

Identificação de fatores determinantes para o cultivo de cacau por meio de regressão espacial no Projeto de Assentamento Itatá, Pará

Camila Totti¹, Vinícius L. Guimarães¹, Antonio Miguel V. Monteiro¹,
Maria Isabel S. Escada¹, Sidnei João S. Sant'Anna¹

¹Instituto Nacional de Pesquisas Espaciais (INPE)
São José dos Campos – SP – Brasil

{camila.andrade, vinicius.lima, miguel.monteiro,
isabel.escada, sidnei.santanna}@inpe.br

Abstract. *Cocoa cultivation, which is important to the economy of Pará, is expanding in the southwest of the state, where the Itatá Settlement Project is located. Mapping this crop is essential to understand its distribution and the spatial factors associated with its occurrence. This study evaluated the influence of variables indicating cocoa cultivation through spatial regression, comparing a Conditional Autoregressive Model and a Geographically Weighted Regression model. Five independent variables derived from a previous study using the AHP technique were analyzed. The results indicate that the GWR presented a greater capacity to capture spatial patterns, with emphasis on the variables ‘influence of land use and occupation’ and ‘distance to roads’ in the distribution of the crop.*

Resumo. *O cultivo de cacau, relevante para a economia do Pará, se expande no sudoeste do estado, onde se localiza o Projeto de Assentamento Itatá. O mapeamento dessa cultura é fundamental para compreender sua distribuição e os fatores espaciais associados à sua ocorrência. Este estudo avaliou a influência de variáveis indicadoras do cultivo de cacau por meio de regressão espacial, comparando um modelo Conditional Autoregressive Model e um modelo Geographically Weighted Regression. Foram analisadas cinco variáveis independentes derivadas de um estudo prévio com a técnica AHP. Os resultados indicam que o GWR apresentou maior capacidade de captar padrões espaciais, destacando-se as variáveis ‘influência do uso e ocupação da terra’ e ‘distância às estradas’ na distribuição do cultivo.*

1. Introdução

O cultivo do cacau (*Theobroma cacao*) no Brasil iniciou-se no Pará, impulsionado por condições climáticas e edáficas favoráveis, especialmente no sudoeste do estado. Desde o século XIX, expandiu-se substituindo pastagens e florestas, com forte participação de migrantes na mão de obra e na ocupação de lotes da reforma agrária [Brondízio and Moran 2012]. Atualmente, o Pará lidera a produção nacional de cacau, respondendo por mais de 50% do total [MAPA 2023]. Municípios próximos à rodovia

Transamazônica e áreas de expansão, como o Projeto de Assentamento (PA) Itatá, passaram por intensas mudanças no uso da terra, com a conversão de florestas para cultivos de cacau, como observado em imagens de satélite [Venturieri et al. 2022] (comunicação pessoal, LiSS¹/INPE, 2024).

Apesar da relevância econômica e ambiental do cacau, seu mapeamento é desafiador devido à semelhança espectral com vegetação secundária e à presença de diferentes sistemas produtivos². Além disso, a cobertura frequente de nuvens na Amazônia limita o uso de imagens ópticas [Venturieri et al. 2022]. Para superar essas barreiras, modelos de regressão espacial vêm incorporando índices espectrais e variáveis indicadoras [Bivand et al. 2013, Fotheringham et al. 2002]. O cultivo de cacau apresenta dependência espacial, com áreas próximas sendo mais similares entre si [Fotheringham et al. 2002]. Assim, variáveis como distância a estradas, classes de uso e cobertura da terra, tamanho de imóveis e textura do solo são consideradas na predição da presença do cultivo [Totti et al. 2024a, Correia 2024, Folhes and Serra 2023]. No sentido de aprimorar metodologias de mapeamento de cacau em áreas tropicais e compreender as dinâmicas espaciais que influenciam a distribuição desta cultura, este estudo analisa a variação espacial da importância de variáveis indicadoras no Projeto de Assentamento Itatá, comparando o Modelo Autoregressivo Condicional (CAR) e a Regressão Geograficamente Ponderada (*Geographically Weighted Regression* - GWR).

2. Material e Métodos

O Projeto de Assentamento (PA) Itatá localiza-se no município de Senador José Porfírio, no sudoeste do estado do Pará, no bioma Amazônia. Sua área é de 52.523,31 hectares [INCRA 2024] e sua extensão de 14.420 km² [IBGE 2022]. O mapa de localização do PA Itatá pode ser visualizado na Figura 1.

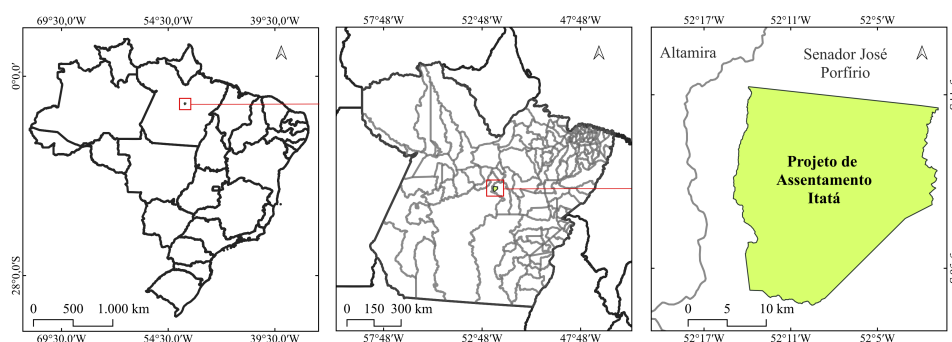


Figura 1. Mapa de localização do Projeto de Assentamento Itatá - Pará.

Para realização deste estudo, coletou-se dados de probabilidade de cultivo de cacau (variável dependente) e dados ambientais e de infraestrutura (variável independente). Um dado matricial de probabilidade da classe Cacau foi utilizado como a variável dependente (VD), obtido previamente por meio de uma classificação com *Random Forest* [Totti et al. 2024b] com quatro classes: Cacau Pleno Sol, Floresta Secundária, Floresta

¹Laboratório de Investigação em Sistemas Socioambientais

²O cacau é cultivado em monocultivo a pleno sol ou em Sistemas Agroflorestais (SAFs), consorciado com outras espécies [Folhes and Serra 2023].

Primária e Pastagem. Esse mapeamento foi realizado com imagens Sentinel-2A, quatro índices de vegetação e 18 medidas texturais, alcançando 93,2% de acurácia global.

As variáveis independentes (VIs) são derivadas de um estudo prévio com o Processo de Análise Hierárquica (AHP) [Totti et al. 2024a]. São estas: uso e cobertura do solo, textura do solo, tamanho dos imóveis, concentração de estabelecimentos rurais e distância às estradas, aplicadas para identificar áreas potenciais de cultivo de cacau no PA Itatá. As cinco variáveis independentes (Figura 2) foram utilizadas considerando suas conexões ao cultivo de cacau. Estradas auxiliam no escoamento e manejo do cacau, justificando a utilização do dado de proximidade de estradas. O cultivo de cacau é realizado em estabelecimentos com menos de 100 ha, o que justifica a utilização do dado de tamanho dos estabelecimentos rurais. Além disso, os tipos de solos mais adequados ao cultivo de cacau são aqueles com textura médio-argilosa, o que contribui para o equilíbrio drenagem/retenção em áreas de alta pluviosidade, como na região do PA Itatá [Totti et al. 2024a, Correia 2024, Folhes and Serra 2023].

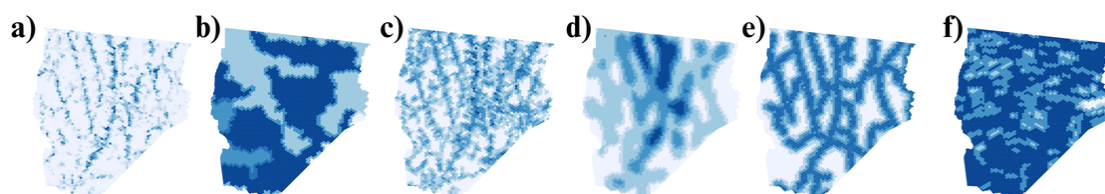


Figura 2. VD: A) probabilidade de cacau; VIs: B) textura do solo; C) uso e cobertura da terra; D) concentração de estabelecimentos rurais; E) distância à estradas; F) tamanho dos imóveis.

Para o alcance dos objetivos propostos neste estudo, foi organizada a seguinte metodologia, dividida em cinco etapas principais (Figura 3): *I*. Obtenção dos dados de entrada (mencionada acima); *II*. Verificação da dependência espacial; *III*. Definição da matriz de vizinhança; *IV*. Geração de um modelo CAR e um modelo GWR; e *V*. Avaliação comparativa dos modelos e da variação da importância das variáveis indicadoras no PA Itatá.



Figura 3. Ilustração do processo metodológico.

Na etapa *II*, foi verificada a dependência espacial da variável dependente, considerando a influência da distribuição espacial do cultivo de cacau sobre os valores observados. Para isso, foram empregados testes estatísticos de autocorrelação espacial, como o Índice de Moran, para identificar padrões espaciais na distribuição do cultivo. Na etapa *III*, a matriz de vizinhança foi definida pelo Índice de Moran aplicado à Regressão por Mínimos Quadrados Ordinários (OLS), para detectar padrões espaciais nos resíduos. Um script em R foi utilizado para identificar a maior grade hexagonal que mantivesse a autocorrelação espacial, resultando na seleção de células com arestas de 400 metros. Essa

grade foi utilizada como unidade espacial para a extração de estatísticas zonais da VD e das VIs, para a execução dos modelos de regressão.

Na etapa IV, foram ajustados dois modelos estatísticos: um modelo CAR e um modelo GWR. O modelo CAR incorporou a estrutura de dependência espacial identificada na etapa anterior, enquanto o modelo GWR permitiu a análise da variação espacial dos coeficientes das variáveis independentes. Em cada modelo final, foram mantidas apenas as variáveis identificadas como significativas dentro de sua respectiva abordagem. Por fim, na etapa V, a avaliação dos modelos foi conduzida por meio do coeficiente de determinação (R^2) e do *Akaike Information Criterion* (AIC)³, além de comparações visuais entre mapas preditos e observados. No modelo GWR, foram analisados os coeficientes locais das variáveis independentes e a significância estatística dessas variáveis foi verificada pelo teste t de Student⁴.

3. Resultados

O teste de Moran I aplicado aos resíduos da regressão indicou dependência espacial significativa ($I = 0,207$; $p < 2,2 \times 10^{-16}$), sugerindo que os resíduos não são distribuídos aleatoriamente. Em seguida, foi gerada uma matriz de vizinhança (células de 400m), extraíndo-se a moda da VD e a mediana das VIs para a regressão. Os modelos CAR e GWR foram ajustados, considerando apenas VIs significativas ($p < 0,05$), de modo que no CAR, todas as VIs foram mantidas, exceto “tamanho dos imóveis”. No GWR, todas foram retidas, pois ao menos 5% das células de cada VI apresentaram significância. A VI com maior área significativa foi “uso e cobertura da terra” (76,7%), seguida de “distância às estradas” (22,4%), “concentração de imóveis” (13,2%), “textura do solo” (11,4%) e “tamanho de imóveis” (7,7%). A Figura 4 apresenta a distribuição espacial dos coeficientes locais e dos testes t de Student.

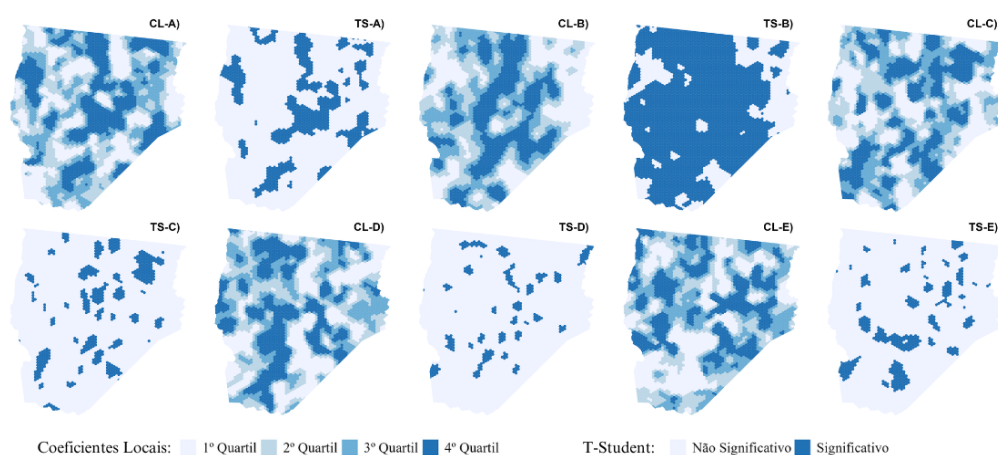


Figura 4. Coeficientes locais (CL) e dos testes t de Student (TS) das VIs: A) distância às estradas, B) uso e cobertura da terra, C) concentração dos estabelecimentos rurais, D) tamanho dos imóveis, E) textura do solo.

³O *Akaike Information Criterion* (AIC) avalia a qualidade do ajuste de um modelo, penalizando a complexidade. Valores menores indicam melhor equilíbrio entre ajuste e parcimônia.

⁴O teste t de Student avalia se há diferença significativa entre médias de amostras, assumindo normalidade dos dados.

A variável “uso e cobertura da terra” exibe os coeficientes mais elevados e a maior área de significância, destacando sua influência na modelagem espacial. “Distância às estradas” apresenta um padrão mais fragmentado, concentrado em regiões específicas. As demais variáveis exibem padrões distintos, com áreas significativas dispersas, indicando variações locais em sua influência. A maior significância de “uso e cobertura da terra” pode estar relacionada a sua característica de identificar diretamente a classe predominante em cada pixel, diferenciando áreas de cultura perene das demais categorias de uso e cobertura da terra. Em contraste, as demais variáveis refletem fatores que influenciam a presença do cultivo, mas não o caracterizam diretamente, o que pode explicar sua menor significância estatística.

O modelo GWR permite observar a distribuição espacial do coeficiente de determinação e avaliar a qualidade do ajuste em diferentes regiões. Observa-se na Figura 5 uma variação espacial significativa, com áreas de maior ajuste (tons avermelhados) concentradas principalmente nas regiões centrais a norte e sul da área de estudo, enquanto áreas de menor ajuste (tons azulados) ocorrem em outros setores.

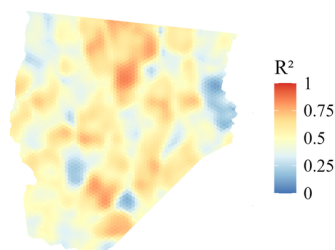


Figura 5. Mapa do coeficiente de determinação (R^2) do modelo GWR

Quando comparados, o modelo GWR apresenta maior complexidade devido à sua abordagem local, que permite a variação espacial dos coeficientes, enquanto o modelo CAR impõe uma estrutura de dependência espacial fixa entre as observações. A Figura 6 ilustra o ajuste dos modelos CAR (B) e GWR (C) em relação à variável dependente (A). O GWR obteve um coeficiente de determinação (R^2) de 0,61, superior aos 0,46 do CAR, além de um menor AIC (-5.549,32 contra -4.697,01), indicando melhor adequação ao conjunto de dados.

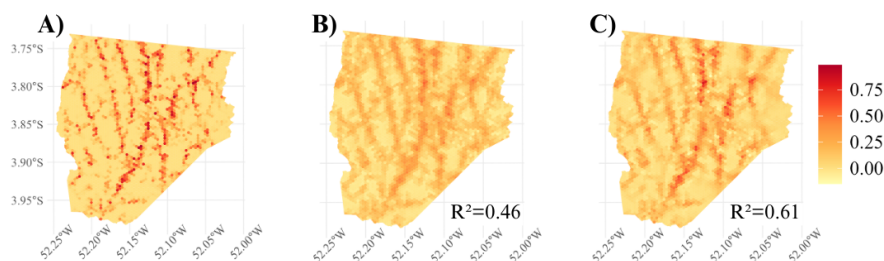


Figura 6. Mapas da variável dependente (A) e dos valores preditos pelo modelo global CAR (B) e local GWR (C).

4. Conclusão

Neste estudo, foi avaliada a influência de variáveis geográficas no cultivo de cacau no Projeto de Assentamento Itatá, comparando abordagens de regressão espacial global (CAR) e

local (GWR). Os resultados indicam que o GWR apresentou melhor desempenho ao capturar a heterogeneidade espacial da relação entre as variáveis independentes e a ocorrência do cultivo, com melhor ajuste global (menor AIC e maior R^2). O uso e cobertura da terra e a proximidade a estradas foram identificados como os fatores mais influentes, embora sua relação com a presença do cultivo varie espacialmente. Os resultados apresentados ressaltam a importância de metodologias que incorporem a variabilidade espacial para a compreensão dos padrões de uso da terra e o aprimoramento das técnicas de mapeamento. Além disso, reforçam a necessidade de estratégias adaptativas para o ordenamento territorial em regiões amazônicas, onde a diversidade ambiental e fundiária influencia a dinâmica agrícola. Como perspectivas futuras, recomenda-se a incorporação de métodos como Geodetector para avaliar interações entre variáveis e simulações Monte Carlo para quantificar incertezas locais. Espera-se que esses avanços contribuam para o desenvolvimento de abordagens mais robustas na análise espacial do cultivo de cacau e de outras atividades agrícolas em áreas tropicais.

Referências

- Bivand, R. S., Pebesma, E., e Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. Springer, New York, second edition edition.
- Brondízio, E. e Moran, E. (2012). Level-dependent deforestation trajectories in the brazilian amazon from 1970 to 2001. *Population and Environment*, 34:69–85.
- Correia, G. G. (2024). Dados multisensor e inferência geográfica para a identificação e análise da dinâmica do cacau na transamazônica paraense. Proposta de dissertação de mestrado, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP. 79p.
- Folhes, R. e Serra, A. (2023). Os efeitos da concorrência de trajetórias tecnológicas na economia cacaueira paraense sobre as promessas de sustentabilidade do setor: um estudo a partir da transamazônica, Pará, Brasil. *Pappers do NAEA*, 1.
- Fotheringham, A. S., Brunson, C., e Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons Ltd, Chichester, England.
- IBGE (2022). Censo demográfico 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/22827-censo-demografico-2022.html?t=sobre>. Acesso em: 9 out. 2024.
- INCRA (2024). Acervo fundiário. Disponível em: <https://acervofundiario.incra.gov.br/acervo/acv.php>. Acesso em: 9 out. 2024.
- MAPA (2023). Plano agrícola e pecuário 2023/2024. Disponível em: <https://www.gov.br/agricultura/pt-br>. Acesso em: 28 jul. 2024.
- Totti, C. et al. (2024a). Identificação de áreas potenciais de ocorrência de cultivo de cacau no Projeto de Assentamento Itatá, no Pará. (submetido à publicação).
- Totti, C. et al. (2024b). Uso de atributos texturais e de contexto com random forest na classificação de áreas de cultivo de cacau no Projeto de Assentamento Itatá (PA). (submetido à publicação).
- Venturieri, A. et al. (2022). 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. Disponível em: <https://www.scirp.org/journal/paperinformation?paperid=118147>. Acesso em: 20 jul. 2024.