

Regressão geograficamente ponderada aplicada ao estudo da relação pecuária e desmatamento no sudeste do estado do Pará.

Magno Roberto Alves Macedo

Instituto de Desenvolvimento Econômico, Social e Ambiental do Pará – IDESP

Av. Municipalidade, 1461 – Marco – Belém – Pará CEP: 66050-350

magno.macedo@idesp.pa.gov.br / mrmacedo@museu-goeldi.br

Abstract: Understanding the deforestation of the Brazilian Amazon, is a challenge for many researchers, many Methodology are applied in pursuit of understanding the dynamics of this process. The livestock is widely accused of being primarily responsible for the removal of forest. The methodologies developed in the quest to understand this process, we highlight the regression models. This study aimed to compare the results of the traditional global model regression model (OLS) with a regression model that considers the spatial relationships of dependence. The results showed that when compared, the models weighted geographically, describe the spatial analysis, obtaining better results than global models, and thus contributed to the development of more reliable models. The regressions applied in southeast Pará showed a strong correlation between the cattle industry and its influence on deforestation in the region.

Palavras Chave: Deforestation, Regression Models, , Desmatamento, Modelos de Regressão

1. Introdução

Modelos de regressão são modelos matemáticos que relacionam o comportamento de uma variável Y com outra X. Quando a função f que relaciona duas variáveis é do tipo $f(X) = a + bX$ temos o modelo de regressão simples. A variável X é a variável independente da equação enquanto $Y = f(X)$ é a variável dependente das variações de X. O modelo de regressão é chamado de simples quando envolve uma relação causal entre duas variáveis. O modelo de regressão é multivariado quando envolve uma relação causal com mais de duas variáveis. Isto é, quando o comportamento de Y é explicado por mais de uma variável independente X_1, X_2, \dots, X_n .

Estes modelos são bastante utilizados para estimar o comportamento futuro de algum fenômeno da realidade. Neste caso extrapolam-se para o futuro as relações de causa-efeito – já observadas no passado – entre as variáveis, sendo permitido simular os efeitos sobre uma variável Y em decorrência de alterações introduzidas nos valores de uma variável X usando estes modelos.

Há algum tempo, os modelos estatísticos vêm influenciando e sendo influenciados pelas representações espaciais, nascendo a chamada geoestatística. O corpo metodológico deste tipo de análise consiste basicamente em um conjunto de instrumentos estatísticos que quantificam a grandeza espacial do objeto de estudo, em modelos de interpolação espacial tendo por base a sua variabilidade estrutural, e em modelos de simulação estocástica que quantificam a incerteza ligada ao fenômeno espacial (Soares, 2000).

Neste escopo, Brunson et al (1996), desenvolveram a chamada *Geographically Weighted Regression - GWR* (Regressão Geograficamente Ponderada - RGP), esta ferramenta foi desenvolvida com o intuito de se estudar fenômenos que sofrem variação de acordo com a área que é estudada, é a chamada heterogeneidade espacial (Charlton e Fotheringham, 2002).

A RGP tem como base o ajustamento de um modelo de regressão para cada ponto no conjunto de dados, ponderando as observações em razão da distância a este ponto (Carvalho et al., 2006). O fundamento é a chamada primeira lei da geografia que diz que pontos mais próximos do ponto de estudo tenham maiores influências nos parâmetros da regressão do que aqueles obtidos mais distantes (Carvalho et al., 2006), isso faz com que os resultados sejam um conjunto

Sabendo da eficiência da variável espacial nas regressões Câmara et al. (2005) nos mostra que *“A idéia é ajustar um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto. Serão feitos tantos ajustes quantas observações existirem e o resultado será um coeficiente de ajuste para cada localização. Estes coeficientes podem ser apresentados visualmente para mostrar como se comportam espacialmente os relacionamentos entre as variáveis. Esta técnica é denominada regressão ponderada espacialmente e permite, através da inferência dos coeficientes de regressão locais, estimar as variações espaciais das medidas territorializadas”*.

Estes mesmos autores ressaltam que a grande maioria dos problemas de gestão pública no Brasil depende da distribuição espacial do fenômeno subjacente, dentre estes se enquadra o desmatamento na região amazônica. É de fundamental importância conhecermos para cada região quais os fatores que influenciam na conversão da floresta amazônica em diferentes usos e tipos de ocupação.

É sabido que nos últimos anos vários estudos tiveram como objeto esclarecer o desmatamento, seus atores e processos envolvidos na região amazônica (Rivero et al. 2009), utilizando diversas metodologias que envolveram modelos econométricos, regressões etc.

Estas pesquisas, em sua maioria trabalham em cima de dados tabulares, onde executam suas expressões, e interpretam seus resultados por meio de gráficos, índices e novas tabelas. Geralmente utilizam modelos de regressão tradicional com dados sócio-econômicos, estes modelos ignoram sua localização e produzem resultados gerais que se pretendem válidos para toda a área de estudo. É preciso perceber que a maior parte das ocorrências, sejam de quais características for, se ocorridas no espaço quase que inevitavelmente apresentam entre si uma relação que depende da distância (Câmara et al., 2005).

Coadunado com esta linha de raciocínio este trabalho se propõe a mostrar a relação entre o desmatamento e a atividade pecuária no sudeste do estado do Pará utilizando dados de fontes secundárias como o INPE, o IBGE. Para tal, este estudo fez uso da estatística espacial e da geoestatística, utilizando-se de um modelo de regressão onde a variável dependente é o desmatamento, e as variáveis explicativas são dados inerentes ao desenvolvimento das atividades pecuárias, além dados de ocupação do solo, sociais etc. o que difere então estes trabalhos dos já realizados? A resposta é que neste modelo de regressão será levado em conta o componente espacial, pois todos os cálculos tomarão como referência a área sem floresta dos municípios paraenses, isso faz com que seja diminuída a generalização dos resultados destes modelos, pois apenas mostrar que a pecuária tem relação direta com o aumento do desmatamento não é o suficiente, é preciso mostrar onde ocorrem estes processos.

2. Metodologia de Trabalho

2.1. Regressão espacial

Os modelos de regressão espacial podem variar de acordo como o interesse de estudo, e podem ser aplicados com efeito espacial global e efeito espacial local. Nesse estudo foi utilizado o efeito local, visto que, quando o processo espacial é não-estacionário, os coeficientes de regressão precisam refletir a heterogeneidade espacial, neste caso o desmatamento foi modelado observando a tendência espacial de forma contínua com os parâmetros variando no espaço, isto pode ser ilustrado por:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

- Variável Dependente (y): é o fenômeno, o objeto, e o que você está querendo relacionar / modelar / prever (o desmatamento, por exemplo)
- Variáveis explicativas (x): São as variáveis que influenciam diretamente, ou que hipoteticamente ajudam a explicar a variável dependente; (rebanho bovino, Número de estabelecimentos agropecuários com bovinos etc.);
- Coeficientes (β): São os valores, calculados pela ferramenta de regressão, que indicam a relação e o peso de cada variável explicativa para a variável dependente;
- Resíduos: (ϵ): a parte da variável dependente que não é explicada pelo modelo.

2.2. Definindo a área e as variáveis de atuação do modelo

Por o tema central deste estudo ser o desmatamento no estado do Pará, nada mais óbvio que o modelo de regressão espacial tome como referência, não toda a extensão do Estado, e sim, somente as áreas onde já ocorreram os desmatamentos. Para isso foram selecionados os dados do projeto PRODES do Instituto Nacional de Pesquisas Espaciais do ano de 2005. A cada polígono de desmatamento foi calculada sua respectiva área e foi associada a identificação do município que ele faz parte. No modelo a área desmatada ($area_desm$) foi a variável dependente.

Como variáveis explicativas, foram utilizadas as informações do censo agropecuário de 2006 referentes ao número de estabelecimentos agropecuários com pastagens (EstPast), População Rural (PopRur), e efetivo total do rebanho bovino (Rebov), pessoal ocupado em estabelecimentos da agropecuária (PesOcup). O uso das informações de 2006 para tentar explicar o desmatamento de 2005, é baseado nas observações do Dr. Jonas Bastos da Veiga, especialista em pecuária na Amazônia, onde afirma que “*primeiro chega o pasto e depois chega o boi*”. O retrocesso temporal de 5 anos justifica-se pela temporalidade dos dados do censo agropecuário.

2.3. Ajustamento do conjunto de dados

Apesar da RGP ser uma ferramenta eficiente na tentativa de explicar um fenômeno, ela não deve ser automaticamente a primeira escolha de qualquer exercício de modelos de regressão. É preciso ajustar as informações e verificar o nível de importância de uma ou de um conjunto de variáveis em relação ao fenômeno. Para tal, foi utilizado modelo de regressão linear por Mínimos Quadrados Ordinários - MQO que consiste em uma técnica de otimização que procura encontrar o melhor ajustamento para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados, onde tais diferenças são chamadas resíduos. O cálculo foi realizado com as ferramentas de estatística espacial “OLS” do módulo *ToolBox* do Software ArcGis 9.3 (ERSI, 2008). O MQO foi aplicado inicialmente em cada variável explicativa, para posteriormente executar em combinações entre estas mesmas variáveis.

2.4. Autocorrelação espacial

Após o teste das variáveis, foram executados para cada operação a verificação da autocorrelação espacial entre os resíduos, este procedimento é necessário pois, sempre que os resíduos apresentarem uma estrutura de aglomeração (*cluster*) ou (*dispersion*), isso significa que no modelo ainda está faltando principais variáveis explicativas, e você não pode confiar em seus resultados. O objetivo é que seus resíduos apresentem um padrão aleatório (*random*),

se o resultado da autocorrelação for este seu modelo está no caminho certo. Neste trabalho índice de autocorrelação espacial utilizado foi o Morans I das ferramentas de estatística espacial do ArcGis.

2.5. Regressão Geograficamente Ponderada

Qualificadas as variáveis, a RGP foi calculada, o método seguiu o mesmo utilizado pelo modelo MQO, primeiro foi aplicado individualmente a cada variável explicativa posteriormente foram aplicada às mesmas variáveis associadas. Os cálculos foram executados com a ferramenta *Geographically Weighted Regression* encontrada nas ferramentas de estatística espacial do ArcGis.

3. Resultados e Discussão

3.1. Desmatamento dependente e rebanho bovino explicativa

Quando aplicado modelo de regressão MQO encontrou-se o $r^2 = 0,83$, o que é considerado um bom resultado pela teoria do modelo, qualificando a variável como de boa relação. Quando aplicado o índice de Morans aos residuais, a resposta foi um $Z\ score = 2,52$, o que colocou os resíduos em um padrão tendendo a agrupado, isso indica que neste modelo apenas o rebanho como variável explicativa, não é suficiente para dar confiabilidade às respostas encontradas.

Quando aplicada a RGP, o R-Squared (r^2) apresentou um índice de 0,93 aumentando significativamente o poder de explicação da variável, o índice de Morans quando aplicado aos residuais geograficamente ponderados, passou de um padrão quase agrupado para o padrão aleatório, o que deu mais credibilidade que há forte relação entre o número de bovinos em uma região e o desmatamento.

3.2. Desmatamento dependente e Número de pessoas ocupadas em atividade agropecuária explicativa.

Os resultados para esta relação apresentaram no método MQO um $r^2 = 0,59$ o que mostra que em relação à anterior esta variável tem um menor poder explicativo para a relação desmatamento pecuária, se comparada com a anterior. Contudo ainda é um bom resultado, por este ser superior a 0,5. Quando analisados os seus residuais, estes apresentaram um padrão aleatório com $Z\ score = 1,21$ o que infere importância a esta variável.

Já os resultados para a RGP mostraram um $r^2 = 0,64$ melhorando seu poder explicativo, contudo o comportamento dos residuais tendeu ao semi agrupado, com o $Z\ score$ subindo para 1.71 deixando essa variável bastante complexa para contribuir com um modelo de regressão confiável.

3.3. Desmatamento como dependente e Número de Estabelecimentos agropecuários com pastagem explicativa.

Apesar de aparentemente essa variável ter relação direta com o desmatamento, quando aplicado o modelo MQO o índice do r^2 foi abaixo do desejável, 0,45, contudo o residual, quando avaliado pelo índice de Morans apresentou o $Z\ Score = 1,32$ o que a coloca em padrão de aleatório. Porém as variáveis com estas características não se enquadram entre as melhores quando se buscam resultados confiáveis em um modelo,

A qualificação desta variável muda completamente quando ela é submetida a RGP, com a ponderação espacial aplicada, o r^2 sobe para 0,60 dando um grau de importância a ela no estudo pecuária desmatamento. Concomitante a isso a análise residual mostrou um excelente resultado, mostrando um padrão totalmente aleatório com um $Z\ score = 0,41$.

3.4. Desmatamento como dependente e População Rural como explicativa.

Se olhássemos esta variável comente baseado no modelo de regressão MQO, a mesma seria considerada insignificante, pois seu r^2 apresentou resultado bem abaixo de desejado apenas 0,26. Apesar de sua residual apresenta característica espacial aleatória, uma variável com estes valores não consegue explicar o fenômeno, explicitamente serão necessárias outras variáveis para dar robustez ao modelo.

Surpreendentemente, a historia muda completamente quando esta variável é inserida dentro da RGP, ela passa a ser uma variável significativa para a análise desmatamento pecuária, pois seu $r^2 =$ sobe para 0,7 e depois de ajustado ficou em 0,58. A componente espacial também melhorou a análise sobre os resíduos, passando o $Z\ Score$ para -0,2 assim permanecendo no padrão aleatório.

3.5. Primeira associação de variáveis, desmatamento, rebanho bovino e população rural.

Algumas tentativas de associação de mais de uma variável independente foram feitas, em sua maioria verificamos ainda necessitamos de outras variáveis para chegarmos a uma regressão espacial multivariada confiável, porém este estudo conseguiu aproximar as variáveis explicativas rebanho e população rural. O primeiro resultado (MQO) apresentou o $r^2 = 0,85$, estatisticamente muito bom, contudo a análise residual tendeu para a aglomeração apresentando um $Z\ Score$ de 2,05. Esta aglomeração acaba influenciar na qualidade da regressão.

Novamente com a RGP os resultados foram mais favoráveis, o r^2 subiu par 0,91 e os residuais agora aparecem com o $Z\ Score$ de 0,8 dando a estes o padrão desejável que é o aleatório.

4. Considerações finais

Sabe-se que o estudo do desmatamento e suas causas são bastante complexos por se tratar de um processo dinâmico. A necessidade dos cientistas em desenvolver ferramentas de respostas rápidas e confiáveis faz com que inúmeras metodologias se cruzem, neste intuito, os resultados mostraram que o uso da ponderação espacial em modelos de regressão que estudam as causas do desmatamento, e particularmente o papel da pecuária neste processo, é de suma importância. Pois de modo geral, quando observamos os resultados das aplicações do modelo geograficamente ponderado, constatou-se que esta é uma importante ferramenta de análise espacial de dados. Em comparação como o modelo de regressão tradicional global, o MQO, a ponderação geográfica proporcionou uma melhoria nos resultados obtidos, dando na maioria dos casos maiores significâncias às variáveis. A dependência espacial contribui sem dúvida para a redução da superestimação das previsões o melhorou a qualidade da modelagem.

É óbvio que não podemos afirmar com esta simples pesquisa qual é o papel da pecuária no desmatamento, mas podemos contribuir com uma linha de pensamento que tende a acreditar que a pecuária tem forte relação com os desmatamentos na Amazônia o que se confirma no sudeste do estado do Pará.

Referencias Bibliográficas

Brusdon, C.; Fotheringham, A. S.; Charlton, M.E. Geographically weighted regression: a method for exploring spatial nonstationarity. **Geographical Analysis**, 28(4): 281-298, 1996.

Camara, G.; Monteiro A. M.; Sposati, A.; Roman, R.; Koga, D.; Aguiar, A.P.D de. Territórios Digitais: As Novas Fronteiras do Brasil. **Seminário Temático Preparatório para a 3ª CNCTI**. Brasília, 2005.

Carvalho, L. E. X.; Silva, H. N.; Loureiro, C. F. G.; Menezes H. B. Regressão Linear Geograficamente Ponderada em ambiente SIG. **Transportes**, Volume XIV, nº 2: 18-26, 2006.

Rivero. S.; Almeida. O.; Ávila. S.; Oliveira. W.; Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento na Amazônia. **Nova Economia** nº 19 (1) 41-66,2009.

Soares A. **Geoestatística para as Ciências da Terra e do Ambiente**. Instituto Superior de Técnico, IST Press. Lisboa, Portugal, 2000.