



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Uso da regressão espacial para o estudo da expansão do cultivo de soja nos municípios de Santarém, Belterra e Mojui dos Campos - PA

Guilherme Ignácio Reis

Monografia apresentada como requisito parcial da disciplina de Análise Espacial de Dados Geográficos ministrada por Dr. Miguel Vieira Monteiro e Dr. Eduardo Camargo, do Curso de Pós-Graduação em Sensoriamento Remoto no Instituto Nacional de Pesquisas Espaciais

RESUMO

Com o início das operações das estações de transporte de grãos da Cargill no município de Santarém em 2004, verifica –se um crescimento considerável na área destinada ao cultivo da soja em Santarém e nos municípios vizinhos de Belterra e Mojuí dos campos. De acordo com a Cargill, em 2018 5% do total de grãos recebidos no terminal são produzidos nesses 3 municípios. Contudo, a entrada desses sistemas produtivos e de nova lógica de organização do território, surgem na região novos padrões de paisagens. O presente trabalho teve como objetivo principal utilizar a Geographically Weighted Regression (GWR) para analisar quais fatores podem estar associados a expansão das áreas de cultivo de soja no período entre 2004 e 2014, primeiros 10 anos de operação do terminal da Cargill. Verificou –se que as variáveis preditoras Áreas de baixio, Áreas de vertente e Áreas de floresta em 2004, Assentamentos Incra, Classe de declividade plana, Classe de declividade moderadamente ondulada, Áreas de platô e Áreas de floresta secundária e pastagem em 2004 apresentaram uma relação significativa com a variável resposta Expansão das áreas de cultivo de soja. O GWR possibilitou especializar e mapear os coeficientes betas encontrados para cada uma das variáveis preditoras, assim como os coeficientes de determinação locais

Palavras-chave: GWR. Uso e cobertura da terra. Áreas de Cultivo de Soja.

SUMÁRIO

1	Introdução.....	4
2	Área de estudo.....	5
3	Materiais e métodos	5
4	Resultados e discussões	12
5	Considerações finais.....	22
	REFERÊNCIAS BIBLIOGRÁFICAS.....	23

1 INTRODUÇÃO

A partir do início dos anos 2000, a Bacia do Curuá – Una se torna uma das regiões onde as atividades relacionadas ao cultivo de soja mais se concentram no estado do Pará. A intensificação desta atividade na região deve-se muito o início das operações do terminal de transportes de grãos da Cargill no município de Santarém em 2004. Segundo a Empresa, em 2018 5% do total de grãos recebidos no terminal são produzidos nos municípios de Santarém, Belterra e Mojuí dos Campos, todos eles inseridos nos limites territoriais da bacia.

A entrada dos sistemas produtivos intensivos associados ao cultivo de soja promoveu uma nova lógica de organização do território que modificou a paisagem da região, dando origem a novos padrões de paisagem que variam em função dos atores presentes (grandes e pequenos produtores) e das formas de produção (patronal e camponesa) Paula (2021). Além disso, antes da vigência da moratória da soja, verificava –se que parte das novas áreas de cultivo de soja se instalaram sobre áreas de floresta recém desmatada.

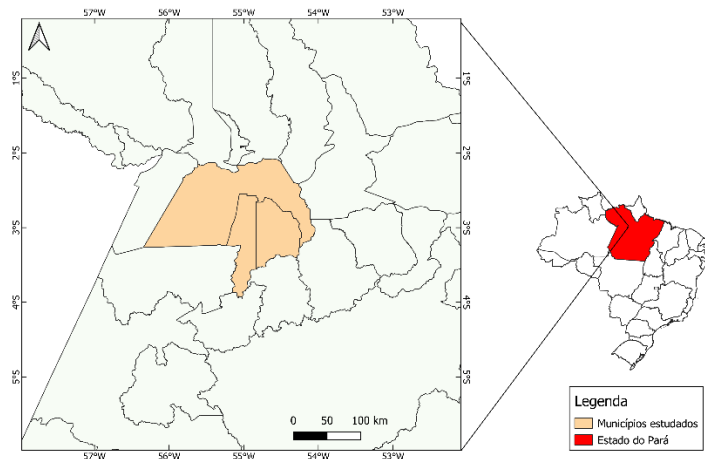
A distribuição espacial das áreas de cultivo de soja é condicionada a uma série de fatores locacionais. Na revisão bibliográfica realizada Frey et al. (2018) é apresentado que a escolha das áreas onde se instalarão as atividades de plantio de soja levam em consideração 4 fatores principais: a aptidão agrícola do local, as condições de acessibilidade até locais onde a soja é armazenada ou processada e a escolha de locais onde a conversão da cobertura da terra anterior não irá acarretar sanções comerciais.

Neste sentido, o presente trabalho tem como proposta utilizar a Geographically Weighted Regression (GWR) para identificar quais fatores podem estar associados à expansão do cultivo de soja entre os anos de 2004 e 2014 nos municípios de Santarém, Belterra e Mojuí dos Campos

2 ÁREA DE ESTUDO

O presente trabalho tem como área de estudo a os municípios de Santarém, Belterra e Mojuí dos Campos (Figura 1). Sob a área de estudo foi gerada uma grade regular de 5 por 5 km.

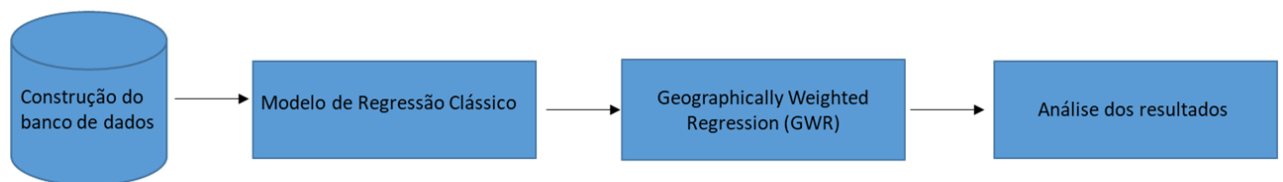
Figura 1 – Área de estudo



3 MATERIAIS E MÉTODOS

A estrutura da metodologia do presente trabalho é apresentada na Figura 2.

Figura 2 – Fluxograma da metodologia proposta



Os softwares utilizados no desenvolvimento do trabalho foram: (i) QGIS 3.16.4 para a manipulação dos dados geográficos utilizados; (ii) TerraView 5.6 para extração dos dados geográficos para a grade celular; (iii) RStudio 4.0.3 para a execução e avaliação do modelo de regressão clássico; (iv) GWR4 para a execução e avaliação do modelo de regressão espacial.

Nas seções seguintes as etapas da metodologia proposta serão apresentadas mais detalhadamente.

3.1 Preparação dos dados

A variável resposta usada no modelo foi a área de expansão do cultivo de soja entre os anos de 2004 e 2014 dentro de cada célula. Ela foi gerada a partir da diferença das áreas cultivo de soja nos anos de 2004 e 2014 fornecidas pelo dado de uso e cobertura da terra do programa TerraClass.

As variáveis preditoras foram selecionadas a partir de uma revisão bibliográfica de trabalhos que utilizaram modelos de regressão espacial para avaliar os fatores que podem estar associados a expansão de áreas agrícolas na Amazônia (ESPINDOLA et al. (2012); AGUIAR et al. (2007); FREY et al. (2018)). As variáveis preditoras selecionadas, assim como suas premissas e unidades são apresentadas na Tabela 1.

Tabela – Apresentação das variáveis preditoras utilizadas.

Variável	Premissa	Unidade	Fonte
Classe de declividade plana (0 – 3%)	De acordo com a EMBRAPA (2003) em áreas planas a mecanização agrícola é bastante viável	% da área da célula	TOPODATA
Classe de declividade moderadamente ondulada (8 – 13%)	De acordo com a EMBRAPA (2003) em áreas de declividade superior a 8% a mecanização agrícola torna-se muito inviável	% da área da célula	TOPODATA
Áreas de Platô	Os platôs são áreas mais altas, pouco íngremes, com um lençol freático profundo e solos predominantemente secos. Apresentam condições favoráveis para o cultivo de soja.	% da área da célula	Pinheiro, Rennó e Escada (2015)
Áreas de vertente	As vertentes são áreas íngremes e de transição entre platôs e baixios. Apresentam condições pouco favoráveis para o cultivo de soja	% da área da célula	Pinheiro, Rennó e Escada (2015)
Áreas de baixio	Planícies aluviais ao longo dos igarapés, onde se encontram os solos predominantemente úmidos. Apresenta condições	% da área da célula	Pinheiro, Rennó e Escada (2015)

	pouco favoráveis para o cultivo de soja		
Áreas de floresta em 2004	As áreas de florestas podem ser um impeditivo para a expansão da soja, devido a vigoração da moratória da soja a partir de 2006	% da área da célula	TerraClass
Classes de transição mais provável	De acordo com Frey et al (2018), as classes de pastagem e floresta secundária são as classes mais prováveis de serem convertidas para áreas de cultivo de soja.	% da área da célula	TerraClass
Assentamentos	Em áreas de assentamentos predominam atividades agrícolas de pequena escala	% da área da célula	INCRA
Tempo de viagem a silos graneleiros (2004)	As áreas de cultivo de soja tendem a se localizarem em locais com boa acessibilidade à essas unidades de armazenamento e processamento de grãos. Calculado a partir da metodologia proposta por SCHIELEIN et al (2021)	Tempo médio de viagem em segundos em cada célula	CONAB, IMAZON e TerraClass
Tempo de viagem a BR – 163 (2004)	As áreas de cultivo de soja tendem a se localizar próximas a BR – 163 por ser uma rodovia asfaltada e que fornece acesso facilitado ao porto de Santarém e aos silos da região. Calculado a partir da metodologia proposta por SCHIELEIN et al (2021)	Tempo médio de viagem em segundos em cada célula	IMAZON e TerraClass
Densidade de vias (2004)	As áreas de cultivo de soja tendem a se instalarem em áreas com uma boa disponibilidade de vias para facilitar o escoamento da produção	Comprimento total em metros das vias que interceptam as células	IMAZON,
Diferença do tempo de viagem a silos graneleiros (2004 – 2014)	A expansão das áreas de cultivo de soja pode ter ocorrido em regiões onde o tempo de viagem até os silos diminuiu devido a instalação de novas vias	Diferença do tempo médio de viagem em segundos em cada célula	CONAB, IMAZON, TerraClass

Diferença na densidade de vias (2004 – 2012)	A expansão das áreas de cultivo de soja pode ter ocorrido em regiões onde houve o aumento na densidade das vias	Diferença do Comprimento total em metros das vias que interceptam as células	IMAZON,
--	---	--	---------

3.2 Regressão Linear Clássica.

O modelo de regressão é uma ferramenta estatística que utiliza relacionamento existente entre duas ou mais variáveis para buscar estimar o quanto uma delas (variável resposta) pode ser descrita, ou ter o seu valor estimado, a partir das demais (variáveis preditoras). Os modelos de regressão linear podem ser genericamente descritos a partir da seguinte equação:

$$Y = \beta_0 + \beta_1 X_i + \dots + \beta_n X_{ni} + \varepsilon_i \quad (1)$$

Onde:

Y = Variável resposta em sua i -ésima observação;

β_0 = Coeficiente do intercepto no eixo vertical. Valor de Y quando $X = 0$

β_1 = Coeficiente de inclinação do modelo linear. Indica a mudança na média de Y quando X é acrescido de uma unidade.

X_i = Valor da variável preditora na i -ésima observação.

ε_i = Termo de erro (resíduo) aleatório

Normalmente, uma análise de regressão é realizada para: (i) Determinar como duas ou mais se relacionam entre si; (ii) Estimar uma função que tenta determinar a relação entre as variáveis independentes com a dependente; (iii) Estimar uma função que consiga estimar valores futuros da variável dependente. Portanto, a hipótese padrão é que as observações não são correlacionadas e que os resíduos do modelo também são independentes e não-correlacionados com a variável dependente, além de apresentarem uma variância constante uma distribuição normal com média zero (DRUCK et al., 2004). O modelo de regressão linear proposto foi desenvolvido no software RStudio com base na grade celular com as variáveis já integradas. Inicialmente foi verificado a multicolinearidade entre as variáveis preditoras selecionada para compor o modelo. A

multicolinearidade é uma condição que ocorre quando algumas variáveis preditoras no modelo estão correlacionadas entre si e que pode provocar a perda da significância do coeficiente β (beta) associado à uma das variáveis independentes do modelo. Para realizar essa verificação, foi utilizada a função Variance Inflation Factor (VIF) (Equação 2).

$$VIF = \frac{1}{1-r_k^2} \quad (2)$$

Onde r_k^2 é o coeficiente de determinação obtido da regressão entre X_k e as demais variáveis independentes. As variáveis que apresentaram um valor VIF superior a 10 foram desconsideradas, uma vez que valores acima deste limiar podem indicar uma alta multicolinearidade (MINITAB, 2019).

Em seguida, as variáveis que não apresentaram multicolinearidade foram inseridas em um modelo de regressão linear múltipla. A seleção das variáveis que compõem o modelo final foi feita a partir do uso da técnica *stepwise*. Ela consiste em adicionar e remover iterativamente as variáveis preditoras no intuito de encontrar o subconjunto de variáveis que resulta no melhor desempenho do modelo. Foi utilizado o modo *both*, onde a cada passo testa-se a entrada e a retirada de cada variável preditora. O critério para a seleção do conjunto de variáveis que resulta no modelo com melhor desempenho foi o Critério de Informação de Akaike (AIC) (Equação 3).

$$AIC = 2p + n \log\left(\frac{SQE}{n}\right) \quad (3)$$

O AIC estima a quantidade relativa de informação perdida por um determinado modelo: quanto menos informações um modelo perde, maior a qualidade desse modelo e menor a pontuação AIC. Logo, quanto menor for o valor AIC, melhor o modelo.

Após a seleção das variáveis e construção do modelo, foi realizada a análise da normalidade dos resíduos usando o teste de Bera – Jarque. O método dos mínimos quadrados usado para estimar os parâmetros dos coeficientes β parte do pressuposto que o fator de erro do modelo (resíduos) possui uma distribuição normal. Desta forma, se os

resíduos não apresentam uma distribuição normal, pode haver uma estimativa pouco confiável para os coeficientes.

Foi verificado que os resíduos não apresentaram uma distribuição normal. Buscando normalizar os resíduos, foi utilizada a transformação BoxCox na variável resposta (YU et al, 2013). Um dos pressupostos para a aplicação da transformação Box – Cox é que os dados não devem possuir valores 0. Por isso, nas células onde não foi observada a expansão do cultivo de soja, foi atribuído um valor aleatório entre 0,000001 e 0,0000019. Na sequência foi utilizado o pacote Forecast¹ presente no software R para calcular o valor λ que é usado como potência para produzir a melhor transformação de ajuste na variável resposta. Após a aplicação da transformação Box – Cox, o modelo de regressão foi refeito e os novos resíduos gerados foram analisados, sendo constatada desta vez a normalidade na distribuição.

Em uma análise de regressão envolvendo dados espaciais, onde está presente a dependência espacial, é pouco provável que o pressuposto da independência e não correlação dos resíduos seja verificado. No caso mais comum os resíduos continuam apresentando a autocorrelação espacial presente nos dados, que pode se manifestar por diferenças regionais sistemáticas nas relações do modelo, ou ainda por uma tendência espacial contínua. Desta forma, as estimativas do modelo devem incorporar a estrutura espacial do dado (DRUCK et al., 2004).

A verificação da autocorrelação espacial dos resíduos é o primeiro passo para se realizar uma regressão espacial, que leva em consideração o componente espacial dos dados geográficos. No presente trabalho, foi construída uma matriz de vizinhança *queen* entre as células da grade para o cálculo do Índice Global de Moran. Este índice retorna um resultado entre -1 e 1. Valores mais próximos de 1 evidenciam a existência de um padrão espacial com autocorreção espacial positiva, caracterizada por agrupamentos espaciais. Valores mais próximos de -1, evidenciam um padrão espacial de dispersão de que o padrão espacial é de dispersão. Já os valores próximos de 0 indicam aleatoriedade na distribuição espacial para a variável em questão (DRUCK et al., 2004).

¹ <https://cran.r-project.org/web/packages/forecast/forecast.pdf>

3.2 Geographically Weighted Regression (GWR)

Ao contrário da regressão linear clássica, a GWR incorpora a componente espacial dos dados geográfico para explorar a relação entre uma variável resposta e uma única, ou múltiplas, variáveis preditoras ao longo do espaço. Além disso, enquanto a regressão linear clássica fornece um único modelo de regressão global para toda a áreas de estudo, a GWR ajusta um modelo de regressão local para cada feição geográfica que compõem a área de estudo. Portanto, na GWR o coeficiente de determinação e os coeficientes beta continuamente no espaço (RIMBA et al.,2021).

A equação 5 apresenta a equação do modelo da GWR.

$$Y_j = \beta_0(u_j, v_j) + \sum_{i=1}^p \beta_i(u_j, v_j) x_{ij} + \varepsilon_i \quad (4)$$

Onde (u_j, v_j) são as coordenadas do local j ; $\beta_i(u_j, v_j)$ representa o coeficiente beta para a variável preditora x_i no local j ; $\beta_0(u_j, v_j)$ representa o coeficiente do intercepto; ε_i representa o termo de erro. O coeficiente $\beta_i(u_j, v_j)$ é estimado a partir da Equação 5.

$$\beta_0(u_j, v_j) = \sum_{k=1}^n w_{jk} (y_k - \beta_0(u_j, v_j) - \sum_{i=1}^p \beta_i(u_j, v_j) x_{ik}) \quad (5)$$

Onde w_{jk} representa a função de decaimento da distância entre os locais i e k . Esta função pode ser calculada a partir de um limiar de distância (Equação 6)

$$w_{jk} = \left(\frac{-d_{jk}^2}{b^2} \right) \quad (6)$$

Onde d_{jk} representa a distância entre os locais i e k , enquanto que b representa o largura de banda de uma função kernel. É necessário escolher uma largura de banda que minimize o critério de informação de Akaike do modelo proposto.

No presente trabalho, o conjunto de variáveis preditoras usadas no modelo GWR é o mesmo que foi usado na construção do modelo de regressão linear clássico proposto na etapa anterior. A função kernel utilizada foi a gaussiana adaptativa e a escolha da melhor largura de banda foi realizada a partir do método Golden Selection, que realiza sucessivos testes para encontrar o valor de largura de banda que produz o menor AIC para o modelo.

4 RESULTADOS E DISCUSSÕES

Na análise da multicolinearidade das variáveis preditoras (Tabela 2) foi verificado que a duas variáveis apresentaram colinearidade: Tempo de viagem até silos graneleiros e Tempo de viagem a BR – 163. Optou –se por desconsiderar a variável Tempo de viagem a BR – 163 e manter a variável Tempo de viagem até silos graneleiros, pois de acordo com a bibliografia as condições de acessibilidade até os silos graneleiros possui um peso maior na tomada de decisão dos agricultores na escolha de um local para iniciar o cultivo da soja.

Tabela 2 – Resultado da análise da multicolinearidade entre as variáveis preditoras

Variável	Valor VIF
Classe de declividade plana (0 – 3%)	3.286044
Classe de declividade moderadamente ondulada (8 – 13%)	6.838146
Áreas de platô vertente e baixio	3.192446
Áreas de baixio	2.671664
Áreas de vertente	4.617179
Áreas de floresta em 2004	7.251777
Áreas de floresta secundária e pastagem em 2004	5.275275
Assentamentos Incra	1.918776
Tempo de viagem até silos graneleiros (2004)	151.314688
Tempo de viagem até a BR – 163 (2004)	145.838384
Densidade de vias (2004)	1.992527
Diferença na densidade de vias (2004 – 2012)	1.577507
Diferença do tempo de viagem a silos graneleiros (2004 – 2014)	1.512329

O método spetwise retornou o conjunto de variáveis que produz o melhor modelo de regressão (Tabela 3). Foram selecionadas 8 variáveis: Classe de declividade plana (0 – 3%), Classe de declividade moderadamente ondulada (8 – 13%), Áreas de platô, Áreas de baixio, Áreas de vertente, Áreas de floresta em 2004, Áreas de floresta secundária e pastagem em 2004, Assentamentos Incra. O modelo obteve um coeficiente de determinação ajustado de 0,6188 e todas as variáveis apresentaram significância estatística (p -valor $< 0,05$).

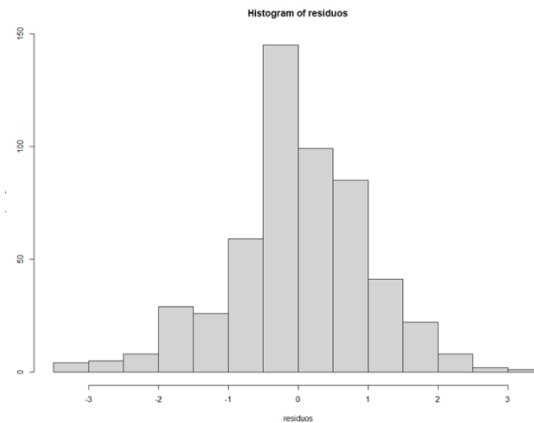
Tabela 3 – Resultado da regressão linear clássica

Variável	β Estimado	P Valor
Intercepto	0.846307	2e-16
Classe de declividade plana (0 – 3%)	0.192828	7.15e-07
Classe de declividade moderadamente ondulada (8 – 13%)	- 0.055418	0.000425
Áreas de platô	0.046322	5.67e-12
Áreas de baixio	-0.054201	1.88e-05
Áreas de vertente	-0.054079	0.002326
Áreas de floresta em 2004	-0.028370	0.002076
Áreas de floresta secundária e pastagem em 2004	0.116517	2e-16
Assentamentos Incra	-0.022639	1.44e-08

R² ajustado: 0,6188

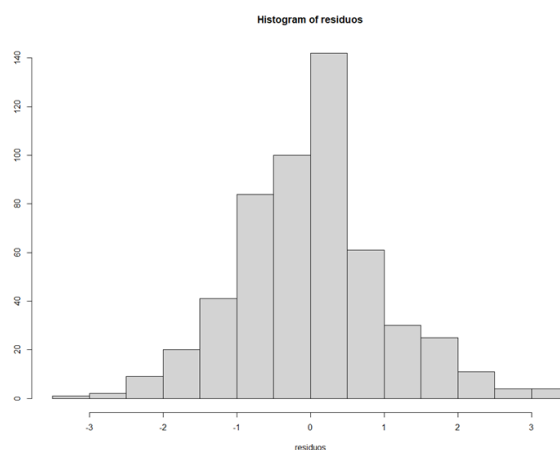
Na sequência foi realizada a análise a normalidade dos resíduos do modelo de regressão clássica. Inicialmente foi gerado um histograma para visualizar a distribuição dos resíduos (Figura 3). Em uma rápida análise visual nota –se que a distribuição não se assemelha muito a uma curva gaussiana. Em um segundo momento foi utilizado o teste estatístico Bera – Jarque para verificar se há normalidade na distribuição. Como o teste tem como hipótese nula a normalidade da distribuição e foi encontrado um p – valor de 0.0003799 no teste, a hipótese nula é rejeitada considerando um nível de significância de 0,01.

Figura 3 – Histograma com a distribuição dos resíduos



Visando tornar a distribuição dos resíduos do modelo mais próxima de uma normal, foi utilizada a transformação Box – Cox na variável resposta. O pacote Forecast retornou um valor λ de -0,343434. Após realizar a transformação na variável resposta, o modelo de regressão foi refeito e os resíduos foram novamente analisados. A Figura 3 apresenta o histograma de distribuição dos resíduos do modelo transformado. É observado que a distribuição se assemelha mais a uma curva gaussiana do que o histograma apresentado na Figura 4. Aplicando novamente o teste Bera – Jarque, verificou-se um p valor de 0.011417. Desta forma, a hipótese nula de normalidade da distribuição pode ser aceita em um nível de significância de 0,01.

Figura 4 – Histograma com a distribuição dos resíduos após a transformação Box - Cox



Após a análise da normalidade do modelo transformado, foi realizada a análise da autocorrelação espacial dos resíduos utilizando o teste de Moran Global. Foi obtida uma estatística com valor 0,734 utilizando 999 permutações, indicando uma forte autocorrelação espacial nos resíduos. Verificada a autocorrelação espacial dos resíduos, a GWR foi executada.

A Golden Selection retornou que a melhor largura de banda para o modelo proposto são os 59 vizinhos mais próximos de cada uma das células da grade. Na Tabela 4 são apresentados os valores mínimos, máximos e médios de cada um dos coeficientes betas estimados. Em relação aos valores médios dos coeficientes, verifica-se que as variáveis Áreas de baixo, Áreas de vertente e Áreas de floresta em 2004 e Assentamentos Incra apresentam coeficientes médios negativos, indicando uma influência negativa na variável resposta, ou seja, a expansão do cultivo de soja não ocorreu com tanta intensidade em regiões onde há a predominância de áreas de baixo, vertente, com presença de floresta em 2004 e áreas onde existem assentamentos rurais. Já as variáveis Classe de Declividade Suave – Ondulada, Áreas de Platô, Áreas de Declividade Plana e Áreas de floresta secundária e pastagem em 2004 apresentaram coeficientes betas médios positivos, indicando uma correlação proporcional com a variável resposta, ou seja, a expansão do cultivo de soja ocorreu principalmente em regiões de declividade plana e suave – ondulada, áreas de platô em locais onde havia a predominância de áreas de pastagem e floresta secundária. Contudo, é importante ressaltar que o coeficiente médio verificado para a variável Classe de Declividade Suave Ondulada vai contra a premissa apresentada na Tabela 1.

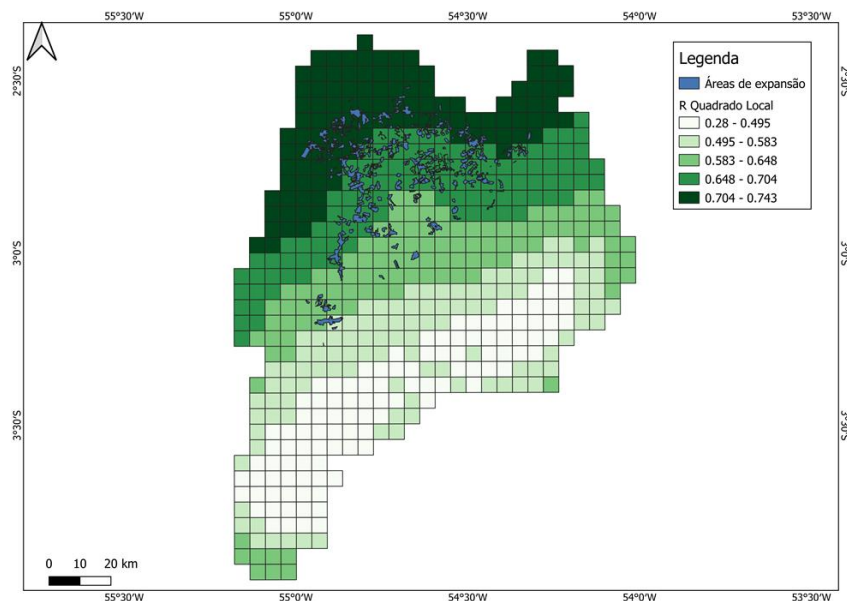
Tabela 4 – Coeficientes estimados na GWR.

Variável	β Mínimo	β Máximo	Média
Intercepto	0.827654	0.883270	0.848309
Classe de declividade plana (0 – 3%)	-0.030357	0.233622	0.118784
Classe de declividade moderadamente ondulada (8 – 13%)	-0.037221	0.091456	0.018228
Áreas de platô	-0.004993	5.67e-12	0.041486

Áreas de baixio	-0.103177	-0.008411	-0.048351
Áreas de vertente	-0.128020	0.010117	-0.049010
Áreas de floresta em 2004	-0.045652	0.031132	-0.005367
Áreas de floresta secundária e pastagem em 2004	0.052946	0.175497	0.124541
Assentamentos Incra	-0.053946	-0.002578	-0.026753

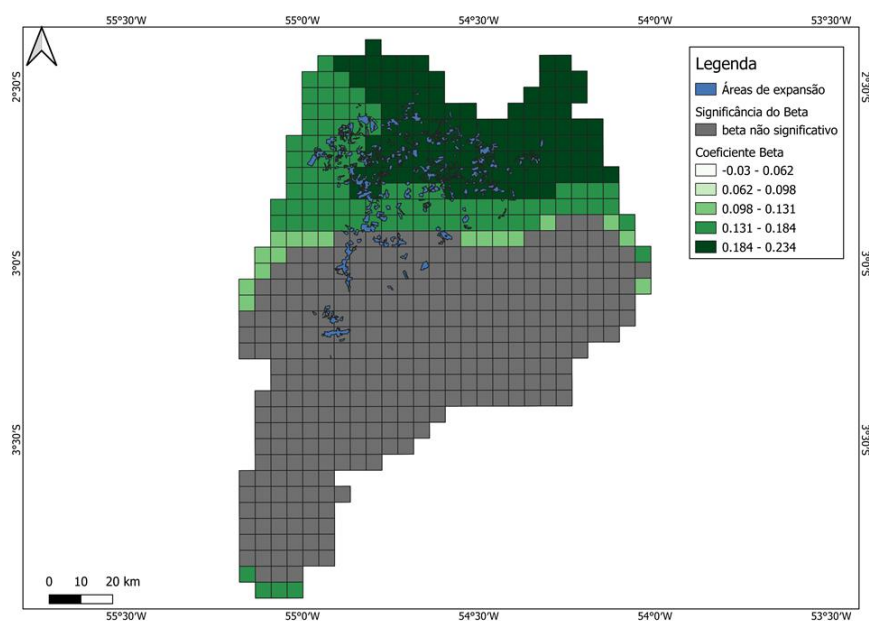
Em relação a variação do coeficiente de determinação (Figura 5), observa-se que os maiores valores encontram-se agrupados na porção norte da área de estudo, enquanto que os menores valores estão localizados na porção sul. Ou seja, os melhores ajustes encontrados para os modelos locais estão localizados na porção norte da área de estudo. Uma hipótese para esse ocorrido é que nessa região a maior parte dos coeficientes betas das variáveis preditoras são significativos. Em contrapartida, na porção sul da área de estudo verificou-se que boa parte dos coeficientes betas das variáveis preditoras não apresentou significância estatística.

Figura 5 – Distribuição do coeficiente de determinação ao longo da área de estudo.



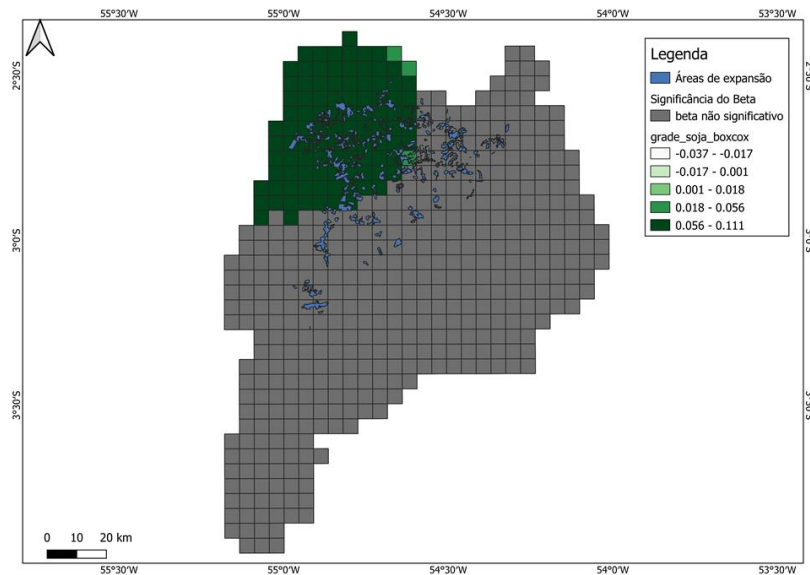
Os coeficientes betas da variável Classe de declividade plana (Figura 6) foram significantes sobretudo na região norte da área de estudo, recobrando um total de 40% da área de estudo. No geral, os valores dos coeficientes desta variável foram positivos, os maiores valores prevalecem na porção norte e nordeste do território, diminuindo em direção ao centro da área de estudo. Desta forma, percebe –se que na porção norte e nordeste da área de estudo, há forte associação positiva entre a variável preditora e a variável resposta

Figura 6 – Distribuição do coeficiente beta da variável classe de declividade plana



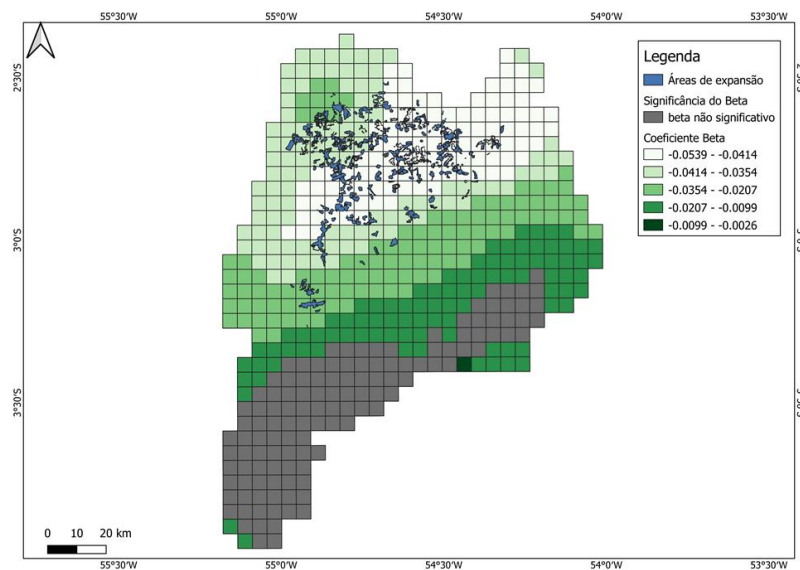
Ao analisar a distribuição espacial dos coeficientes betas encontrados para a variável suave – ondulada (Figura 7), nota-se que ela foi significativa em na porção noroeste da área de estudo, aproximadamente 9% da área total. Nesta região, os valores dos coeficientes foram altos e positivos, indo contra as premissas apresentadas para a variável, que existiria uma relação inversa entre esta variável preditora e a variável resposta.

Figura 7 - Distribuição do coeficiente beta da variável classe suave – ondulada



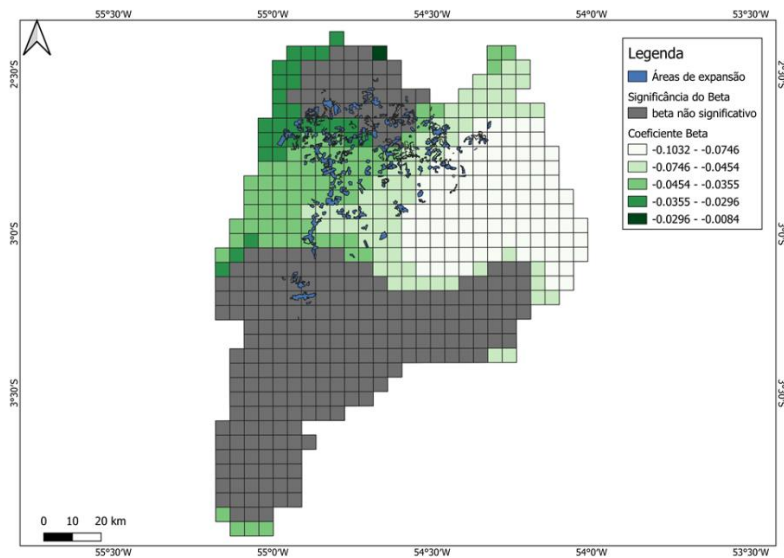
Os coeficientes betas da variável Assentamentos Inkra (Figura 8) são significantes na maior parte da área de estudo (76% da área total), com exceção da porção sul. Todos os valores são negativos, os maiores se concentram próximos ao extremo sul da área de estudo e tendem a diminuir conforme se aproximam da parte central e norte.

Figura 8 - Distribuição do coeficiente beta da variável Assentamentos Inkra



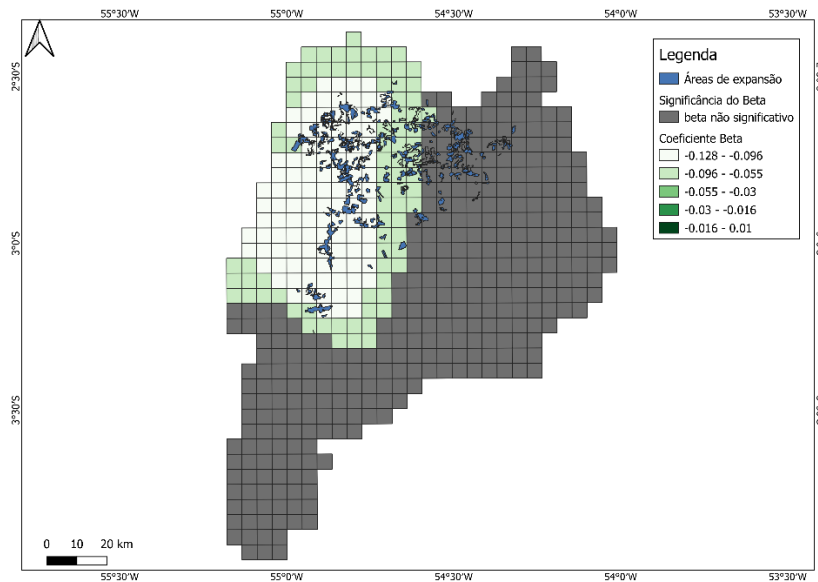
A variável Áreas de Baixo (Figura 9) apresenta coeficientes betas significativos sobretudo na porção central e nordeste da área de estudo (51% da área total). Os valores encontrados são todos negativos, os menores se encontram concentrados na porção central – leste e nordeste da área de estudo, enquanto que os maiores localizam –se na porção central – oeste e em algumas poucas células na região noroeste.

Figura 9 - Distribuição do coeficiente beta da variável Áreas de Baixo



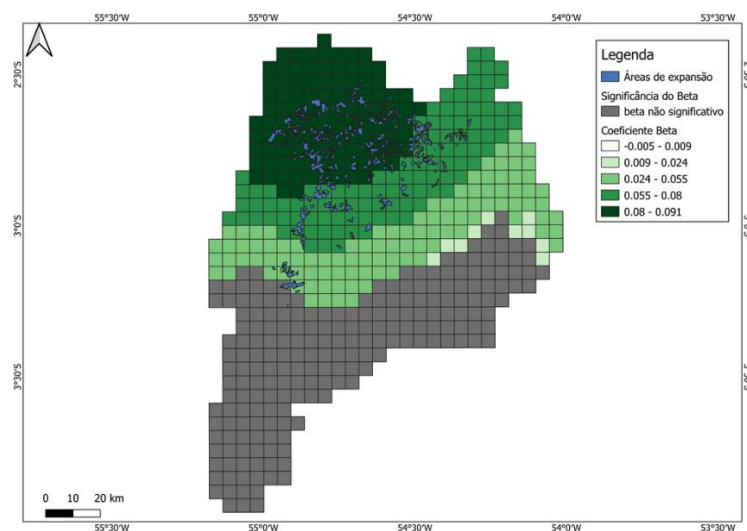
Os coeficientes betas da variável Áreas de Vertente (Figura 10) foram significativos apenas na faixa centro – oeste e noroeste da área de estudo (35% do total da área). Apresentaram valores negativos em todos os locais onde foram significativos. Nota –se que os maiores valores encontram – se agrupados na faixa central do cluster onde os coeficientes são significativos, enquanto que os valores menores encontram –se ao longo da borda do cluster

Figura 10 - Distribuição do coeficiente beta da variável Área de Vertente



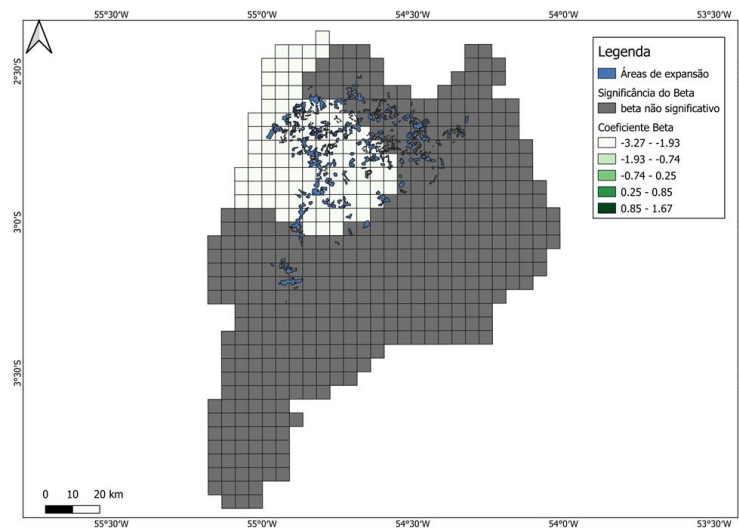
A variável Áreas de Platô (Figura 11) apresenta uma forte correlação positiva com a variável resposta na porção norte da área de estudo. Observa-se que o valor do coeficiente beta tende a diminuir progressivamente conforme se desloca para o centro da área de estudo. A variável se mostrou significativa sobretudo nas poções norte e central (61% da área total).

Figura 11 - Distribuição do coeficiente beta da variável Área de Platô



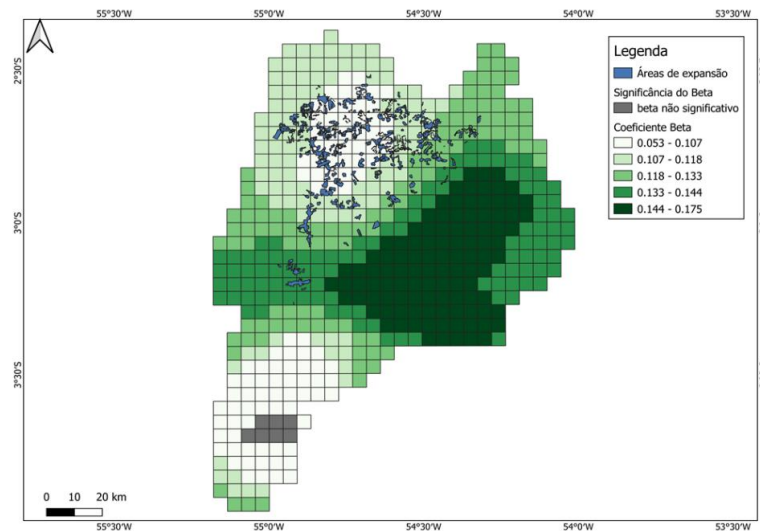
A localização da distribuição dos coeficientes (Figura 12) betas significativos da variável Áreas de Floresta em 2004 foi bastante semelhante ao observado na variável Classe de Declividade Suave – Ondulada, se concentrando sobretudo na região noroeste da área de estudo. Foram verificados altos valores negativos dos coeficientes betas no local onde eles foram significativos.

Figura 12 - Distribuição do coeficiente beta da variável Áreas de Floresta



A variável Áreas de Pastagem e Floresta Secundária (Figura 13) foi a que apresentou a maior porcentagem de área total com coeficientes betas significantes, 98% da área total. Todos os valores encontrados são positivos, os maiores se concentram na porção centro-leste da área de estudo e tendem a diminuir conforme se deslocam em direção a porção sul e norte.

Figura 13 - Distribuição do coeficiente beta da variável Áreas de pastagem e floresta secundária em 2004



5 CONSIDERAÇÕES FINAIS

A espacialização dos coeficientes betas das variáveis com a GWR permitiu analisar quais variáveis preditoras possuem uma significância mais abrangente na área de estudo, assim como visualizar como os valores variam ao longo do espaço. Ao utilizar um modelo de regressão espacial local, pode-se ter uma compreensão melhor do efeito da associação das variáveis preditoras selecionadas com a variável resposta do que com um modelo de regressão espacial global. Um modelo de regressão espacial global não possibilitaria a visualização dos locais onde as variáveis são significantes ou não.

Observou-se que as variáveis Áreas de Floresta, Declividade Suave Ondulado, Declividade suave e Áreas de Vertente, apresentaram uma significância bastante restrita na área de estudo. Já as variáveis foram significantes em mais de 50% da área total, apresentando um bom potencial de utilização para análise da expansão das áreas de cultivo de soja na área de estudo como um todo.

REFERÊNCIAS BIBLIOGRÁFICAS

PAULA, D. S. Expansão da agricultura de larga escala na Bacia do Rio Curuá-Una (PA): implicações ambientais e socioeconômicas. 2021. 145p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2021.

YU, D.L.; PETERSON, N.A.; REID, R.J. Exploring the impact of non-normality on spatial non-stationarity in geographically weighted regression analyses: tobacco outlet density in New Jersey. *GIScience & Remote Sensing*, v.46, n. 3, pp. 329-346, 2009. Disponível em: <https://www.tandfonline.com/doi/abs/10.2747/1548-1603.46.3.329?casa_token=9wQPAqnuCn0AAAAA:WQA5pkNiMYYYVvM3PbFDpxg28-ciq9wrv4HSQNA-ACffuNG3-OVUxepytsUvUVHzO-KRdEMZeNzl9w>. Acesso em 10 dez. 2021

AGUIAR, A. P. D.; CÂMARA, G.; ESCADA, M. I. S. Spatial statistical analysis of land-use determinants in the brazilian amazonia: Exploring intra-regional heterogeneity. *Ecological modelling*, Elsevier, v. 209, n. 2, p. 169–188, 2007.

DRUCK, S.; CARVALHO, M.; CÂMARA, G.; MONTEIRO, A. M. V. Análise espacial de dados geográficos. Brasília: Embrapa, 2004.

ESPINDOLA, G. M. D.; AGUIAR, A. P. D. D.; PEBESMA, E.; CÂMARA, G.; FONSECA, L. Agricultural land use dynamics in the brazilian amazon based on remote sensing and census data. *Applied Geography*, Elsevier, v. 32, n. 2, p. 240–252, 2012.

FREY, G.P.; WEST, T.A.P.; HICKLER, T.; RAUSCH, L.; GIBBS, H.K.; BÖRNER, J. Simulated Impacts of Soy and Infrastructure Expansion in the Brazilian Amazon: A Maximum Entropy Approach. *Forests*, v. 9, n.10, 2018, <https://doi.org/10.3390/f9100600>.

RIMBA, A. B., MOHAN, G., CHAPAGAIN, S. K., ARUMANSAWANG, A., PAYUS, C., FUKUSHI, K. Impact of population growth and land use and land cover (LULC) changes on water quality in tourism-dependent economies using a geographically weighted regression approach. *Environmental Science and Pollution Research*, v. 28, n. 20, p.25920–25938, 2021. doi:10.1007/s11356-020-12285-8.

MINITAB. Basta! Lidando com a multicolinearidade na análise de regressão, 2019. Disponível em: < <https://blog.minitab.com/pt/basta-lidando-com-a-multicolinearidade-na-analise-de-regressao> >.

PINHEIRO, T. F.; RENNÓ, C. D.; ESCADA, M. I. S. Mapeamento de ambientes de

terra firme no Distrito Florestal Sustentável da BR-163 utilizando o descritor de terreno HAND. In: PINHEIRO, T. F.; RENNÓ, C. D.; ESCADA, M. I. S. Distrito Florestal Sustentável da BR-163: dinâmicas sociais, mudanças ambientais e produção florestal. Belém: MPEG, 2015.

SCHIELEIN, J. et al. The role of accessibility for land use and land cover change in the Brazilian Amazon, v. 1, n. 11, p. 657-664, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0143622821000357>>. Acesso em: 17 Jul. 2021