

IDENTIFICAÇÃO DE FATORES DETERMINANTES NO CULTIVO DE CACAU POR MEIO DE REGRESSÃO ESPACIAL NO PROJETO DE ASSENTAMENTO ITATÁ, PARÁ

Camila Totti

Programa de Pós-Graduação em Sensoriamento Remoto (PGSER), Coordenação de Ensino, Pesquisa e Extensão (COEPE), Instituto Nacional de Pesquisas Espaciais (INPE), Av. dos Astronautas 1758, São José dos Campos - SP, 12227-010. camila.andrade@inpe.br

RESUMO

O cultivo de cacau (*Theobroma cacao*) desempenha um papel central na economia do Pará e tem apresentado uma expansão recente considerável no Projeto de Assentamento Itatá, localizado no sudoeste do estado. Devido à importância econômica e ambiental dessa cultura, o mapeamento das áreas de cultivo torna-se essencial. Contudo, desafios persistem, como a semelhança espectral entre o cacau cultivado em sistemas agroflorestais e as florestas primárias e secundárias, além da constante cobertura de nuvens em regiões tropicais. Para contornar essas dificuldades, variáveis geográficas contextuais, como distância às estradas, uso e cobertura do solo e declividade, têm sido utilizadas para identificar áreas potenciais de cultivo. Este estudo investiga se a importância dessas variáveis se mantém uniforme ao longo da área de estudo, incorporando o conceito de dependência espacial. Modelos de regressão espacial foram aplicados para avaliar a variabilidade espacial da relação entre as variáveis geográficas e o cultivo de cacau. Dois modelos foram comparados: o modelo global *Conditional Autoregressive Model* (CAR), que assume relações constantes em toda a área, e o modelo local *Geographically Weighted Regression* (GWR), que permite variações locais nos coeficientes. Os resultados indicaram que o modelo GWR apresentou melhor desempenho em comparação ao modelo CAR, com maior precisão (R^2 e AIC) e maior capacidade de capturar a heterogeneidade espacial das relações. Entre as variáveis analisadas, Distância às Estradas e Uso e Cobertura do Solo destacaram-se no modelo local, exibindo influência variável sobre a probabilidade de cultivo de cacau. A análise reforça a importância de considerar a variabilidade espacial ao modelar dinâmicas de uso da terra.

Keywords – cocoa mapping, data cube, time series, Sentinel-2, Pará.

1. INTRODUÇÃO

O cultivo do cacau (*Theobroma cacao*) no Brasil teve suas origens no estado do Pará, onde as condições climáticas e de solo favoráveis, especialmente na região sudeste do estado, impulsionaram o desenvolvimento dessa cultura. Entre o final do século XIX e o início do século XX, o cacau passou por uma significativa expansão territorial, substituindo monoculturas de cana-de-açúcar e avançando sobre áreas florestais. Esse crescimento foi amplamente impulsionado pela chegada de migrantes provenientes de outras regiões do país, que supriram a alta demanda por mão de obra no cultivo dessa *commodity* agrícola [1–4].

Nas últimas décadas, o Pará consolidou sua posição de liderança como o maior produtor de cacau do Brasil, respondendo por mais de 50% da produção nacional e gerando um valor bruto de produção estimado em cerca de R\$ 1,8 bilhão [5, 6]. Entre os anos de 1990 e 2016, a área cultivada no estado cresceu mais de 100%, enquanto a Bahia, tradicional polo produtor, registrou uma redução de aproximadamente 10% no mesmo período [7]. Esse protagonismo no cenário nacional se fortalece ainda mais por meio de iniciativas voltadas à sustentabilidade, como o Projeto Cacau 2030, que busca, por meio de investimentos e práticas agrícolas mais eficientes, ampliar a produção anual para além de 400 mil toneladas até 2030, alinhando-se aos Objetivos de Desenvolvimento Sustentável (ODS) [8].

Entretanto, em regiões amazônicas, como o Projeto de Assentamento Itatá, localizado no sudoeste do Pará, intensas transformações no uso e cobertura da terra têm sido observadas nos últimos anos, impulsionadas, principalmente, pela conversão de áreas de floresta primária em categorias de uso voltadas à agricultura perene, com destaque para o cultivo do cacau [9]. Esse fenômeno está associado, em grande parte, ao deslocamento de ex-meeiros¹ de plantações de cacau de municípios vizinhos, como Altamira, para o interior da área de assentamento [5] (Comunicação pessoal, INPE, 2024).

Apesar da importância econômica e ambiental do cultivo do cacau, o mapeamento preciso das áreas cultivadas ainda representa um desafio significativo. A principal dificuldade reside na semelhança espectral entre as plantações de cacau cultivadas em sistemas agroflorestais, que possuem sombreamento, e as áreas de florestas primárias e secundárias. Além disso, a cobertura persistente de nuvens durante os períodos chuvosos, característica comum em regiões tropicais, compromete a qualidade das imagens de satélite, dificultando a aplicação de técnicas tradicionais de sensoriamento remoto [11, 12].

Como alternativa para contornar esses desafios, têm sido utilizadas variáveis contextuais na identificação de áreas potenciais de cultivo de cacau. Variáveis geográficas como a distância a estradas, o uso e cobertura do solo, a declividade, entre outras características relacionadas ao contexto regional, são citadas na literatura por apresentarem relação com a ocorrência dessa cultura [13–15]. No entanto, a relação entre essas variáveis explicativas e a presença do cacau pode não ser uniforme em toda a área de estudo. Sob a perspectiva da dependência espacial, baseada na Primeira Lei da Geografia, que afirma que “tudo está relacionado a todo o resto, mas as coisas próximas estão mais relacionadas do que as distantes” [16], questiona-se se essas variáveis mantêm a mesma dimensão e hierarquia de importância ao longo da área estudada. Essa possível variação espacial torna-se crucial para a compreensão das dinâmicas geográficas que influenciam a distribuição das áreas cultivadas.

Para responder a essa questão, modelos de regressão espacial surgem como ferramentas adequadas, pois permitem incorporar a dependência espacial nas análises. Entre esses modelos, destacam-se os globais e os locais. Os modelos globais, como o *Conditional Autoregressive Model* (CAR), consideram a dependência espacial, mas partem do pressuposto de que a relação entre as variáveis explicativas e a variável dependente é constante em toda a área de estudo [17]. Em contraste, os modelos locais, como o *Geographically Weighted Regression* (GWR), permitem que os coeficientes variem espacialmente, capturando diferenças locais na

¹Meeiros de cacau são famílias que cultivam a planta em terras de outros, os proprietários. Essa prática, comum em algumas regiões produtoras, envolve um acordo informal em que a colheita é dividida igualmente entre o meeiro e o dono da terra. [10]

importância das variáveis [18].

Diante desse contexto, o presente estudo tem como objetivo principal avaliar a variação espacial da importância das variáveis geográficas relacionadas ao cultivo de cacau no Projeto de Assentamento Itatá, no sudoeste do Pará. Para tanto, serão utilizadas técnicas de regressão espacial, com ênfase na comparação entre o desempenho do modelo global CAR e do modelo local GWR na representação das relações entre as variáveis geográficas e a presença de cacau. A análise proposta visa contribuir para uma compreensão mais detalhada das dinâmicas espaciais que influenciam o cultivo de cacau na região, oferecendo subsídios para aprimorar as metodologias de mapeamento em áreas tropicais e para orientar políticas de uso da terra e gestão sustentável da agricultura.

2. MATERIAL E MÉTODOS

O município de Senador José Porfírio possui uma extensão territorial de 14.419,916 km² [19], limitando-se ao norte com o município de Altamira, que desempenha um papel importante na produção estadual de cacau [6, 20]. O Projeto de Assentamento Itatá, localizado no município de Senador José Porfírio, no sudoeste do estado do Pará, compreende uma área de 52.523,31 hectares no bioma Amazônia [21]. A região é composta por um mosaico de florestas primárias, florestas secundárias, áreas de agricultura perene e pastagens, sendo representativa das transformações no uso e cobertura do solo típicas dessa porção do território paraense [9]. A área de estudo (Figura 1) inserida em um contexto de expansão da cacauicultura e modificações na paisagem amazônica, oferece um cenário adequado para análises espaciais relacionadas ao cultivo de cacau e às dinâmicas associadas ao uso da terra.

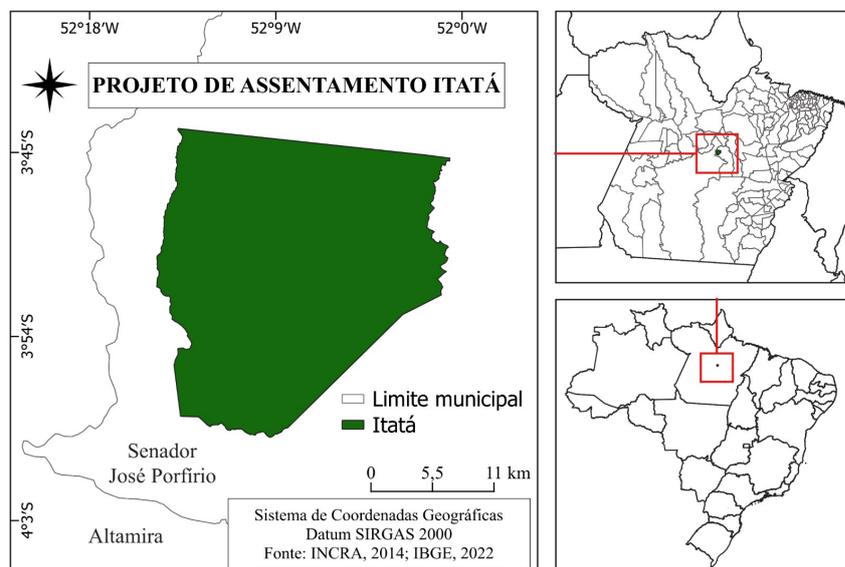


Figura 1: Mapa de localização do Projeto de Assentamento Itatá - Pará.

A metodologia do presente trabalho foi estruturada em três etapas principais (Figura 2). Primeiramente, realizou-se a seleção e preparação dos dados, incluindo a verificação da presença de dependência espacial e a definição do tamanho da matriz de vizinhança a ser utilizada no processo de regressão espacial. Na segunda etapa, foram gerados os modelos de

regressão espacial global e local, com destaque para a aplicação do teste de Lagrange (LM) e dos modelos CAR e GWR. Por fim, na terceira etapa, procedeu-se à comparação dos modelos e à avaliação da variação da importância das variáveis ao longo do Projeto de Assentamento (PA) Itatá, utilizando testes de significância e a obtenção dos coeficientes locais.

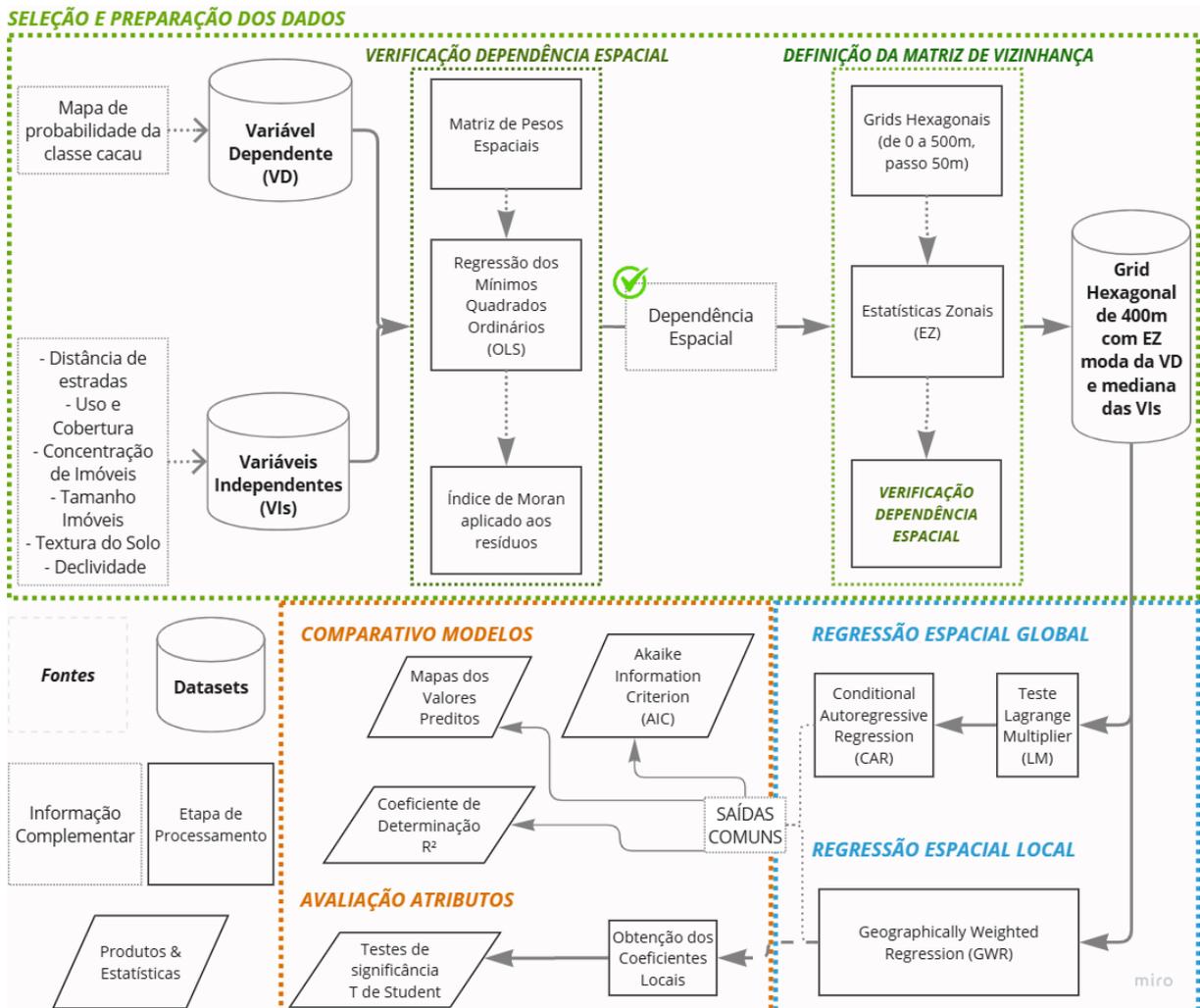


Figura 2: Ilustração do processo metodológico empregado neste trabalho.

2.1. Seleção das variáveis

Neste estudo foi utilizado o dado matricial de probabilidade da classe Cacau oriundo do trabalho de Totti [22], onde foi realizada uma classificação supervisionada com Random Forest utilizando imagens Sentinel-2 (óptico) e uso de índices de vegetação e medidas texturais de matriz de co-ocorrência de nível de cinza. A acurácia global do modelo foi de 93,2% e as acurácias do produtor e do usuário para a classe Cacau foram respectivamente de 90,1% e 86,2%.

As variáveis independentes selecionadas para aplicação da análise de regressão espacial foram derivadas do trabalho de Totti [23], onde o Processo de Análise Hierárquica (AHP) foi aplicado na identificação de áreas potenciais de ocorrência de cultivo de cacau no PA Itatá. No trabalho foram utilizadas 6 variáveis, sendo estas: Uso e cobertura da terra [9]; Declividade [24]; Textura do Solo [25]; Tamanho dos imóveis registrados [26]; Concentração de estabelecimentos rurais [27]; Distância à estradas [28]. O critério de seleção das

variáveis envolveu consulta com especialistas do Laboratório de Investigação em Sistemas Socioambientais (LiSS) e de literatura referente ao tema [13, 14]. As variáveis foram normalizadas em uma escala de 1 a 4 e foram atribuídos pesos que podem ser observados na Tabela 1.

Critérios	Pesos
Distância Estradas	0.293
Uso e Cobertura	0.278
Concentração de Imóveis	0.227
Tamanho Imóveis	0.120
Textura Solo	0.042
Declividade	0.041

Tabela 1: Peso dos critérios de potencial para o cultivo de cacau. Adaptado de [23].

A fórmula resultante da variável resposta e das variáveis independentes é apresentada a seguir:

$$VD_{CocoaVotes} = VI_{LULC} + VI_{TextSolo} + VI_{DistEstradas} + VI_{TamImveis} + VI_{Conc.Imveis} + VI_{Decli} \quad (1)$$

Na Fórmula 1, $VD_{CocoaVotes}$ representa os votos para a classe cacau obtidos a partir do modelo Random Forest. As variáveis independentes são compostas por VI_{LULC} que corresponde em inglês ao *Land Use Land Cover*, sigla em inglês para Uso e Cobertura do Solo; $VI_{TextSolo}$ referente à textura do solo; $VI_{DistEstradas}$, que representa a distância às estradas; $VI_{TamImveis}$ associado ao tamanho dos imóveis; $VI_{Conc.Imveis}$ referente à concentração de imóveis; e VI_{Decli} que corresponde à declividade.

2.2. Definição da matriz de vizinhança

A definição da matriz de vizinhança foi realizada com base na análise de autocorrelação espacial dos resíduos, utilizando o Índice de Moran aplicado à Regressão dos Mínimos Quadrados Ordinários (*Ordinary Least Squares - OLS*). O Índice de Moran é uma medida estatística que avalia o grau de autocorrelação espacial em um conjunto de dados, indicando se valores semelhantes estão espacialmente agrupados. O índice está descrito na Fórmula 2, onde e é o vetor de resíduos, e W a matriz de pesos espaciais. Ele varia de -1 a 1, onde valores próximos de 1 indicam alta autocorrelação positiva (valores semelhantes próximos), valores próximos de -1 indicam autocorrelação negativa (valores diferentes próximos) e valores próximos de 0 indicam ausência de padrão espacial.

$$I = \frac{e'W e}{e'e} \quad (2)$$

A regressão OLS, descrita na Fórmula 3, contém a variável dependente (resposta) Y_i , a variável independente (explicativa) X_i , o coeficiente angular na reta β_1 , o intercepto β_0 e o erro ou termo residual ξ_i .

$$Y_i = \beta_0 + \beta_1 X_i + \xi_i \quad (3)$$

Inicialmente, verificou-se a existência de autocorrelação espacial com a resolução espacial original das variáveis no modelo. Constatada uma forte autocorrelação espacial, desenvolveu-se um script em linguagem R para determinar o maior tamanho de grade que preservasse essa relação sem comprometer a representatividade dos dados. O script seguiu os seguintes passos: (1) criação de grades hexagonais com tamanhos variando de 50 a 500 metros, em intervalos de 50m; (2) extração de estatísticas zonais, sendo a moda aplicada à variável resposta e a mediana às variáveis independentes; (3) geração de uma matriz de vizinhança para cada grade com estatísticas zonais; (4) geração de uma matriz de pesos espaciais para cada grade com estatísticas zonais; (5) aplicação de regressão por mínimos quadrados ordinários em cada grade; (6) Aplicação do índice de Moran aos resíduos do modelo para verificação da presença de autocorrelação espacial dos resíduos. Após essa análise, a grade hexagonal de 400 metros foi selecionada como a mais adequada.

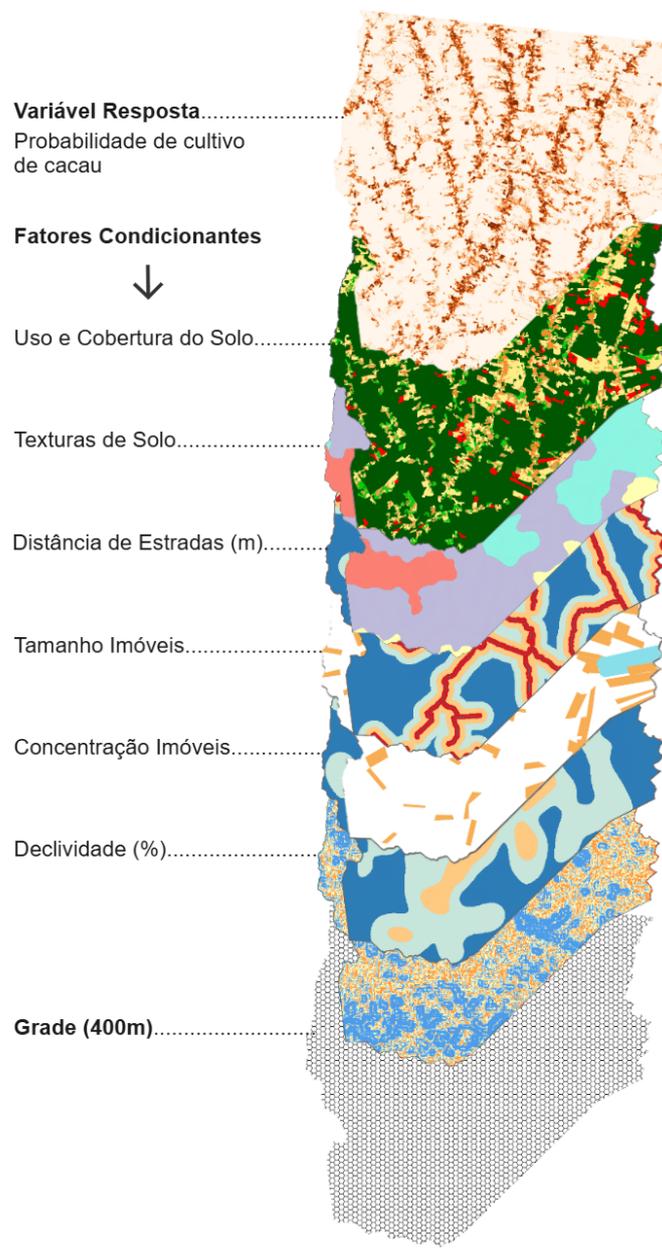


Figura 3: Esquema de agregação dos fatores condicionantes à variável resposta.

2.3. Verificação dos pressupostos e teste Lagrange Multiplier

Para avaliar qual o modelo de regressão espacial global mais adequado, aplicou-se o teste dos multiplicadores de Lagrange (*Lagrange Multiplier* - LM) [29, 30], que inclui as estatísticas LMerr (dependência no erro) e LMlag (dependência no lag). Caso ambos os testes fossem significativos, foram considerados os testes robustos RLMerr e RLMLag, que ajustam possíveis interdependências entre os efeitos espaciais. A seleção do modelo adequado foi baseada na estatística robusta mais significativa, priorizando o modelo de erro espacial (*Spatial Error* - CAR) ou de lag espacial (*Spatial Lag* - SAR). Essa abordagem assegurou a escolha mais adequada do modelo espacial global a ser empregado.

2.4. Regressão Global (*Spatial Error Model*)

O elemento da dependência espacial pode ser definido pela Fórmula 4, onde p é o parâmetro de correlação espacial, que mede a força da relação espacial; W é a matriz de pesos espaciais, que define como cada unidade está conectada às outras; e ξ é a componente do erro que não tem dependência espacial.

$$\varepsilon = pW\varepsilon + \xi \quad (4)$$

Na regressão espacial, o termo de erro é ajustado para incorporar a dependência espacial, como observado na Fórmula 5:

$$Y = X\beta + \varepsilon \quad (5)$$

A regressão global foi realizada utilizando o modelo *Spatial Error Model* (*Conditional Autoregressive Regression* - CAR), uma forma de modelo autoregressivo espacial. Este modelo assume que a dependência espacial está presente nos erros do modelo e, portanto, a correlação espacial entre as observações é considerada ao modelar a variável dependente. O CAR, como parte da análise, permite capturar as relações espaciais locais de uma maneira robusta, utilizando o comportamento espacial dos resíduos para melhorar a precisão da estimação.

Nesta etapa, as variáveis independentes não significativas foram removidas do modelo, resultando na exclusão da variável *declividade*. Dessa forma, a fórmula 6, ajustada para as variáveis significativas, ficou conforme apresentada abaixo:

$$V_{DCocoaVotes} = V_{ILULC} + V_{ITextSolo} + V_{IDistEstradas} + V_{ITamImoveis} + V_{IConc.Imoveis} \quad (6)$$

2.5. Regressão Local (*Geographically Weighted Regression*)

A *Geographically Weighted Regression* (GWR) foi utilizada como uma abordagem de regressão espacial local, permitindo a análise das variações espaciais nos coeficientes de regressão em diferentes áreas do Projeto de Assentamento Itatá. A GWR é uma técnica que estima os parâmetros de regressão de forma local (Fórmula 7, levando em consideração a dependência

espacial (ε) entre as observações (u_i, v_i) , o que permite capturar a heterogeneidade espacial e a não-estacionaridade dos dados.

$$Y_i = \beta_0(u_i, v_i) + X_{1i}\beta_1(u_i, v_i) + X_{2i}\beta_2(u_i, v_i) + \dots + X_{ki}\beta_k(u_i, v_i) + \varepsilon_i \quad (7)$$

Ao contrário dos modelos globais, que estimam um único conjunto de coeficientes para toda a região, a GWR fornece uma estimativa diferente para cada localização, oferecendo uma visão mais detalhada e precisa das relações entre as variáveis em diferentes partes do território estudado. Esse modelo possibilita a identificação de áreas com padrões distintos de cultivo de cacau, refletindo as influências espaciais locais de variáveis como proximidade de estradas, declividade e uso do solo.

2.6. Métricas avaliativas

Para avaliar o desempenho dos modelos foram utilizadas as métricas R^2 e o *Akaike Information Criterion* (AIC). O R^2 mede a proporção da variância da variável dependente que é explicada pelo modelo, indicando o quão bem as variáveis explicativas descrevem os dados observados. Valores mais próximos de 1 indicam um melhor ajuste do modelo [31]. Já o AIC é uma métrica que combina a qualidade do ajuste com a complexidade do modelo, penalizando modelos com um número excessivo de parâmetros [32]. Modelos com menor AIC são preferidos, pois indicam um equilíbrio entre simplicidade e precisão [33].

Também foi comparado em uma avaliação visual o mapa dos valores preditos pelo modelo global e pelo modelo local em relação ao mapa da variável dependente. Além disso foram gerados mapas dos coeficientes locais das variáveis independentes do modelo GWR e dos resultados do teste de significância T de Student [34] aplicado aos coeficientes. O teste T de Student é representado pela Fórmula 8, onde $\hat{\beta}_1$ é o valor estimado do coeficiente da variável independente, e $S(\hat{\beta}_1)$ representa o erro-padrão associado à estimativa do coeficiente. Esse teste é utilizado para verificar se o coeficiente $\hat{\beta}_1$ é significativamente diferente de zero, indicando se a variável correspondente exerce um efeito estatisticamente relevante no modelo.

$$t^* = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} \quad (8)$$

3. RESULTADOS

3.1. Teste de *Lagrange Multiplier*

Os resultados do teste de Lagrange Multiplier (Tabela 2) indicaram significância para dependência espacial nos resíduos do modelo, com valores elevados para as estatísticas LMerr (484,64; $p < 2,2e-16$) e LMlag (373,42; $p < 2,2e-16$). Os testes robustos confirmaram a preponderância do efeito de erro espacial, com RLMerr apresentando um valor expressivo (119,53; $p < 2,2e-16$) em relação ao RLMlag (8,30; $p = 0,0039$). Assim, o modelo de erro espacial (*Spatial Error* - CAR) foi considerado mais apropriado para capturar os padrões de dependência espacial no conjunto de dados analisados.

Teste	Estatística	p-valor
RSerr (LMerr)	484.64	<2.2e-16
RSlag (LMlag)	373.42	<2.2e-16
adjRSerr (RLMerr)	119.53	<2.2e-16
adjRSlag (RLMlag)	8.3038	0.003956

Tabela 2: Resultados dos testes de Lagrange Multiplier.

3.2. Regressão Global VS Local

Para avaliar a adequação dos modelos à análise espacial, foram analisados os valores de AIC e R^2 para os modelos GWR e CAR. O modelo GWR apresentou melhor ajuste global, com um AIC de -5549,32, indicando um menor erro de previsão penalizado em comparação ao modelo CAR, cujo AIC foi de -4697,01. Além disso, o R^2 do modelo GWR (62%) foi superior ao do modelo CAR (46%), reforçando sua melhor adequação no contexto analisado, considerando as variáveis e a estrutura espacial. A Figura 4 apresenta o gráfico de dispersão de ambos os modelos com seus respectivos valores de R^2 .

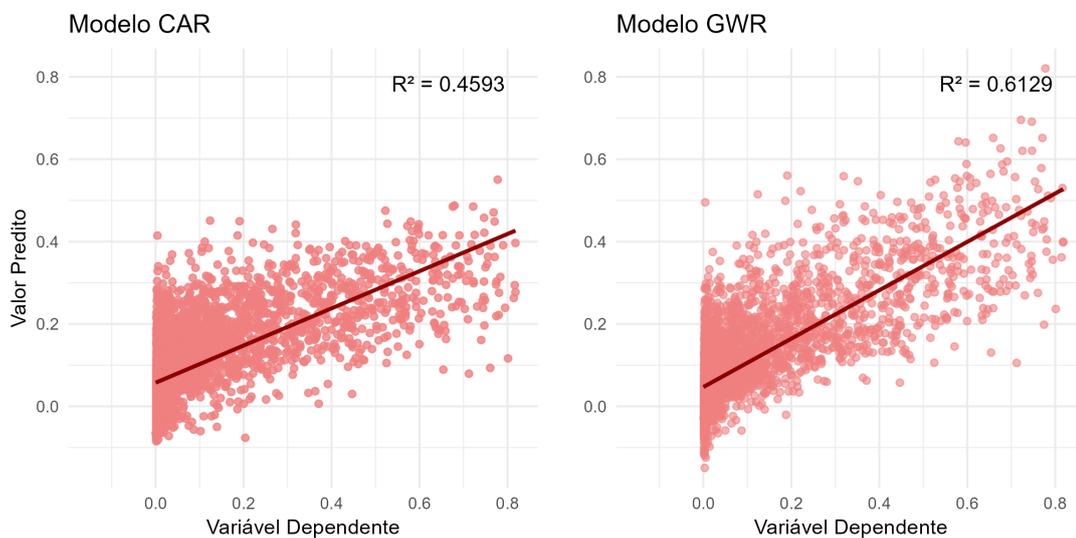


Figura 4: Gráficos de dispersão do coeficiente de determinação R^2 dos modelos CAR e GWR.

Ao comparar o mapa da variável resposta com os mapas gerados pelos modelos global e local (Figura 5), verificou-se que ambos representaram adequadamente os padrões de distribuição espacial da probabilidade de cultivo de cacau. No entanto, o modelo local GWR demonstrou maior precisão na representação da intensidade da probabilidade. Essa superioridade é evidenciada por sua maior semelhança visual entre o mapa predito e o mapa da variável resposta.

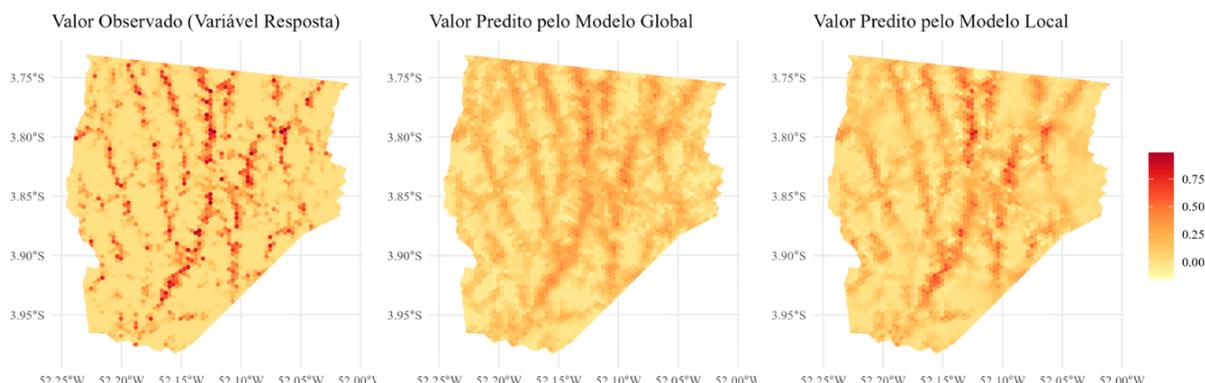


Figura 5: Mapas da variável resposta e dos valores preditos pelo modelo Global e Local.

A Tabela 3 apresenta os resultados dos modelos GWR e CAR para as principais variáveis independentes, com destaque para os valores de R^2 , estimativas dos parâmetros, erros padrão, valores Z e P -valores, agrupados por intervalos de R^2 . O foco da análise está na significância estatística das variáveis, representada pelos P -valores. O modelo global (CAR) identificou quatro das cinco variáveis como estatisticamente significativas (P -valor $< 0,05$), indicando sua relevância no ajuste geral do modelo. Por outro lado, o modelo local (GWR) destacou apenas duas variáveis como significativas: "Distância de Estradas" e "Uso e Cobertura do Solo".

Variável	R ² Agrupado GWR	Estimativa		Erro Padrão		Z-valor		P-valor	
		GWR	CAR	GWR	CAR	GWR	CAR	GWR	CAR
Concentração Imóveis	0-0.25	-0.0030		0.0570		-0.0526		9.5807e-01	
	0.25-0.5	0.0123		0.0498		0.2467		8.0517e-01	
	0.5-0.75	0.0072	0.0260	0.0450	0.0052	0.1602	5.0187	8.7273e-01	5.2000e-07
	0.75-1	-0.0011		0.0401		-0.0267		9.7866e-01	
Uso e Cobertura do Solo	0-0.25	0.0440		0.0323		1.3648		1.7231e-01	
	0.25-0.5	0.1098	0.1408	0.0356	0.0048	3.0874	29.4323	2.0189e-03	0.0000e+00
	0.5-0.75	0.1746		0.0347		5.0345		4.7900e-07	
	0.75-1	0.2501		0.0343		7.2959		2.9700e-13	
0-0.25	-0.0122		0.0484		-0.2519		8.0115e-01		
Textura do Solo	0.25-0.5	-0.0115	0.0088	0.0331	0.0040	-0.3476	2.2151	7.2816e-01	2.6751e-02
	0.5-0.75	-0.0012		0.0282		-0.0441		9.6485e-01	
	0.75-1	-0.0439		0.0276		-1.5889		1.1208e-01	
	0-0.25	0.0141		0.0274		0.5170		6.0513e-01	
Tamanho Imóveis	0.25-0.5	-0.0010	0.0060	0.0355	0.0048	-0.0278	1.2451	9.7786e-01	2.1308e-01
	0.5-0.75	0.0050		0.0342		0.1467		8.8340e-01	
	0.75-1	-0.0202		0.0356		-0.5665		5.7107e-01	
	0-0.25	0.0079		0.0225		0.3502		7.2618e-01	
Distância de Estradas	0.25-0.5	0.0372	0.0727	0.0290	0.0122	1.2833	5.9593	1.9940e-01	2.5300e-09
	0.5-0.75	0.0521		0.0312		1.6678		9.5352e-02	
	0.75-1	0.0811		0.0357		2.2741		2.2962e-02	
	0-0.25	0.0079		0.0225		0.3502		7.2618e-01	

Tabela 3: Resultados dos modelos GWR e CAR para as principais variáveis, incluindo valores de R^2 , estimativas dos parâmetros, erros padrão, valores Z e P -valores, agrupados por intervalos de R^2 .

Essa diferença na identificação das variáveis relevantes pode ser explicada pela abordagem distinta de cada modelo. O modelo CAR, de natureza global, considera a influência de todas as variáveis de forma agregada, o que pode amplificar sua significância ao não levar em conta as variações espaciais locais. Já o modelo GWR, por ser um modelo local, analisa os padrões

espaciais de forma detalhada, permitindo que a significância das variáveis varie espacialmente. Nesse contexto, as variáveis "Distância de Estradas" e "Uso e Cobertura do Solo" emergem como mais importantes, sugerindo que elas têm uma influência espacialmente consistente sobre o fenômeno estudado.

3.3. Regressão Local

A Figura 6 apresenta o mapa do coeficiente de determinação (R^2) do modelo GWR, ilustrando a distribuição espacial da capacidade explicativa do modelo em diferentes regiões. O R^2 , agrupado em intervalos que variam de 0,25 a 1, indica o ajuste local do modelo em diferentes áreas do território. As regiões com valores mais altos de R^2 (em tons de vermelho) demonstram um melhor ajuste do modelo, sugerindo que as variáveis independentes explicam de forma mais consistente a variável resposta nessas localidades. Por outro lado, as áreas com valores mais baixos (em tons claros) indicam uma menor capacidade do modelo de capturar a variabilidade local, possivelmente devido à heterogeneidade espacial ou à ausência de fatores explicativos relevantes para essas regiões. Observa-se ainda que o ajuste é melhor nas áreas mais afastadas das regiões com maior probabilidade de cultivo de cacau, sugerindo que nessas áreas a relação entre as variáveis independentes e a variável dependente é mais estável e menos sujeita a influências de fatores locais complexos. Essa visualização reforça a utilidade do modelo GWR em capturar variações locais de desempenho, contrastando com modelos globais que não consideram essa variação.

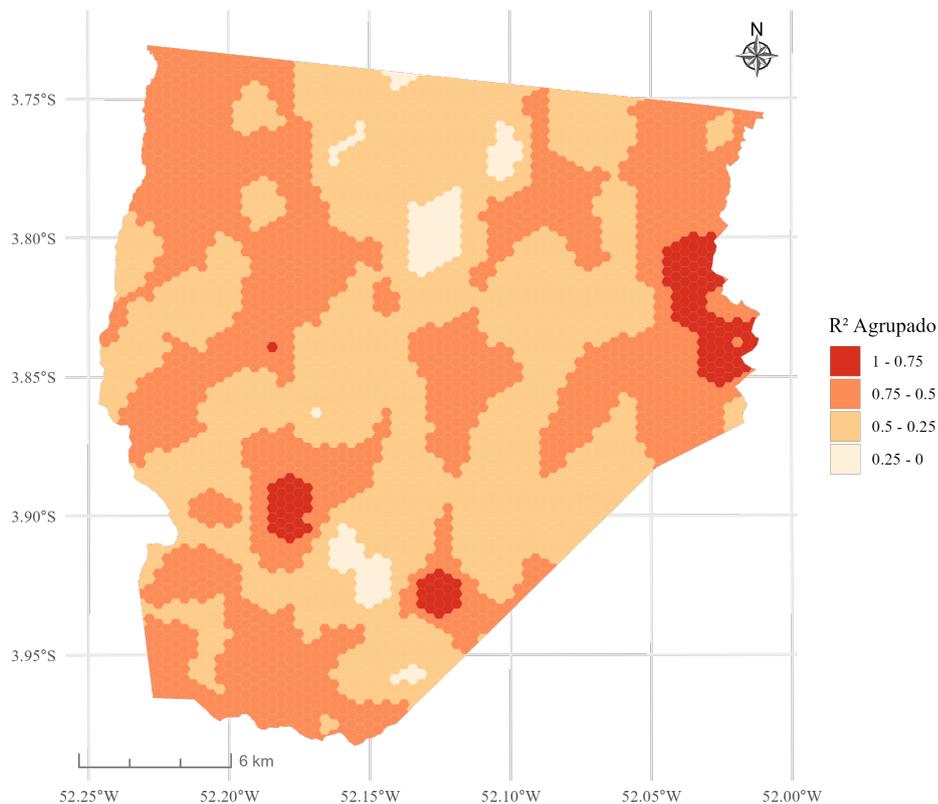


Figura 6: Mapa do coeficiente de determinação (R^2) do modelo GWR, com agrupamento por faixas de desempenho local.

Conforme verificado na Tabela 3, as variáveis independentes "Tamanho de Imóveis", "Textura do Solo" e "Concentração de Imóveis" não apresentaram valores p significativos para o modelo local GWR. A Figura 7 apresenta, na parte superior, a distribuição espacial dos coeficientes locais dessas variáveis, divididos em quartis, enquanto a parte inferior exibe os resultados dos testes T de Student. As áreas em destaque no mapa inferior indicam as regiões onde os coeficientes são estatisticamente significativos, enquanto as demais áreas, representadas em tons mais claros, correspondem a valores não significativos. Observa-se que a maior parte das regiões não apresenta significância estatística, corroborando os resultados da tabela e evidenciando que essas variáveis têm um impacto limitado no modelo em nível local. Além disso, as áreas consideradas significativas estão distribuídas de forma dispersa, sem a formação de grandes zonas de continuidade espacial, o que reforça a baixa influência dessas variáveis no padrão observado.

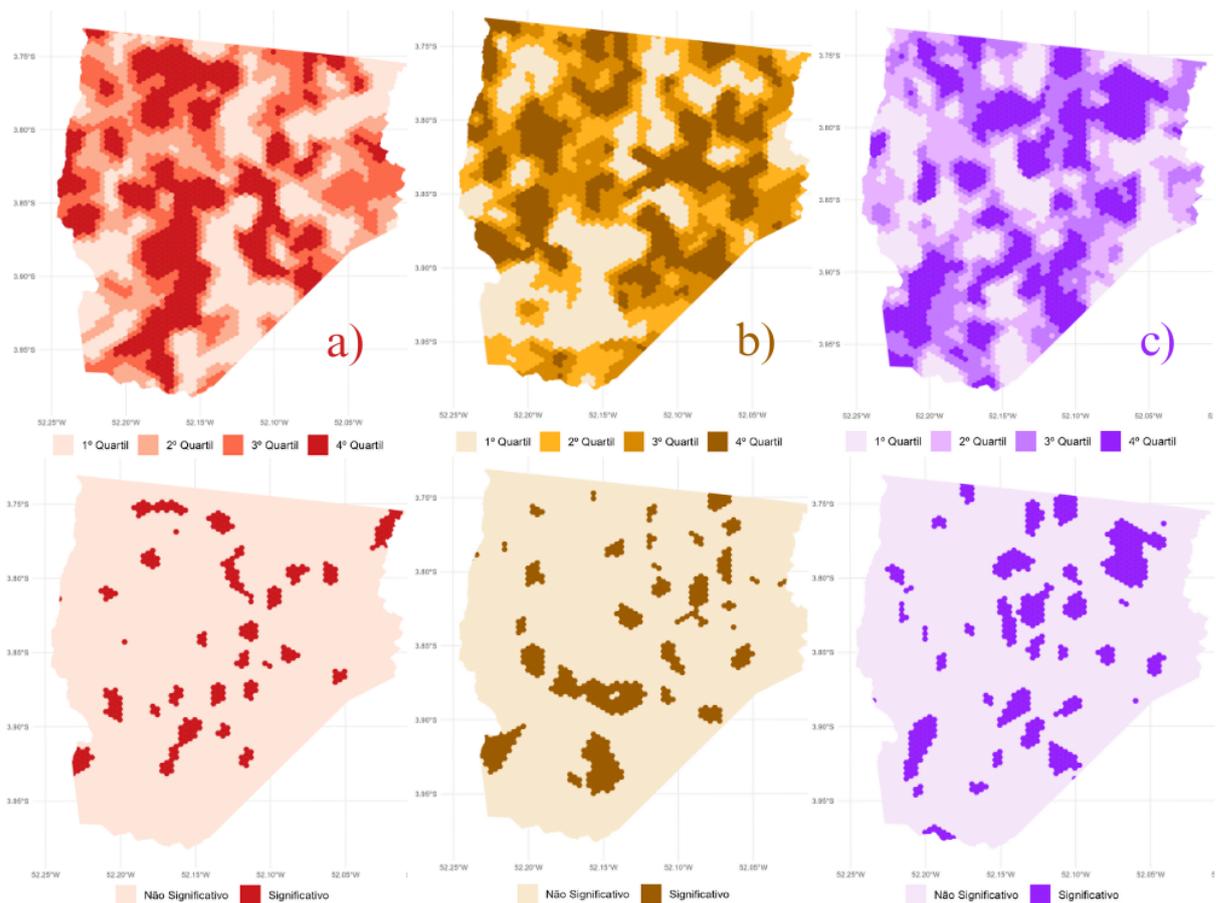


Figura 7: Coeficientes locais (acima) e testes T de Student (abaixo) das variáveis independentes não significativas: a) Tamanho de imóveis; b) Textura do Solo; c) Concentração de Imóveis.

As variáveis que apresentaram valores de p significativos no modelo GWR incluem 'Distância de Estradas' e 'Uso e Cobertura do Solo'. A Figura 8 ilustra, à esquerda, a distribuição espacial dos coeficientes locais da variável 'Distância de Estradas' e, à direita, os resultados do teste de significância T de Student. Observa-se que, diferentemente das variáveis independentes que não apresentaram significância no modelo local, a variável 'Distância de Estradas' exibe uma maior quantidade de zonas com significância estatística, que, além de serem mais numerosas, apresentam maior contiguidade espacial.

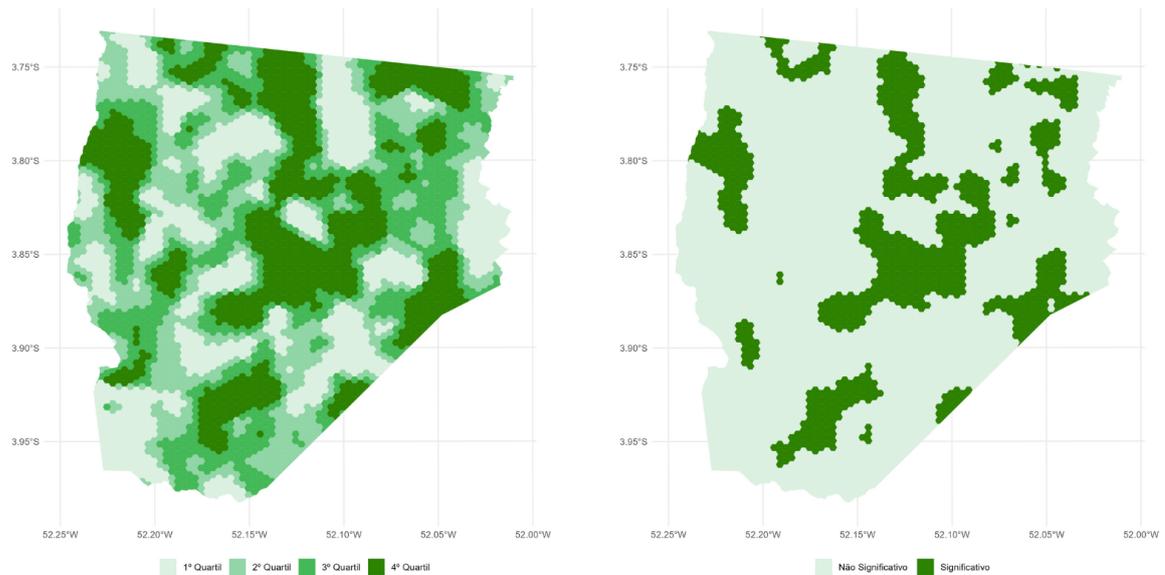


Figura 8: Distribuição espacial dos coeficientes locais (à esquerda) e respectivos resultados do teste de significância T de Student (à direita) para a variável Distância de Estradas no modelo GWR.

A Figura 9 apresenta a distribuição espacial dos coeficientes locais (à esquerda) e os respectivos resultados do teste T de Student (à direita) para a variável 'Uso e Cobertura do Solo' no modelo GWR. Nota-se que as zonas significativas estão amplamente distribuídas no espaço, com uma predominância em áreas centrais e no sul da região analisada. Em comparação com a variável 'Distância de Estradas', as zonas significativas da variável 'Uso e Cobertura do Solo' exibem um padrão mais fragmentado, com áreas não significativas dispersas no mapa.

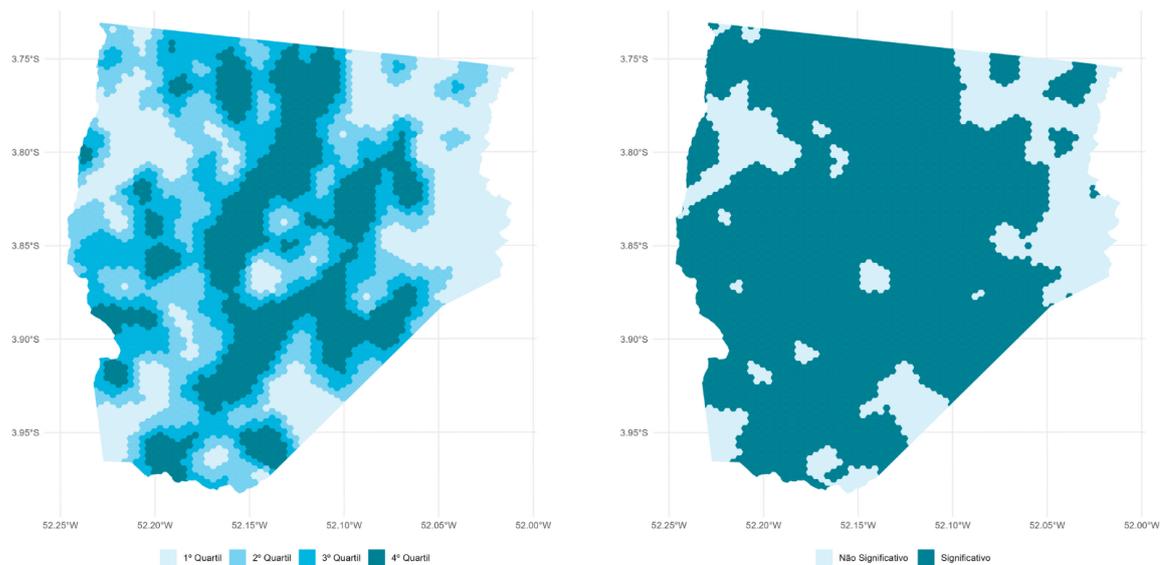


Figura 9: Distribuição espacial dos coeficientes locais (à esquerda) e respectivos resultados do teste de significância T de Student (à direita) para a variável Uso e Cobertura do Solo no modelo GWR.

4. CONCLUSÃO

Este estudo analisou a variação espacial da importância das variáveis geográficas relacionadas ao cultivo de cacau no Projeto de Assentamento Itatá, utilizando modelos de regressão espacial para comparar a adequação dos modelos global (CAR) e local (GWR). Os resultados destacaram a superioridade do modelo GWR em capturar padrões espaciais heterogêneos, com um melhor ajuste global, evidenciado pelos menores valores de AIC e maior coeficiente de determinação (R^2) em relação ao modelo CAR. Essa diferença pode ser atribuída à capacidade do modelo GWR de adaptar os coeficientes às características locais, minimizando o impacto de outliers e considerando as nuances espaciais na área de estudo.

A análise das variáveis independentes evidenciou que, enquanto o modelo global identificou quatro das cinco variáveis como estatisticamente significativas, o modelo local destacou apenas duas: Distância de Estradas e Uso e Cobertura do Solo. Essas variáveis emergiram como as mais relevantes no contexto local, apresentando uma influência mais consistente e espacialmente significativa sobre a probabilidade de cultivo de cacau. Essa divergência reflete a natureza dos modelos analisados: o CAR assume uma relação uniforme para todas as áreas, o que pode amplificar a significância de variáveis menos influentes, enquanto o GWR permite que as relações variem espacialmente, revelando padrões locais mais detalhados.

Os mapas gerados pelo GWR também demonstraram que o ajuste do modelo é melhor em áreas mais afastadas das regiões com maior probabilidade de cultivo de cacau. Isso sugere que, nessas áreas, as relações entre as variáveis independentes e a variável dependente são mais estáveis e menos afetadas por fatores locais complexos. Por outro lado, as zonas de menor ajuste podem indicar a ausência de variáveis explicativas relevantes ou a influência de fatores locais não considerados no modelo.

Com base nos resultados, este trabalho reforça a importância de modelos locais, como o GWR, para a análise de fenômenos espaciais que apresentam heterogeneidade em suas relações. A utilização de variáveis geográficas como Uso e Cobertura do Solo e Distância de Estradas mostrou-se promissora para o mapeamento de cacau em áreas tropicais, embora a inclusão de outras variáveis ou o uso de abordagens multiescala possam aprimorar ainda mais os resultados.

Por fim, o estudo demonstra que a análise de regressão local é uma ferramenta útil para a compreensão de dinâmicas espaciais complexas, especialmente em áreas de maior heterogeneidade, como o Projeto de Assentamento Itatá. Trabalhos futuros podem explorar a integração de dados adicionais, como variáveis climáticas e temporais, bem como a aplicação de modelos híbridos que combinem características de abordagens locais e globais para capturar diferentes dimensões da relação entre o cultivo de cacau e o ambiente.

Por fim, o estudo demonstra que a análise de regressão local é uma ferramenta útil para a compreensão de dinâmicas espaciais complexas, especialmente em áreas de maior heterogeneidade, como o Projeto de Assentamento Itatá. Como sugestão para trabalhos futuros, recomenda-se a aplicação do método Geodetector, que se destaca por não ser sensível à colinearidade entre as variáveis explicativas, permitindo avaliar e explorar a influência de fatores condicionantes sobre o cultivo de cacau. Além disso, sugere-se a aplicação do método de simulação Monte Carlo para avaliar a robustez dos resultados obtidos. Esse método realiza múltiplas iterações dos dados, gerando distribuições de probabilidade para os coeficientes e permitindo a identificação de padrões consistentes ao longo das simulações. A inclusão

dessa abordagem contribui para a obtenção de estimativas mais confiáveis e estatisticamente robustas, reduzindo o impacto de incertezas e fornecendo uma compreensão mais abrangente e aprimorada das dinâmicas observadas no Projeto de Assentamento Itatá.

5. REFERÊNCIAS

- [1] G. Bondar. *O cultivo do cacau*. Tipografia Naval, Salvador, 1956.
- [2] Amílcar Baiardi. *Subordinação do trabalho ao capital na lavoura cacauzeira da Bahia*. Editora Hucitec, São Paulo, 1984.
- [3] M. Rosário, T. Perrucho, R. L. Fowler, and J. C. Sales. *Cacau: história e evolução no Brasil e no mundo*. CEPLAC, Ilhéus, 1978.
- [4] F. B. Piasentin and C. H. Saito. Os diferentes métodos de cultivo de cacau no sudeste da bahia, brasil: Aspectos históricos e percepções. *Boletim do Museu Paraense Emílio Goeldi: Ciências Humanas*, 9:61–78, 2014.
- [5] A. Venturieri et al. The sustainable expansion of the cocoa crop in the state of pará and its contribution to altered areas recovery and fire reduction. *Journal of Geographic Information System*, 14(3):361–386, 2022. Disponível online. Acesso em: 20 jul. 2024.
- [6] Ministério da Agricultura, Pecuária e Abastecimento. Plano agrícola e pecuário 2023/2024, 2023. Disponível online. Acesso em: 28 jul. 2024.
- [7] E. C. Landau, G. A. da Silva, and L. Moura. Evolução da produção de cacau (*theobroma cacao*, malvaceae). *Revista Brasileira de Agroecologia*, 16(1):1–14, 2020.
- [8] World Cocoa Foundation. Cocoaaction brasil. Disponível online. Acesso em: 03 nov. 2024.
- [9] Instituto Nacional de Pesquisas Espaciais (INPE). TerraClass - mapeamento de uso e cobertura do solo. Disponível online. Acesso em: 12 jul. 2024.
- [10] R. Folhes and A. Serra. Os efeitos da concorrência de trajetórias tecnológicas na economia cacauzeira paraense sobre as promessas de sustentabilidade do setor: um estudo a partir da transamazônica, Pará, Brasil. *Pappers do NAEA*, 1, 2023.
- [11] J. E. Batista et al. Optical time series for the separation of land cover types with similar spectral signatures: Cocoa agroforest and forest. *International Journal of Remote Sensing*, 43(9):3298–3319, 2022.
- [12] I. O. Abu et al. Detecting cocoa plantations in cote d’ivoire and ghana and their implications on protected areas. *Ecological Indicators*, 129:107863, 2021.
- [13] Serviço Nacional de Aprendizagem Rural (SENAR). *Cacau: produção, manejo e colheita*. Coordenação de Produção e Distribuição de Materiais Instrucionais: B. H. B. Araújo, Brasília, 2018.
- [14] G. G. Correia. Dados multisensor e inferência geográfica para a identificação e análise da dinâmica do cacau na transamazônica paraense. Dissertação de mestrado, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, 2024. 79p.

- [15] Osei-Gyabaah et al. Land suitability analysis for cocoa (theobroma cacao) production in the sunyani municipality, bono region, ghana. *Smart Agricultural Technology*, 5:100262, 2023. Disponível online. Acesso em: 02 jul. 2024.
- [16] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46(Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods):234–240, 1970. Disponível online. Acesso em: 06 jul. 2024.
- [17] Roger S. Bivand, Edzer Pebesma, and Virgilio Gómez-Rubio. *Applied Spatial Data Analysis with R*. Springer, New York, second edition edition, 2013.
- [18] A. Stewart Fotheringham, Chris Brunson, and Martin Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons Ltd, Chichester, England, 2002. Disponível online. Acesso em: 06 jul. 2024.
- [19] Instituto Brasileiro de Geografia e Estatística (IBGE). Censo demográfico 2022, 2022. Disponível online. Acesso em: 9 out. 2024.
- [20] Pecuária e Abastecimento Ministério da Agricultura. Plano agrícola e pecuário 2023/2024. Disponível online. Acesso em: 28 jul. 2024.
- [21] Instituto Nacional de Colonização e Reforma Agrária (INCRA). Acervo fundiário, 2024. Disponível online. Acesso em: 9 out. 2024.
- [22] C. Totti et al. Uso de atributos texturais e de contexto com random forest na classificação de áreas de cultivo de cacau no Projeto de Assentamento Itatá (PA), 2024. (submetido à publicação).
- [23] C. Totti et al. Identificação de áreas potenciais de ocorrência de cultivo de cacau no Projeto de Assentamento Itatá, no Pará, 2024. (submetido à publicação).
- [24] Instituto Nacional de Pesquisas Espaciais (INPE). Topodata - modelo digital de elevação. Disponível online. Acesso em: 29 jul. 2024.
- [25] Instituto Brasileiro de Geografia e Estatística (IBGE). Dados de textura do solo: Escala 1:250.000, 2023. Disponível online. Acesso em: 30 jul. 2024.
- [26] SICAR. Dados do Cadastro Ambiental Rural (CAR): Tamanho do Imóvel em Formato Vetorial, 2024. Disponível online. Acesso em: 30 jul. 2024.
- [27] Instituto Brasileiro de Geografia e Estatística (IBGE). Cadastro Nacional De Endereços Para Fins Estatísticos (CNEFE), 2022. Disponível online. Acesso em: 31 jul. 2024.
- [28] OpenStreetMap contributors. Openstreetmap data, 2024. Disponível online. Acesso em: 30 jul. 2024.
- [29] Luc Anselin. *Spatial econometrics: methods and models*. Kluwer Academic Publishers, Dordrecht, 1988.
- [30] Luc Anselin. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27:247–267, 2002.
- [31] Andy Field. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications, 2013.

- [32] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [33] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, Taylor Francis Group, Boca Raton, FL, 2016. Chapman & Hall/CRC Interdisciplinary Statistics Series.
- [34] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908. Disponível online. Acesso em: 14 dez. 2024.