maior significância estatística, representado pelos menores valores de p-valor, foram distância à área desmatada (dsm_dist), total e densidade de borda de área desmatada (dsm_te e dsm_ed, respectivamente), e o número de fragmentos de desmatamento (dsm_n_frag), portanto estas foram as variáveis selecionadas para comporem a análise de regressão linear múltipla.

Vár. Ind.	R^2	p-valor
focos_n	0.085	4.43e-19
focos_dens	0.081	7.29e-16
focos_dist	0.076	4.21e -21
uc_area	0.038	2.95e-03
uc_dist	0.001	1.74e-23
dsm_dist	0.224	1.78e-42
dsm_area	0.038	2.68e-27
dsm_te	0.28	3.21e-35
dsm_ed	0.288	3.16e-39
dsm_n_frag	0.292	1.49e-46
dist_est	0.017	1.12e-14
dist_ti	0.012	3.74e-08
dist_polo	0.012	8.37e-17

Tabela 2. r^2 e p-valor resultado da análise de regressão linear simples entre cada variável independente e a variável dependente intensidade de degradação florestal. Em destaque, as variáveis com maiores r^2 e valor-p (dsm_dist, dsm_te, dsm_ed e dsm_n_frag) selecionadas para comporem a análise de regressão linear múltipla.

Através do índice de variância inflada (VIF) foi constatada multicolinearidade entre as variáveis total de borda, densidade de borda e número de fragmentos de

desmatamento. Desta forma, as variáveis densidade de borda e número de fragmentos de desmatamento foram excluídas da análise com o método *stepwise backward*. Portanto, a análise de regressão múltipla (Tabela 3) foi composta pela variável dependente intensidade de degradação florestal (deg_int) e as variáveis independentes total de borda de desmatamento (dsm_te) e distância à área desmatada (dsm_dist).

As variáveis independentes apresentaram contribuição significativa para o modelo (estatística F) e foram capazes de explicar aproximadamente 29% (r^2-ajustado: 0.2976) da intensidade de degradação florestal, além de coeficientes estatisticamente significativos (p-valor) (Tabela 3).

```
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  cell_dados
Dependent Variable  :     deg_int  Number of Observations:  415
Mean dependent var  :    0.328193  Number of Variables   :    3
S.D. dependent var  :    0.213633  Degrees of Freedom    :  412

R-squared           :    0.301046  F-statistic           :     88.7262
Adjusted R-squared  :    0.297653  Prob(F-statistic)     :9.04451e-033
Sum squared residual:     13.2383  Log likelihood        :     126.012
Sigma-square        :   0.0321318  Akaike info criterion :    -246.025
S.E. of regression  :    0.179253  Schwarz criterion     :     -233.94
Sigma-square ML     :   0.0318995
S.E of regression ML:    0.178604

--------------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
--------------------------------------------------------------------------------
          CONSTANT      0.331157       0.012163         27.226   4.8023e-019
            dsm_te      0.010200       0.824274        12.4578   0.1996e-013
          dsm_dist     -1.844300       0.673641        -2.7378   1.5371e-08
--------------------------------------------------------------------------------
```

Tabela 3. Resultado da análise de regressão linear múltipla com R^2-ajustado 0.298 e coeficientes estatisticamente significativos (p-valor), com contribuição significativa das variáveis dependentes para o modelo (estatística F).

O teste de Moran I, aplicado ao resíduo do modelo de regressão múltipla, apresentou índice 0.2434, p-valor 0 (zero) e z-score 9,91 indicando haver dependência espacial com significância estatística nas variáveis utilizadas no modelo e, portanto, justificando a necessidade do uso de modelos de regressão que incorporem os efeitos espaciais dos dados na modelagem da relação entre as variáveis. O mapa de agrupamentos LIZA (Figura 2) possibilita a visualização da distribuição espacial dos resíduos gerados pelo modelo de regressão múltipla e, consequentemente, a constatação visual da dependência espacial dos dados utilizados devido a ocorrência de agrupamentos.

Uma vez que houve a constatação da dependência espacial dos dados, o teste Multiplicadores de Lagrange (LM) possibilitou a escolha do modelo de regressão espacial global mais adequado às variáveis (Tabela 4). Desta forma, a partir da análise do p-valor do Multiplicadores de Lagrange (LM) ambos os modelos CAR e lag apresentaram significância estatística, desta forma a decisão foi tomada considerando o menor p-valor do teste multiplicadores de lagrange robusto (RLM), e portanto, o mais adequado foi o modelo Spatial Lag.

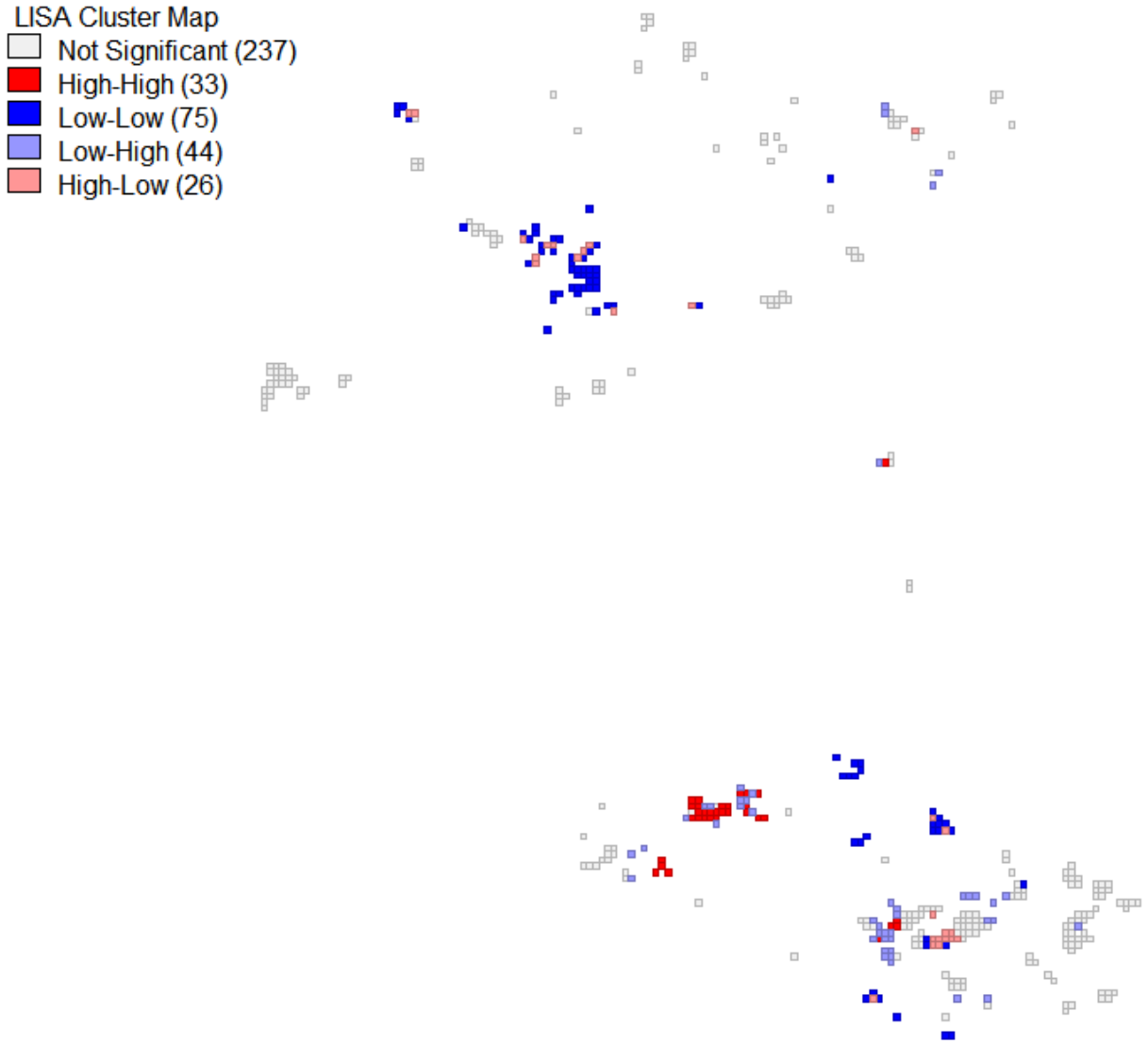

Figura 2. Mapa de resíduos da análise de regressão múltipla permitindo a visualização da ocorrência de agrupamentos de elevados e baixosvalores de intensidade de degradação florestal e, portanto, a constatação da dependência espacial dos dados.

		p-valor
LMerro	1279.3	2.2E-16
RLMerro	0.619	0.2776
LMlag	1782.4	2.2E-16
RLMlag	338.5	2.2E-16

Tabela 4. Resultados do teste Multiplicadores de Lagrange (LM) e Multiplicadores de Lagrange Robusto (RLM) para a escolha do modelo de regressão espacial melhor ajustado aos dados. Segundo o p-valor do

teste LM ambos os modelos são significativos, portanto a análise baseou-se no RLM determinando, por fim, a escola do modelo Lag.

O resultado da análise de regressão espacial utilizando o modelo Lag é apresentado na Tabela 5. Comparando o desempenho do Lag ao desempenho da análise de regressão linear múltipla (Tabela 6) é possível constatar superioridade do modelo Lag à medida que este apresentou maiores r^2-ajustado e log likelihood, e menores Akaike info criterion (AIC) e Schwarz criterion (SC).

```
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : cell_dados
Spatial Weight      : viz_nova
Dependent Variable  :     deg_int  Number of Observations:  415
Mean dependent var  :    0.328193  Number of Variables   :    4
S.D. dependent var  :    0.213633  Degrees of Freedom    :  411
Lag coeff.   (Rho)  :    0.311901

R-squared           :    0.409209  Log likelihood        :     153.839
Sq. Correlation     : -            Akaike info criterion :    -299.679
Sigma-square        :    0.026963  Schwarz criterion     :    -283.566
S.E of regression   :    0.164204

-----------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
-----------------------------------------------------------------------------
         W_deg_int     0.311901       0.0392708        7.9423     0.00000
          CONSTANT     0.233818       0.0166345       14.0562     0.00000
            dsm_te     0.896903    7.67182e-006       11.6909     0.00000
          dsm_dist    -0.695729     6.19521e-006      -2.53905    0.00000
-----------------------------------------------------------------------------
```

Tabela 5. Resultado do modelo de regressão espacial global lag com coeficientes e p-valor estatisticamente significativos.

	Regressão Linear Múltipla	Regressão Espacial (lag)
R^2 Ajustado	0.301	0.409
Log Vizinhança	126.012	153.839
AIC	-246.025	-299.679
SC	-233.94	-283.566

Tabela 6. Indicadores de qualidade do modelo de regressão linear múltipla e regressão espacial lag demonstrando superioridade do modelo lag em relação ao modelo linear múltiplo em todos os indicadores.

4. CONCLUSÃO

Na região de Novo Progresso (PA), para o período entre 2009 e 2011, as regiões de floresta mais próximas a áreas desmatadas estão mais propensas a sofrerem degradação de mais elevada intensidade, bem como regiões com a paisagem florestal mais fragmentada (aqui representadas pelo total de borda de área desmatada). Desta forma, foi possível constatar sinergia entre os processos de desmatamento e degradação florestal no período de análise para a área de estudo.

Futuramente, para que se obtenham melhores resultados em termos do poder explicativo da modelagem da intensidade de degradação florestal em Novo Progresso,

recomenda-se a introdução de outras variáveis relacionadas, por exemplo, com a expansão das áreas de pastagem, além de centros urbanos e projetos de colonização, buscar novas aburdagens para a ocorrência de incêndios florestal, dado que este fenômeno é comumente relacionado com a degradação florestal, e ainda conciliar o período de análise com momentos importantes para a região, como as etapas da colonização e planos de controle e combate ao desmatamento e incêndios florestais.

5. REFERÊNCIAS

ANSELIN, Luc. Under the hood issues in the specification and interpretation of spatial regression models. **Agricultural economics**, v. 27, n. 3, p. 247-267, 2002.

BALCH, J. K., NEPSTAD, D. C., CURRAN, L. M., BRANDO, P., PORTELA, O., GUILHERME, P., CARVALHO Jr., O. Size, species, and fire behavior predict tree and liana mortality from experimental burns in the Brazilian Amazon. Forest Ecology and Management, 2011.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). Monitoramento da cobertura florestal da Amazônia por satélites: Sistemas PRODES, DETER, DEGRAD e QUEIMADAS 2007-2008. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2008a. 47 p. Disponível em: . Acesso em: 03 out. 2015.

INSTITUTO DE PESQUISA AMBIENTAL DA AMAZÔNIA (IPAM). Avaliação ambiental integrada do território da BR-163. Itaituba: Instituto de Pesquisa Ambiental da Amazônia, 2006. 45 p.

PEARSON, T.R.H.; BROWN, S.; CASARIM, F.M. Carbon emissions from tropical forest degradation caused by logging. Environmental Research Letters, v. 9, n. 3, p. 034017, 2014.

PEREIRA, Denys et al. Fatos florestais da Amazônia 2010. Belém: Imazon, 2010. 124 p. Instituto do Homem e Meio Ambiente da Amazônia.

PINHEIRO, Taise P. Padrões e trajetórias de degradação florestal em fronteiras madeireiras da Amazônia. 17 de dezembro de 2015. 207f. Tese – Instituto Nacional de Pesquisas Espaciais.

SHIMABUKURO, Y.E.; BEUCHLE, R.; GRECCHI, R.C.; ACHARD, F. Assessment of forest degradation in Brazilian Amazon due to selective logging and fires using time series of fraction images derived from Landsat ETM+ images. Remote sensing letters, v. 5, n. 9, p. 773-782, 2014.