




Padrões territoriais da diversificação produtiva municipal em Mato Grosso do Sul - evidências a partir de Random Forest Regression. Territorial patterns of municipal agricultural diversification in Mato Grosso do Sul - evidence from Random Forest Regression.

[Rafael Rossi](#)¹

¹- Docente e Pesquisador no Programa de Pós-Graduação em Recursos Naturais, Universidade Federal de Mato Grosso do Sul – UFMS , *Campus* Campo Grande/MS – Brasil. E-mail: r.rossi@ufms.br

Resumo

O presente artigo é instrumento por meio do qual investigamos os padrões territoriais da diversificação produtiva municipal em Mato Grosso do Sul a partir da análise de seis culturas alimentares tradicionais (banana, batata-doce, feijão, mandioca, melancia e milho) com dados da Pesquisa Agrícola Municipal (IBGE, 2024). A diversidade foi mensurada por meio do Índice de Shannon, enquanto os regimes produtivos dominantes foram identificados a partir da maior contribuição padronizada (z-score) entre as culturas. Aplicou-se análise de Contingência Bayesiana para avaliar a associação entre regime dominante e níveis de diversificação, além de Random Forest Regression (RFR) para modelar a estrutura multivariada da diversidade produtiva. No âmbito espacial, utilizou-se o modelo BYM2 para estimar o efeito espacial latente do Shannon e o Moran's I Global para avaliar autocorrelação dos valores preditos. Conclui-se que a diversificação produtiva municipal se conforma em fenômeno estruturalmente organizado, cuja compreensão se beneficia da integração entre modelagem estatística multivariada e territorialização cartográfica.

Palavras-chave: Diversificação agrícola. Estatística Geográfica Inferencial. Ciências Agrárias.

Abstract

This article investigates the territorial patterns of municipal agricultural diversification in the state of Mato Grosso do Sul, Brazil, based on the analysis of six traditional food crops (banana, sweet potato, beans, cassava, watermelon, and maize) using data from the Municipal Agricultural Survey (IBGE, 2024). Diversification was measured through the Shannon Index, while dominant productive regimes were identified based on the highest standardized contribution (z-score) among the analyzed crops. Bayesian contingency analysis was applied to assess the association between dominant regime and levels of diversification, and Random Forest Regression (RFR) was employed to model the multivariate structure of productive diversity. From a spatial perspective, the BYM2 model was used to estimate the latent spatial effect of the Shannon Index, and Global Moran's I was calculated to evaluate spatial autocorrelation of predicted values. It is concluded that municipal agricultural diversification constitutes a structurally organized phenomenon, whose understanding benefits from the integration of multivariate statistical modelling and cartographic territorialization.

Keywords: Agricultural diversification. Inferential Geographic Statistics. Agrarian Sciences.



Introdução

Um dos desafios contemporâneos à pesquisa no âmbito das Ciências Agrárias, e também, na Geografia, diz respeito à identificação e compreensão de diversos regimes produtivos alimentares em nível municipal. Nesse sentido, é importante reconhecermos que cada técnica quantitativa ou método teórico apresenta sempre as suas potencialidades, mas, igualmente, os seus limites.

A análise estatística, junto com o conhecimento científico geográfico, apresenta potencialidades para desvendar padrões, tendências, rupturas, agrupamentos ou fragmentações. Por isso mesmo, abordagens que se predisponham a traduzir tendências essenciais de dinâmicas agrárias, podem se beneficiar de análises como a Contingência Bayesiana e a Regressão Random Forest.

A abordagem Bayesiana tem sido cada vez mais utilizada em análises agrícolas e territoriais por permitir avaliar evidência probabilística de associação entre variáveis categóricas e estruturais (KRUSCHKE, 2015; MCELREATH, 2020). No que se refere à diversidade, aplicamos o índice de Shannon, comumente, utilizado para avaliar diversidade e estrutura produtiva por permitir capturar simultaneamente riqueza e equitabilidade dos componentes de um sistema (SHANNON, 1948; MAGURRAN, 2013).

A Contingência Bayesiana é um modelo probabilístico que avalia a associação entre duas variáveis categóricas estimando a força da evidência a favor da dependência ou independência entre elas, por meio do Fator de Bayes. Diferentemente do teste qui-quadrado clássico, ela não se limita a rejeitar ou não uma hipótese nula; ela quantifica o quanto os dados favorecem um modelo de associação em relação a um modelo de independência. Isso permite avaliar, por exemplo, se determinados regimes produtivos dominantes estão estruturalmente associados a níveis específicos de diversificação, sem reduzir a interpretação a um simples p-valor. Trata-se de uma leitura probabilística da organização territorial das categorias produtivas.

A análise de Contingência Bayesiana tem sido utilizada como alternativa probabilística aos testes clássicos de independência, permitindo avaliar a evidência de associação entre variáveis categóricas por meio de fatores de Bayes e probabilidades posteriores (KRUSCHKE, 2015; WAGENMAKERS et al., 2018).

Já a Regressão Random Forest é um método de aprendizado de máquina baseado na combinação de múltiplas árvores de decisão (ensemble learning), que estima uma variável contínua a partir de relações potencialmente não lineares e interativas entre preditores. Sua principal força é capturar estruturas complexas sem impor forma funcional prévia (como linearidade), além de fornecer métricas de importância das variáveis e valores preditos, município a município (BREIMAN, 2001; JAMES et al., 2013).

O objetivo, com o presente artigo, foi investigar os padrões territoriais de seis culturas alimentares tradicionais nos municípios de Mato Grosso do Sul: banana, batata-doce, feijão, melancia, mandioca e milho. Nesse aspecto, o estudo pode contribuir com trabalhos futuros para modelar a diversificação produtiva ou outros indicadores estruturais como resultado emergente de combinações produtivas, revelando hierarquias latentes e possibilitando posterior territorialização das predições.

Material e métodos

Os dados utilizados neste estudo foram obtidos na Pesquisa Agrícola Municipal (PAM) do Instituto Brasileiro de Geografia e Estatística (IBGE), disponíveis no Sistema IBGE de Recuperação

Automática – SIDRA. A PAM fornece informações anuais sobre área plantada, área colhida, quantidade produzida e valor da produção agrícola para todos os municípios brasileiros. Para este trabalho, foram selecionadas as culturas alimentares de banana, batata-doce, feijão, mandioca, melancia e milho, referentes ao ano de 2024, considerando sua relevância estrutural na composição produtiva municipal e sua representatividade territorial no estado analisado (IBGE, 2024).

As variáveis de produção municipal foram transformadas por meio da função logarítmica natural na forma $\ln(x + 1)$, procedimento amplamente utilizado para reduzir assimetria, estabilizar variâncias e minimizar a influência de valores extremos em distribuições tipicamente não normais (OSBORNE, 2002). Posteriormente, as variáveis foram padronizadas via escore-z (z-score), permitindo comparabilidade entre culturas com diferentes magnitudes produtivas e evitando que variáveis de maior escala dominassem as análises multivariadas subsequentes (FIELD, 2013). Essa dupla transformação assegura maior robustez estatística na modelagem inferencial.

Para cada município, foi identificado o regime produtivo dominante, definido como a cultura com maior valor padronizado (z-score) entre as seis analisadas. Esse procedimento permite classificar os municípios segundo sua especialização relativa, preservando comparabilidade estatística entre variáveis previamente normalizadas. A estratégia está alinhada à literatura sobre especialização produtiva regional, que utiliza critérios de predominância relativa para identificar padrões estruturais territoriais (KRUGMAN, 1991). O regime dominante foi posteriormente utilizado como variável categórica em análises probabilísticas.

A diversidade produtiva municipal foi mensurada por meio do Índice de Shannon, originalmente desenvolvido na teoria da informação (SHANNON, 1948) e amplamente aplicado em estudos ecológicos e econômicos para mensurar diversidade estrutural. O índice foi calculado a partir da proporção relativa de cada cultura no conjunto produtivo municipal, permitindo capturar simultaneamente riqueza (número de culturas presentes) e equitabilidade (distribuição relativa entre elas). Em contexto produtivo, valores mais elevados de H' indicam maior diversificação agrícola municipal.

Com o objetivo de investigar a associação probabilística entre regime produtivo dominante e níveis de diversificação agrícola (Shannon categorizado em quartis), foi aplicada uma análise de Contingência Bayesiana. Diferentemente da abordagem frequentista tradicional, a inferência Bayesiana permite quantificar evidência relativa entre modelos por meio do Fator de Bayes (Bayes Factor), fornecendo medida contínua de suporte à hipótese de dependência entre as variáveis categóricas (KASS; RAFTERY, 1995). O procedimento possibilita avaliar se determinados regimes produtivos apresentam maior probabilidade de associação com níveis específicos de diversificação municipal.

Para modelar a relação entre as culturas agrícolas e o índice contínuo de Shannon, foi aplicada uma Random Forest Regression (RFR), método de aprendizado de máquina baseado em agregação de múltiplas árvores de decisão construídas sobre amostras bootstrap e seleção aleatória de preditores (BREIMAN, 2001). A RFR permite capturar relações não lineares e interações complexas sem impor pressupostos paramétricos estritos, sendo particularmente adequada para estruturas produtivas potencialmente heterogêneas. O objetivo foi estimar a capacidade preditiva das culturas sobre o nível de diversificação municipal e identificar sua importância relativa na estrutura produtiva.

No âmbito da investigação dos padrões espaciais, foram produzidos dois mapas. O primeiro, utilizando a técnica BYM2 (Besag–York–Mollie reparametrizado), uma formulação Bayesiana hierárquica para dados areais que decompõe a variação espacial em dois componentes: um estruturado (dependente da vizinhança, geralmente modelado por um CAR – *Conditional Autoregressive*) e outro

não estruturado (efeito aleatório i.i.d.). Na parametrização BYM2, esses dois componentes são combinados por meio de um parâmetro de mistura (ϕ) que controla a proporção da variância atribuída à estrutura espacial, permitindo interpretação mais estável e comparável entre modelos. Quando se mapeia o efeito espacial latente, o que se representa é a tendência territorial acima ou abaixo do padrão médio, isto é, uma estrutura espacial não observável diretamente nos dados brutos, mas inferida pelo modelo como padrão residual persistente no território (RIEBLER et al., 2016; SIMPSON et al., 2017).

O segundo mapa elaborado diz respeito ao valor predito da diversificação (Shannon) estimado pela RFR. A cartografia, nesse contexto, não constitui etapa exploratória isolada, mas sim instrumento de interpretação espacial das estruturas produtivas estimadas, permitindo visualizar padrões territoriais derivados da modelagem preditiva.

Resultados e discussão

A partir da identificação do regime produtivo dominante em cada município e, também, da divisão em quartis do Índice Shannon, elaboramos uma tabela de Contingência Bayesiana. Os resultados estão apresentados na tabela 1.

Tabela 1 – Contingência Bayesiana

| Regime | | Q_Shannon | | | | Total |
|----------|----------|-----------|----------|----------|----------|----------|
| | | 1 | 2 | 3 | 4 | |
| Banana | Contagem | 1.00 | 1.00 | 0.00 | 0.00 | 2.00 |
| | % | 5.00 % | 5.26 % | 0.00 % | 0.00 % | 2.53 % |
| Mandioca | Contagem | 11.00 | 6.00 | 4.00 | 1.00 | 22.00 |
| | % | 55.00 % | 31.58 % | 20.00 % | 5.00 % | 27.85 % |
| Melancia | Contagem | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | % | 5.00 % | 0.00 % | 0.00 % | 0.00 % | 1.27 % |
| Milho | Contagem | 7.00 | 12.00 | 16.00 | 19.00 | 54.00 |
| | % | 35.00 % | 63.16 % | 80.00 % | 95.00 % | 68.35 % |
| Total | Contagem | 20.00 | 19.00 | 20.00 | 20.00 | 79.00 |
| | % | 100.00 % | 100.00 % | 100.00 % | 100.00 % | 100.00 % |

A tabela 1 evidencia uma distribuição assimétrica do regime produtivo dominante ao longo dos quartis de diversificação (Q_Shannon), sendo que os quartis 1 e 2 representam os níveis mais altos de diversidade, enquanto 3 e 4 indicam menor diversidade. Observa-se que, nos quartis superiores (1 e 2), há presença expressiva de municípios dominados por mandioca (55% no Q1 e 31,58% no Q2), enquanto o milho, embora ainda relevante (35% no Q1 e 63,16% no Q2), não é hegemônico nos níveis mais altos de diversidade.

Destaca-se que as culturas de batata-doce e feijão não aparecem como regimes produtivos dominantes na tabela 1, pois, embora estejam presentes na base de dados, não constituíram a principal cultura em nenhum dos municípios analisados segundo o critério de dominância adotado. Como a tabela apresenta apenas os regimes produtivos dominantes, essas culturas não foram incluídas na

estrutura de contingência. Essa ausência não indica inexistência produtiva, mas apenas que essas culturas não representaram a maior participação relativa dentro da estrutura produtiva municipal.

Isso sugere que, nos territórios mais diversificados, a estrutura produtiva tende a ser relativamente mais equilibrada ou associada a gêneros alimentares como a mandioca, que historicamente se integra a sistemas agrícolas mais heterogêneos e policulturais.

Em contraste, nos quartis inferiores (3 e 4), que representam menor diversificação, a dominância do milho torna-se praticamente absoluta (80% no Q3 e 95% no Q4), enquanto a mandioca perde participação relativa (20% no Q3 e apenas 5% no Q4). Esse padrão indica uma associação territorial consistente entre especialização em milho e redução da diversidade produtiva municipal.

Os dados sugerem que a especialização cerealista se vincula a estruturas produtivas mais concentradas, ao passo que regimes baseados em mandioca aparecem mais frequentemente nos estratos superiores de diversidade. Trata-se, portanto, de uma evidência empírica coerente com a hipótese central desse artigo: os regimes produtivos dominantes organizam diferencialmente os níveis territoriais de diversificação. A figura 1 apresenta o mapa BYM2 do Índice Shannon.

Efeito Espacial Latente do Índice de Shannon em Mato Grosso do Sul (Modelo BYM2)

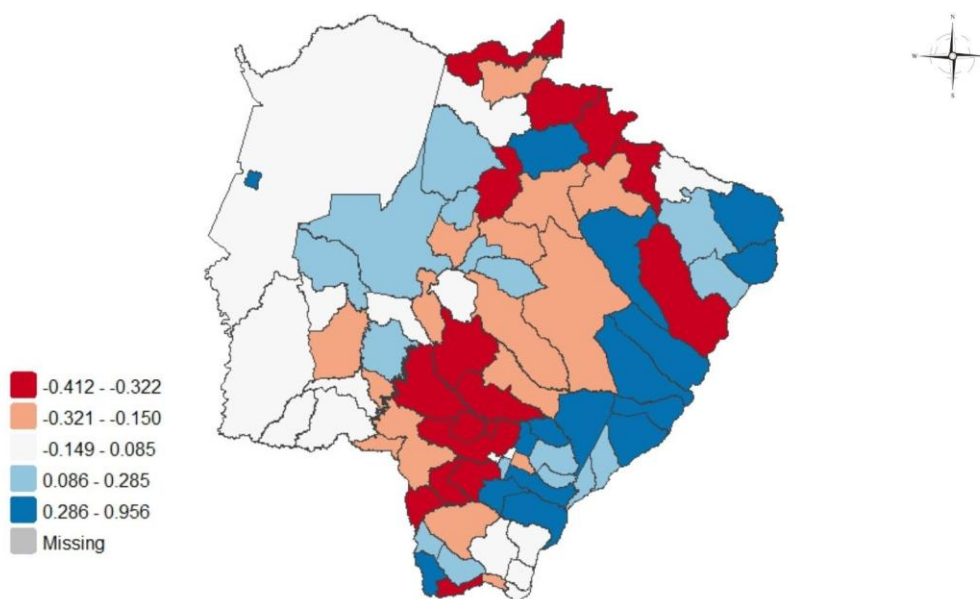


Figura 1 – Mapa BYM2 do Índice Shannon.

O mapa da figura 1 apresenta o efeito espacial latente do Índice de Shannon estimado pelo modelo BYM2, ou seja, uma estrutura territorial inferida a partir da dependência espacial entre municípios, separando a variação estruturada (associada à vizinhança) da variação não estruturada. As áreas em azul indicam tendência territorial acima do padrão médio, enquanto as áreas em vermelho revelam tendência territorial abaixo do padrão médio estadual, não como reflexo direto dos valores observados, mas como padrão residual persistente captado pelo modelo.

Essa abordagem dialoga com a tradição da análise espacial que compreende que processos socioeconômicos e produtivos tendem a apresentar interdependência espacial sistemática, exigindo

modelagem explícita dessa estrutura (ANSELIN, 1988; BESAG; YORK; MOLLIE, 1991). No contexto da produção do conhecimento, essa leitura permite compreender a diversificação produtiva como fenômeno territorialmente organizado, e não apenas como atributo isolado de cada unidade administrativa.

O uso do BYM2 reforça a incorporação de métodos quantitativos espacialmente estruturados na análise das transformações agrárias, ampliando a capacidade de identificar regimes produtivos regionalizados, continuidades espaciais e zonas de transição agrária. Ao evidenciar padrões latentes persistentes, o modelo contribui para a compreensão das dinâmicas de especialização e diversificação agrícola sob uma perspectiva relacional do espaço, na qual as configurações territoriais resultam de processos históricos e econômicos interdependentes (GOODCHILD, 2004; WARD; BROWN, 2009).

Pode-se observar tendências territoriais acima do padrão médio (em tons de azul) concentradas sobretudo na porção leste e sudeste do estado, enquanto padrões abaixo do padrão médio (em tons de vermelho) se destacam com maior intensidade na região central e em parte do norte, formando blocos espaciais relativamente contínuos; o valor médio de $\phi = 0,69$ indica que cerca de 69% da variância do efeito latente é explicada pela componente espacial estruturada, evidenciando forte predominância da dependência territorial na organização da diversificação produtiva. Assim, o mapa não implica causalidade, mas oferece uma representação inferencial da organização espacial da diversificação produtiva, potencialmente útil para análises comparativas e planejamento territorial.

Nesse sentido, no prosseguimento de nossas análises, elaboramos uma *Random Forest Regression* (RFR) estimada com 300 árvores e divisão treino-teste (64 observações para treino e 15 para teste), apresentando estabilidade interna conforme indicado pelo erro out-of-bag. As métricas de desempenho, apresentadas na tabela 2. Inserimos o Índice Shannon como variável resposta e as culturas analisadas enquanto variáveis dependentes.

Tabela 2 – Métricas de performance

| | MSE | MSE | RMSE | MAE / MAD | MAPE | R² |
|---------|------------|------------|-------------|------------------|-------------|----------------------|
| Valores | 0.083 | 0.496 | 0.288 | 0.222 | 183.39% | 0.539 |

De acordo com os dados da tabela 2, percebemos que a RFR indicou capacidade explicativa moderada ($R^2 = 0,539$), com erro quadrático médio ($MSE = 0,083$) e $RMSE = 0,288$. Esses valores sugerem que mais da metade da variação do Índice de Shannon é explicada pela composição produtiva municipal, evidenciando que a diversificação agrícola não ocorre de maneira aleatória, mas apresenta estrutura estatisticamente captável. Conforme Breiman (2001), a agregação de múltiplas árvores reduz variância e melhora a robustez preditiva, permitindo modelar relações não lineares típicas de sistemas produtivos territoriais. A tabela 3 explicita as métricas de importância da RFR de cada variável utilizada.

Tabela 3 – Métricas de importância das variáveis

| | Redução média da acurácia | Aumento total da pureza dos nós | Perda média por omissão da variável |
|----------|----------------------------------|--|--|
| Milho | 0.036 | 0.867 | 0.247 |
| Mandioca | 0.022 | 0.679 | 0.221 |
| Melancia | 0.009 | 0.470 | 0.191 |
| Banana | 0.004 | 0.227 | 0.171 |
| Batata | 8.843×10^{-4} | 0.221 | 0.163 |
| Feijao | 0.002 | 0.199 | 0.168 |

A tabela 3 apresenta as métricas de importância das variáveis, evidenciando o milho como principal preditor da diversificação produtiva, seguido pela mandioca. O milho apresenta os maiores valores tanto na redução média de acurácia quanto no aumento total da pureza dos nós, indicando que sua remoção implicaria maior perda de desempenho do modelo. Esse resultado é relevante, pois indica que determinadas culturas estruturantes ocupam posição central na organização produtiva municipal. Atividades com maior peso econômico podem funcionar como eixos estruturadores do território, articulando diferentes escalas de produção e condicionando arranjos produtivos locais (SANTOS, 1996).

A mandioca, por sua vez, embora com menor magnitude preditiva que o milho, também exerce influência relevante, enquanto culturas como batata-doce e feijão apresentam contribuição marginal na explicação da variabilidade do Índice de Shannon. Esse padrão indica que a diversificação produtiva não depende apenas da presença numérica de culturas, mas da posição estrutural que determinadas atividades ocupam no sistema produtivo.

Conforme argumenta Krugman (1991), efeitos de escala e organização espacial podem gerar complexificação econômica mesmo em contextos de aparente especialização. Assim, a importância relativa das culturas revela uma dinâmica territorial na qual especialização e diversidade não são categorias necessariamente opostas. A figura 2 apresenta o gráfico da correlação preditiva entre os ranks preditos verificados com a RFR e os ranks do próprio Índice Shannon.

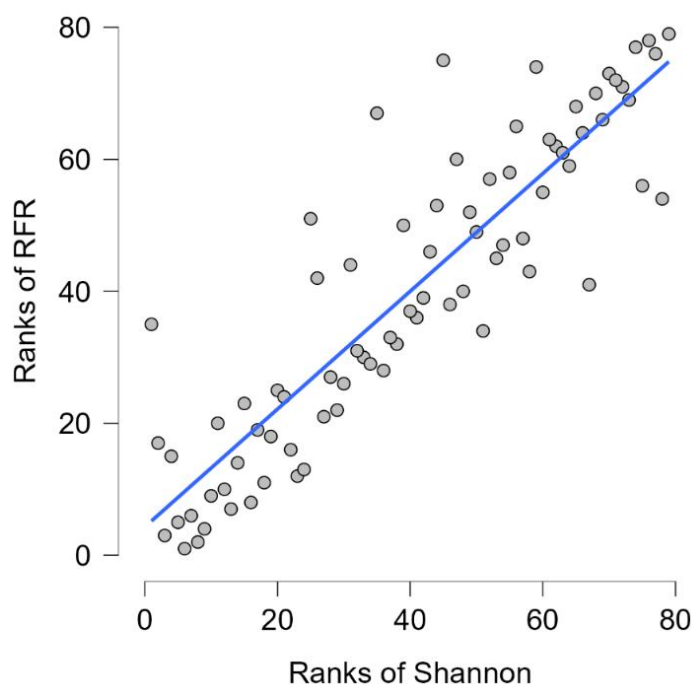


Figura 2 – Gráfico de correlação Bayesiana.

A correlação Bayesiana é uma medida da associação entre duas variáveis estimada a partir da distribuição posterior do coeficiente de correlação, permitindo quantificar a magnitude do efeito e, também, a evidência probabilística a favor ou contra a existência de associação. A correlação Bayesiana entre o Índice de Shannon observado e os valores estimados pela Random Forest Regression (RFR) apresentou Kendall's $\tau = 0,732$, com $BF_{10} = 3,887 \times 10^{18}$, indicando associação monotônica forte e evidência decisiva a favor da hipótese de correlação. Isso pode ser visualizado na

figura 2, pois os pontos estão fortemente alinhados à linha reta indicam elevada concordância entre os rankings do Shannon observado e do predito pela RFR, sugerindo forte associação monotônica e consistência preditiva do modelo.

Importante esclarecer que os ranks da figura 2 são ordenação dos municípios do menor para o maior valor das variáveis (Shannon observado e Shannon predito pela RFR). Esse resultado nos demonstra que a ordenação municipal produzida pelo modelo preditivo reproduz de forma consistente a hierarquia observada de diversificação produtiva. O Fator de Bayes extremamente elevado reforça que a relação não é produto de flutuação amostral, mas reflete estrutura sistemática capturada pela modelagem. Assim, a RFR demonstra não apenas capacidade explicativa (R^2 já explicitado anteriormente), mas também coerência ordinal robusta entre valores previstos e observados, justificando sua utilização como instrumento válido para inferência preditiva e posterior territorialização dos resultados, como apresentado na figura 3.

Diversificação Produtiva Municipal Estimada por Random Forest Regression em Mato Grosso do Sul (2024)

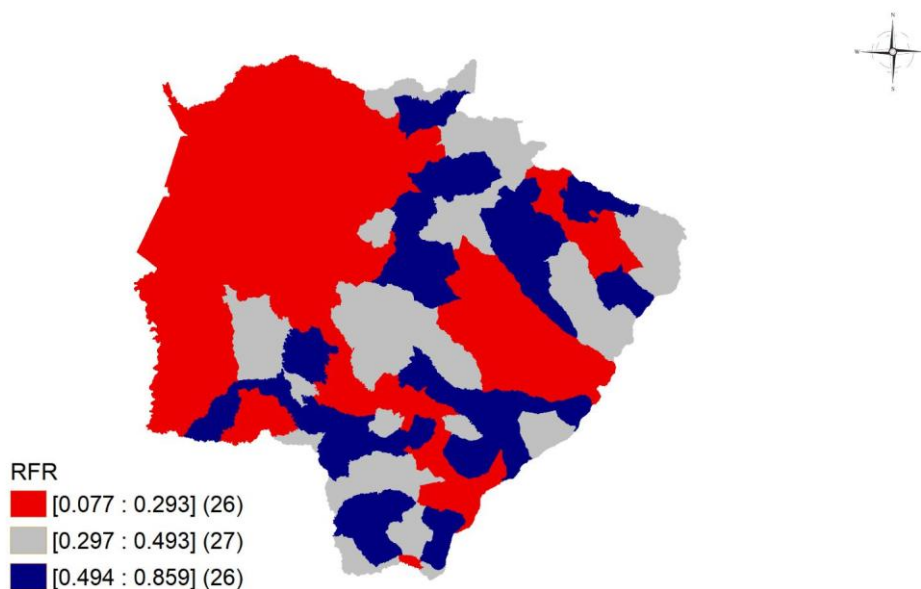


Figura 3 – Mapa da diversificação produtiva.

Com o objetivo de verificar a autocorrelação espacial sobre os valores preditos pela RFR, realizamos um Moran's I Global. Obtivemos um valor de $I = 0,057$, indicando uma autocorrelação espacial positiva muito fraca, sugerindo que os valores preditos pela Random Forest Regression (RFR) apresentam apenas um leve padrão de similaridade entre municípios vizinhos definidos pela matriz Queen de ordem 2. O pseudo p-valor obtido por 999 permutações ($p = 0,076$) não atinge o nível convencional de significância estatística de 5%, o que implica que não há evidência robusta para rejeitar a hipótese nula de aleatoriedade espacial.

Esse resultado significa que a distribuição territorial dos valores preditos pelo modelo RFR não apresenta dependência espacial estatisticamente confirmada, estando mais próxima de um arranjo espacial aleatório do que de um cluster estruturado. O uso do procedimento de permutação reforça essa interpretação, pois o pseudo p-valor deriva da comparação entre o índice observado e a

distribuição empírica gerada por rearranjos aleatórios dos dados no espaço, conforme proposto por Anselin (1995). Assim, embora haja um leve sinal de associação espacial positiva, ele é fraco e marginalmente significativo, recomendando cautela na interpretação de padrões espaciais globais.

O mapa da figura 3, por sua vez, explicita a diversificação produtiva estimada pela RFR evidenciando a distribuição territorial dos níveis previstos de Shannon em três classes, revelando contrastes espaciais significativos entre municípios de baixa, média e alta diversificação. Observamos a presença de extensas áreas com níveis mais baixos de diversificação, intercaladas por corredores e núcleos de maior intensidade estimada, sugerindo que a diversidade produtiva não se distribui de forma homogênea no território estadual.

Diferentemente do mapa BYM2 que explicitou a estrutura espacial latente do Índice Shannon, o mapa da figura 3 traduz uma estrutura inferida pelo modelo estatístico RFR, isto é, a intensidade estimada da diversificação a partir da combinação das culturas analisadas. A territorialização do valor predito permite visualizar como a estrutura produtiva, capturada estatisticamente, se materializa espacialmente, reforçando a importância da integração entre modelagem quantitativa e interpretação geográfica. Nesse sentido, o mapa não representa apenas dados observados, mas a espacialização de uma regularidade estrutural inferida, aproximando estatística preditiva e análise territorial em perspectiva complementar.

Percebemos que a classe de menor diversificação (vermelho) concentra-se predominantemente na porção oeste e centro-oeste do estado, enquanto a classe de maior diversificação (azul) apresenta maior presença no sul e em áreas do leste, com a classe intermediária distribuindo-se de forma mais fragmentada na faixa central e nordeste.

A análise conjunta dos resultados (contingência Bayesiana, BYM2, Random Forest Regression, correlação ordinal robusta e territorialização cartográfica) indica que a diversificação produtiva municipal em Mato Grosso do Sul não se configura como fenômeno aleatório, mas como expressão de uma estrutura produtiva organizada. A associação decisiva entre regimes dominantes e níveis de Shannon, aliada à capacidade explicativa moderada da RFR ($R^2 = 0,539$) e à forte correlação entre valores observados e previstos ($\tau = 0,732$), revela coerência sistêmica entre especialização relativa e diversidade estrutural.

Esse padrão dialoga com a literatura que interpreta os territórios como sistemas complexos, nos quais especialização e diversificação podem coexistir de forma complementar, dependendo da escala, da inserção em mercados e da organização produtiva regional (ILBERY, 1998; MARSDEN, 1998). Assim, a evidência estatística e espacial sugere que culturas estruturantes, como o milho, não necessariamente restringem a diversidade, podendo integrar arranjos produtivos multifuncionais.

Além disso, a integração entre modelagem estatística e leitura espacial reforça abordagens contemporâneas que defendem o uso de métodos quantitativos para revelar regularidades estruturais na organização agrária (WOODS, 2011). A Random Forest, ao captar relações não lineares e hierarquias de importância entre culturas, evidencia que a diversificação emerge da posição estrutural que determinadas atividades ocupam no sistema produtivo, e não apenas da simples contagem de culturas. A territorialização dos valores preditos amplia essa leitura ao demonstrar que padrões estatísticos globais possuem expressão espacial concreta.

Nesse sentido, os resultados contribuem para a pesquisa agrária ao demonstrar empiricamente que a análise integrada entre inferência estatística e cartografia permite compreender como estruturas produtivas se organizam, se distribuem e se articulam territorialmente, oferecendo base analítica mais robusta para interpretação dos regimes agrícolas contemporâneos.

Considerações finais

Os resultados demonstram que a diversificação produtiva municipal em Mato Grosso do Sul não se distribui de maneira aleatória, mas expressa uma organização estrutural associada aos regimes produtivos dominantes. A evidência probabilística fornecida pela Contingência Bayesiana reforça que determinados regimes, — especialmente a dominância do milho, se associam, de modo sistêmico, a níveis inferiores de diversificação, enquanto a mandioca aparece com maior frequência nos estratos superiores de diversidade. Essa configuração sugere que especialização e diversidade não constituem categorias mutuamente excludentes, mas podem organizar-se de forma diferenciada no território.

Do ponto de vista metodológico, a integração entre Random Forest Regression, análise Bayesiana e modelagem espacial (BYM2) evidencia o potencial da Estatística Geográfica Inferencial para revelar estruturas produtivas latentes e hierarquias territoriais. O efeito espacial latente identificado ($\phi = 0,69$) indica predominância da dependência territorial na organização da diversidade, enquanto a fraca autocorrelação dos valores preditos sugere que a estrutura multivariada captada pela RFR não se traduz automaticamente em clusterização espacial global. Assim, o artigo contribui ao demonstrar que a articulação entre inferência estatística robusta e leitura espacial permite compreender com maior precisão a dinâmica dos regimes agrícolas contemporâneos, ampliando o repertório analítico sem recorrer a pressupostos causalistas indevidos.

Referências bibliográficas

- ANSELIN, L. Local Indicators of Spatial Association - LISA. **Geographical Analysis**, v. 27, n. 2, p. 93-115, 1995. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- ANSELIN, L. **Spatial econometrics: Methods and Models**. Dordrecht: Springer, 1988, 284p. <https://doi.org/10.1007/978-94-015-7799-1>
- BESAG, J.; YORK, J.; MOLLIE, A. Bayesian image restoration, with two applications in spatial statistics. **Annals of the Institute of Statistical Mathematics**, v. 43, p. 1-20, 1991. <https://doi.org/10.1007/BF00116466>
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32, 2001. <https://doi.org/10.1023/A:1010933404324>
- FIELD, A. **Discovering statistics using IBM SPSS statistics**. London: Sage Publications, 2013, 952p. https://books.google.com.br/books/about/Discovering_Statistics_Using_IBM_SPSS_St.html
- GOODCHILD, M. F. GIScience, Geography, Form, and Process. **Annals of the Association of American Geographers**, v. 94, n. 4, p. 709-714, 2004. <https://doi.org/10.1111/j.1467-8306.2004.00424.x>
- IBGE. Instituto Brasileiro de Geografia e Estatística. **Pesquisa Agrícola Municipal – PAM 2024**. Rio de Janeiro: IBGE, 2024. <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>
- ILBERY, B. **The Geography of Rural Change**. London: Longman, 1998, 267p. https://books.google.com.br/books/about/The_Geography_of_Rural_Change.html
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in R**. 1st edition. New York: Springer, 2013, 426p. <https://doi.org/10.1007/978-1-4614-7138-7>
- KASS, R. E.; RAFTERY, A. E. Bayes factors. **Journal of the American Statistical Association**, v. 90, n. 430, p. 773-795, 1995. <https://doi.org/10.1080/01621459.1995.10476572>

- KRUGMAN, P. **Geography and Trade**. Cambridge: The MIT Press, 1992, 156p. <https://mitpress.mit.edu/9780262610865/geography-and-trade/>
- KRUSCHKE, J. **Doing Bayesian Data Analysis**. Academic Press, 2015. <https://www.sciencedirect.com/book/monograph/9780124058880/doing-bayesian-data-analysis>
- MAGURRAN, A. E. **Measuring Biological Diversity**. Wiley, 2013, 272p. https://books.google.com.br/books/about/Measuring_Biological_Diversity.html
- MARSDEN, T. New rural territories: regulating the differentiated rural spaces. **Journal of Rural Studies**, v. 14, n. 1, p. 107-117, 1998. [https://doi.org/10.1016/S0743-0167\(97\)00041-7](https://doi.org/10.1016/S0743-0167(97)00041-7)
- MCELREATH, R. **Statistical Rethinking. A Bayesian Course with Examples in R and STAN**. 2nd edition. New York: CRC Press, 2020, 612p. <https://doi.org/10.1201/9780429029608>
- OSBORNE, J. Notes on the use of data transformations. **Practical Assessment, Research & Evaluation**, v. 8, n. 1, p. 1-7, 2002. <https://doi.org/10.7275/4vng-5608>
- RIEBLER, A.; SORBYE, S. H.; RUE, H.; SIMPSON, D. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. **Statistical Methods in Medical Research**, v. 25, n. 4, p. 1145-1165, 2016. <https://doi.org/10.1177/0962280216660421>
- SANTOS, M. **A Natureza do Espaço: Técnica e Tempo, Razão e Emoção**. 4^a edição. São Paulo: EDUSP, 1996, 392p. https://books.google.com.br/books/about/A_Natureza_do_Espaco.html
- SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, n. 3, p. 379-423, 1948. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- SIMPSON, D.; RUE, H.; RIEBLER, A.; MARTINS, T. G.; SORBYE, S. H. Penalising model component complexity: a principled, practical approach to constructing priors. **Statistical Science**, v. 32, n. 1, p. 1-28, 2017. <https://doi.org/10.1214/16-STS576>
- WAGENMAKERS, E.-J.; MARSMAN, M.; JAMIL, T.; LY, A.; VERHAGEN, J.; LOVE, J.; SELKER, R.; GRONAU, Q. F.; ŠMÍRA, M.; EPSKAMP, S.; MATZKE, D.; ROUDER, J. N.; MOREY, R. D. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. **Psychonomic Bulletin & Review**, v. 25, p. 35-37, 2018. <https://doi.org/10.3758/s13423-017-1343-3>
- WARD, N.; BROWN, D. L. Placing the rural in regional development. **Regional Studies**, v. 43, n. 10, p. 1237-1244, 2009. <https://doi.org/10.1080/00343400903234696>
- WOODS, M. **Rural**. 1st edition. London: Routledge, 2011, 352p. <https://www.routledge.com/Rural/Woods/p/book/9780415442404>

Recebido em 25 de fevereiro de 2026

Retornado para ajustes em 11 de março de 2026

Recebido com ajustes em 11 de fevereiro de 2026

Aceito em 12 de março de 2026