

UNIVERSIDADE FEDERAL DE VIÇOSA

DIEGO ANDREÃO FALQUETTO

ANÁLISE DE AGRUPAMENTO DE IMÓVEIS RURAIS PARA ESTIMATIVA DO VALOR DA TERRA
NUA UTILIZANDO K-MEANS

VIÇOSA – MINAS GERAIS

2022

DIEGO ANDREÃO FALQUETTO

ANÁLISE DE AGRUPAMENTO DE IMÓVEIS RURAIS PARA ESTIMATIVA DO VALOR DA TERRA
NUA UTILIZANDO K-MEANS

Relatório final, apresentado a Universidade Federal de Viçosa, como parte das exigências para obtenção do título de Engenheiro Agrícola e Ambiental.

Orientadora: Flora Maria de Melo Villar

Coorientadora: Poliana Maria da Costa
Bandeira

VIÇOSA – MINAS GERAIS

2022

DIEGO ANDREÃO FALQUETTO

ANÁLISE DE AGRUPAMENTO DE IMÓVEIS RURAIS PARA ESTIMATIVA DO VALOR DA TERRA
NUA UTILIZANDO K-MEANS

Relatório final, apresentado a Universidade Federal de Viçosa, como parte das exigências, para obtenção do título de Engenheiro Agrícola e Ambiental.

APROVADO: 22 de março de 2022.

Assentimento:



Diego Andreão Falquetto
Autor



Documento assinado digitalmente
Flora Maria de Melo Villar
Data: 05/04/2022 16:25:14-0300
Verifique em <https://verificador.iti.br>

Flora Maria de Melo Villar
Orientadora

RESUMO

FALQUETTO, Diego Andreão, Bacharel em Engenharia Agrícola e Ambiental, Universidade Federal de Viçosa, março de 2022. **Análise de agrupamento de imóveis rurais para estimativa do valor da terra nua utilizando *k-means***. Orientadora: Flora Maria de Melo Villar. Coorientadora: Poliana Maria da Costa Bandeira.

A análise para precificação de imóveis rurais leva em conta aspectos de terra nua, benfeitorias e aspectos gerais da propriedade em questão. No que se refere ao valor da terra nua (VTN), são analisados fatores quanto a produção, logística, meio ambiente e disponibilidade de recursos. Apesar de haver algumas metodologias para determinação do VTN, cada uma abrange grupos restritos de variáveis, bem como são atribuídos pesos diferentes para uma mesma variável. Diante deste cenário, objetivou-se com esse trabalho consolidar uma base de dados que abrangesse informações de produção, logística, meio ambiente e disponibilidade de recursos para os imóveis rurais contidos na base de dados do Cadastro Ambiental Rural (CAR) do Mato Grosso a fim de aplicar o algoritmo não supervisionado de *Machine Learning K-Means*, agrupando-os e permitindo determinação de valores de referência semelhantes para propriedades contidas nos mesmos *clusters*. Para isso, foram determinadas 21 variáveis a partir de diferentes bases de dados públicas, para os mais de 165 mil imóveis rurais únicos contidos na base de dados do CAR. Foram formados 12 *clusters*, que apresentaram diferentes características para as variáveis de entrada, com destaque para as variáveis área de agropecuária, e as referentes aos tipos de solo areias quartzosas, cambissolo, latossolo, podzólico, plintossolo e concrecionado, que apresentaram as maiores representatividades no modelo. Além disso, os agrupamentos transcenderam os limites regionais do Estado, possibilitando a comparação entre imóveis de diferentes locais. Assim, os *clusters* gerados foram satisfatórios, seja pela distinção notória perante análise individual das características analisadas, seja pela abrangência não regional, possibilitando mais do que comparação entre VTN em regiões distintas, mas possibilitando também futuros trabalhos de determinação de imóveis com características semelhantes para produção agrícola.

Palavras-chave: não-supervisionado; machine learning; preço de imóveis.

ABCSTRACT

FALQUETTO, Diego Andreão, Bacharel em Engenharia Agrícola e Ambiental, Universidade Federal de Viçosa, March, 2022. **Clustering analysis of rural properties to estimate the value of bare land using k-means.** Adviser: Flora Maria de Melo Villar. Co-adviser: Poliana Maria da Costa Bandeira.

The analysis for pricing rural properties takes into account aspects of bare land, improvements and general aspects of the property in question. With regard to the value of bare land (VTN), factors related to production, logistics, environment and resource availability are analyzed. Although there are some methodologies for determining the VTN, each one covers restricted groups of variables, as well as different weights are assigned to the same variable. Given this scenario, the objective of this work was to consolidate a database that included information on production, logistics, environment and availability of resources for rural properties contained in the database of the Rural Environmental Registry (CAR) of Mato Grosso in order to of applying the unsupervised K-Means Machine Learning algorithm, grouping them and allowing the determination of similar reference values for properties contained in the same clusters. For this, 21 variables were determined from different public databases, for the more than 165 thousand unique rural properties contained in the CAR database. Twelve clusters were formed, which presented different characteristics for the input variables, with emphasis on the agricultural area variables, and those referring to the types of soil quartz sands, cambisol, latosol, podzol, plintosol and concreted, which presented the highest representativeness in the model. In addition, the groupings transcended the regional limits of the State, making it possible to compare properties from different locations. Thus, the clusters generated were satisfactory, either by the notorious distinction in the face of individual analysis of the analyzed characteristics, or by the non-regional scope, allowing more than comparison between VTN in different regions, but also allowing future works to determine properties with similar characteristics for production agricultural.

Keywords: unsupervised; machine learning; property price.

SUMÁRIO

1	INTRODUÇÃO	6
2	MATERIAL E MÉTODOS.....	8
2.1	Caracterização da área de estudo.....	8
2.2	Recursos utilizados.....	9
2.3	Seleção de critérios	10
2.4	Pré-processamento	11
2.5	Análise espacial	11
2.5.1	Restrições – Lógica booleana	12
2.5.2	Fatores – Análise escalonar.....	13
2.6	Algoritmo.....	14
2.7	Validação dos clusters.....	15
3	RESULTADOS E DISCUSSÃO	20
4	CONCLUSÕES.....	42
5	REFERÊNCIAS.....	43

1 INTRODUÇÃO

O Brasil possui mais de 5,7 milhões de imóveis rurais particulares, totalizando uma área de 521 milhões de hectares segundo o Instituto de Colonização e Reforma Agrária – INCRA (INCRA, 2016). No estado do Mato Grosso, o número de imóveis rurais é de 171.237, correspondendo a uma área de 92.488.804,15 de hectares, segundo o Sistema Nacional de Cadastro Ambiental Rural – SICAR (SICAR, 2021).

As propriedades rurais, segmentadas por todas as regiões do Brasil contribuem para o Produto Interno Bruto (PIB) do agronegócio brasileiro, calculado pelo Centro de Estudos Avançados em Economia Aplicada, da Esalq/USP, em conjunto com a Confederação da Agricultura e Pecuária do Brasil (CNA), que chegou a quase 2 trilhões de reais em 2020, uma participação de 26,6% em relação ao PIB total do país. A representatividade desse setor se deve principalmente aos recordes de safras e aumento do preço das *commodities*.

A expressividade dos números, seja em número monetário ou de propriedades, através da cadeia produtiva são resultados de diferentes níveis de produção, abrangendo desde a agricultura familiar até grandes fazendas. No Brasil, 81% dos imóveis rurais possuem menos de 50 hectares, enquanto no estado do Mato Grosso esse percentual cai para 35% (BRASIL, 2020).

A proporção do setor agrícola também é expressiva em relação a compra e venda de imóveis rurais. No estado do Mato Grosso são mais de 5.000 anúncios de vendas somente em sites de imobiliárias. Esse mercado além do grande montante fiscal movimentado, também gera impostos, como: o Imposto sobre a Propriedade Territorial Rural (ITR), Imposto de Transmissão de Bens Inter-Vivos (ITBI) e Imposto sobre Ganho de Capital da Alienação de Imóveis Rurais (INSTRUÇÃO NORMATIVA SRF No 84, 2001).

Para auxiliar na precificação da terra nua do Mato Grosso, estado com 65% das propriedades com mais de 50 hectares (BRASIL, 2020), o INCRA forneceu até 2019 um relatório de análise de mercado de terras, a partir da delimitação de mercados Regionais de Terras, realizada por análise de agrupamento utilizando os dados agrícolas dos municípios (CARMEM et al., 2019).

Com este trabalho objetivou-se agrupar imóveis rurais do Mato Grosso com características semelhantes fazendo uso de algoritmo não supervisionado de *Machine Learning* a fim de facilitar a determinação do Valor de Terra Nua (VTN). Para isso, foi gerada base de dados única

a partir de diferentes fontes públicas, seguido do agrupamento dos imóveis rurais com características semelhantes utilizando o algoritmo *K-Means* e relacionamento dos grupos formados com os valores de terra nua calculados pelo Instituto de Terras do Mato Grosso.

2 MATERIAL E MÉTODOS

O trabalho foi realizado no Departamento de Engenharia Agrícola da Universidade Federal de Viçosa (DEA/UFV), com adoção de metodologia de coleta, tratamento e análise dos dados proposta por Aquino (2018). Dessa forma, o desenvolvimento do algoritmo se deu pelas seguintes etapas:

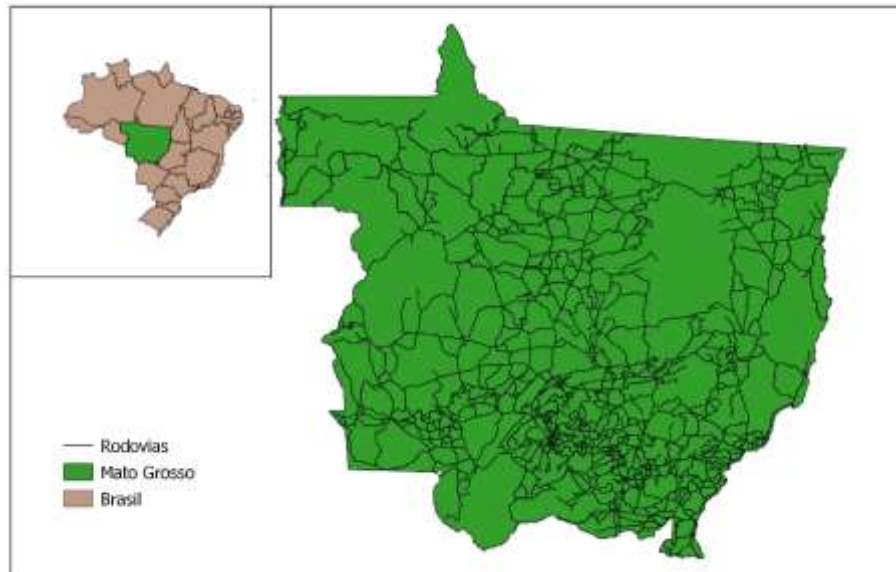
- i. Seleção de critérios: Com base em literatura e norma técnica (NBR 14.653-3) que especifica os aspectos dos imóveis rurais que devem ser caracterizados pelos critérios, foram selecionados 7 critérios com base da disponibilidade de bases de dados públicas referentes ao estado do Mato Grosso para compor o desenvolvimento do presente trabalho, são eles: Limites das propriedades rurais, rodovias, ferrovias, hidrovias, tipos de solo, disponibilidade hídrica e uso e ocupação do solo.
- ii. Obtenção dos dados: Foram utilizadas bases de dados cartográficos disponíveis nos endereços eletrônicos dos seguintes órgãos: MAPBIOMAS, MMA, SICAR.
- iii. Pré-processamento: Os sistemas de coordenadas foram padronizados, seguido por: vetorizar dados no formato *raster*, recortá-los para cada imóvel rural e padronização dos dados.
- iv. Análise Espacial: Foram calculadas as variáveis diretas relacionando as propriedades rurais ao demais critérios.
- v. Aplicação de *K-means*: Foi gerada matriz única e aplicação do algoritmo de agrupamento;
- vi. Análise e interpretação do resultado: O grau de agrupamento foi avaliado, seguido da validação dos resultados obtidos e levantamento de variáveis com maior representatividade no agrupamento.

2.1 Caracterização da área de estudo

O estado do Mato Grosso integra a região Centro-Oeste do Brasil, tendo como capital o município de Cuiabá. O estado possui 141 municípios, sendo responsável por aproximadamente 30% da produção de grãos do Brasil. Com área de 903.357 m², é o segundo maior estado em área territorial do Brasil, possui cerca de 3.567.234 habitantes e PIB de

R\$142,12 bilhões (IBGE, 2021). Na Figura 1 é apresentada localização do Mato Grosso e sua malha viária.

Figura 1. Localização do Mato Grosso e malha viária



Fonte: O autor.

2.2 Recursos utilizados

No presente trabalho foram utilizadas:

- Bases cartográficas oficiais em formato *shapefile* em escala 1:100.000 do estado do Mato Grosso, disponibilizadas pelo Ministério do Meio Ambiente (MINISTÉRIO DO MEIO AMBIENTE, 2020), estas forneceram malha rodoviária, ferroviária e hidroviária;
- Mapa de limites de propriedades rurais do estado do Mato Grosso, disponíveis no Sistema de Cadastro Ambiental Rural (SICAR, 2021);
- Mapa em formato *raster* de uso e ocupação do solo com resolução de 30 por 30 metros, disponibilizados pelo MAPBIOMAS (MAPBIOMAS, 2020), o qual forneceu informações sobre uso e ocupação do solo e disponibilidade hídrica;
- Mapa de pedologia no formato *shapefile*, em escala 1:250.000, disponível no Instituto de Terras de Mato Grosso (INTERMAT, 2000).

Além disso foi utilizado *hardware* um notebook Intel Core i7, 8GB de RAM e SSD de 500 GB, *software* QGIS 3.16.12 (QGIS.ORG, 2021), Python (PYTHON SOFTWARE FOUNDATION, 2013) com as bibliotecas *GeoPandas* 0.9 (GEOPANDAS, 2021) e *Rasterio* (RASTERIO, 2018).

2.3 Seleção de critérios

Para determinação das variáveis utilizadas no agrupamento foi utilizada a NBR 14653-3 (2004), a qual define como critérios de avaliação de terra nua fatores relacionados à homogeneidade da área, capacidade de uso das terras, situação, e outros fatores, como, por exemplo, a disponibilidade de recursos hídricos. A norma técnica permite que fatores externos sejam considerados na determinação do valor de terra nua, a depender das especificidades da região. Ao analisar a área de estudo, foi observada a proporção de 69,84% de rodovias estaduais e federais não pavimentadas no Mato Grosso, e por isso, o fator logístico de transporte também foi considerado no estudo. Assim, a seleção dos critérios buscou abranger características de produção agrícola, logística para transporte de cargas e qualidade de vida dos residentes, e respeito das normas ambientais quanto a área mínima de preservação Tabela 1.

Tabela 1. Delimitação da Área de Reserva Legal

Localização	Área	Percentual
	Floresta	80%
Amazônia Legal	Cerrado	35%
	Campos Gerais	20%
Demais regiões	-	20%

Fonte: Lei n 12.651, de 25 de maio de 2012.

Levando em conta os aspectos homogeneidade, situação, disponibilidade de recursos e logística foram definidas as seguintes variáveis para composição da base de dados: áreas: total, agropecuária e preservação permanente; presença ou não de: ferrovia, hidrovia e recurso hídrico; Distância de rodovias asfaltadas; Áreas por tipos de solo: areias quartzosas, aluviais, cambissolo, gleissolo, orgânico, latossolo, planossolo, podzólico, plintossolo, litólico, concrecionado, solodizado e terra roxa. Assim, objetiva-se desenvolver algoritmo capaz de agrupar os imóveis rurais com características comuns em aspectos de terra nua, a fim de auxiliar precificações destes imóveis.

2.4 Pré-processamento

Em detrimento das diferentes fontes das bases cartográficas descritas no item 2.2, houve a necessidade de padronização de sistema de coordenadas, ou seja, verificação de inconsistências nos dados. As etapas de pré-processamento dos dados são apresentadas a seguir, todas realizadas utilizando *Python* 3 e biblioteca *GeoPandas*.

- Padronizar o sistema de coordenadas por meio de sobrescrita do atributo “*crs*” contido *GeoDataFrame* com a ferramenta “*to_crs*”, todas foram transformadas para o sistema de coordenadas projetadas SIRGAS 2000 (EPSG:4674), utilizando *Python*. Para as etapas de cálculo de área, foi necessário conversão para sistema de coordenadas para WGS 84 (EPSG: 3857);
- Para a camada *raster* de Uso e Ocupação do Solo foi recortada a área de interesse por meio da ferramenta “*Recortar raster pela camada máscara*” do software QGIS;
- A ferramenta “*Poligonizar*” do software QGIS foi utilizada para transformação da camada *raster* de Uso e Ocupação do Solo para vetor;
- Para os dados de pedologia, foram recortadas as áreas de cada imóvel rural e contabilizadas as áreas com cada tipo de solo;
- Os *shapefiles* de rodovias, ferrovias e hidrovias foram avaliados quanto a integridade dos dados e filtradas quanto ao estágio do empreendimento, uma vez que se encontram no arquivo representações de projetos e obras em andamento. Para isso foi utilizado o atributo “*loc*” disponível no *Python* 3.

2.5 Análise espacial

A etapa de análise espacial consistiu em definir variáveis capazes de quantificar a relação entre os limites de propriedades e os demais critérios. Esse procedimento se fez necessário, pois, os critérios levantados e presentes na norma são independentes entre si, cabendo ao avaliador relacioná-los com a propriedade no ato de decisão durante a avaliação do imóvel rural.

Na teoria da decisão, um critério é utilizado como base para medir e avaliar determinada decisão e pode assumir características restritivas ou de fator (AQUINO, 2018). Para as variáveis determinadas no estudo, foram consideradas as seguintes classificações: de fator, quando

puder ser atribuída como porcentagem da área do imóvel rural; ou como restrição, quando ausente no imóvel rural em estudo.

2.5.1 Restrições – Lógica booleana

O processo de definição das restrições foi realizado de acordo com o critério a ser avaliado, seguindo a metodologia especificada em cada item a seguir:

Malha Ferroviária

O *shapefile* de ferrovias possui em seus atributos dados qualitativos e quantitativos para cada ferrovia do estado. As ferrovias podem ser analisadas em razão do atributo situação, o qual pode representar quatro formas: em obras, em operação, planejada e em estudo. Para este trabalho, foram selecionadas somente ferrovias em operação, uma vez que não é intuito do trabalho gerar dados especulativos de preço futuro dos imóveis rurais.

A correlação entre as ferrovias e imóveis rurais foi analisada pela aplicação da ferramenta “*overlay*” pertencente à biblioteca Geopandas. Dessa forma, as propriedades foram classificadas utilizando Lógica Booleana, recendo o valor 1 em caso de interseção e 0 caso contrário.

Malha Hidroviária

O *shapefile* de hidrovias possui em seus atributos dados qualitativos e quantitativos para cada hidrovia do estado. Estas foram analisadas em razão da “classificação”, isto é, o quanto cada hidrovia é utilizada para navegação, em que essa informação consta na tabela de atributos. Assim, foram selecionadas somente hidrovias com uso frequente para navegação. A malha hidroviária resultante foi correlacionada com a base de dados de imóveis a partir da interseção entre as feições, com os imóveis recebendo o valor 1 em caso de interseção e 0 caso contrário.

Disponibilidade Hídrica

A disponibilidade hídrica é um dos principais fatores responsáveis pela variabilidade dos rendimentos da cultura de soja no tempo e no espaço (FARIAS et al., 2001). Nesse cenário, com o estado do Mato Grosso sendo o maior produtor de soja no Brasil, produzindo 26,55%

de toda soja brasileira na safra 2020/21 segundo a Embrapa, a avaliação da disponibilidade hídrica permite instalação de sistemas de irrigação e diversificação com culturas de maior valor agregado, como o arroz.

Assim, foi utilizado o mapa de uso e ocupação do solo para quantificação da disponibilidade hídrica, permitindo mensurar a presença de rios e lagos de forma otimizada. A correlação entre os imóveis e a disponibilidade hídrica foi realizada pela interseção, recendo o valor 1 em caso de interseção e 0 caso contrário.

2.5.2 Fatores – Análise escalonar

Os fatores são critérios que definem o grau de aptidão para a área considerada (AQUINO, 2018). Eles definem áreas em termos de uma medida contínua de aptidão, realçando ou diminuindo a importância de uma alternativa em consideração naqueles locais fora das restrições absolutas (WEBER, 2000). Para isso, o mapa de uso e ocupação do solo vetorizado conforme descrito no item 2.2 possibilitou a obtenção dos critérios e correspondência de variáveis a serem utilizadas no modelo. A seguir são descritos os procedimentos adotados.

Uso e Ocupação do Solo

O uso e ocupação atual do solo é avaliado sob dois aspectos na precificação dos imóveis rurais: a capacidade produtiva e o atendimento da legislação ambiental. Dessa forma, cada propriedade foi analisada quanto ao percentual ocupado com agropecuária, formação natural não florestada e floresta. Tal análise forneceu informação sobre a área produtiva e a conformidade ambiental conforme Tabela 1. Tais cálculos foram realizados com a ferramenta “area” da biblioteca GeoPandas.

Tipos de solo

Para os recortes de cada imóvel rural no mapa de pedologia, foram calculadas as porcentagens de área para os tipos de solo areias quartzosas, aluviais, cambissolo, gleissolo, orgânico, latossolo, planossolo, podzólico, plintossolo, litólico, concrecionado, solodizado e terra roxa. Esses cálculos foram seguidos pelo preenchimento da coluna correspondente com a porcentagem da área total com o respectivo tipo de solo.

Distância de rodovia pavimentada

O *shapefile* de rodovias possui em seus atributos dados qualitativos e quantitativos para cada rodovia do estado. É característico dos imóveis terem acesso via modal rodoviário, assim, as rodovias foram analisadas em razão do atributo cobertura, que representa a pavimentação da via. Outro fator que pode ser analisado é referente a situação da rodovia, logo, foram selecionadas somente rodovias pavimentadas e em funcionamento.

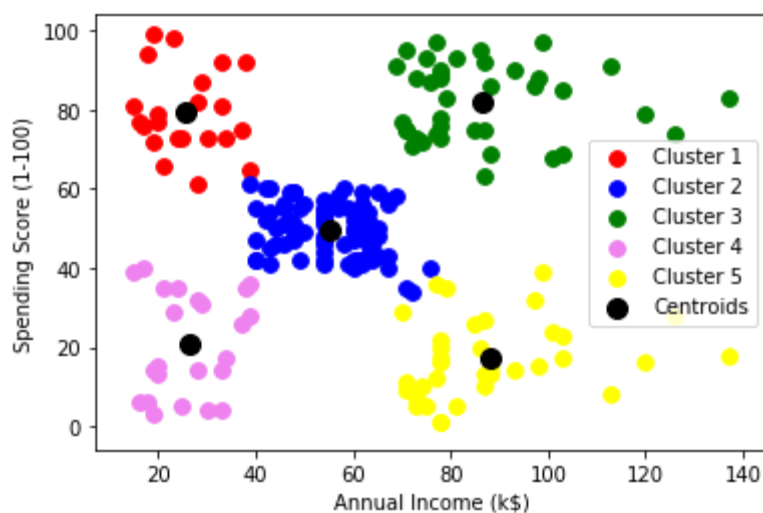
A correlação entre as rodovias e imóveis rurais foi analisada pela aplicação da ferramenta “*buffer*” pertencente à biblioteca Geopandas. Dessa forma, as propriedades receberam *buffers* constantes a cada um quilômetro, até verificação de interseção com alguma rodovia, recebendo assim o valor correspondente à quilometragem que foi verificada interseção.

2.6 Algoritmo

Esta etapa constituiu na aplicação do algoritmo de *Machine Learning K-means*, um dos mais populares algoritmos não supervisionados para *clusterização* (SINAGA; YANG, 2020). O método “*Clusterizar*” serve para encontrar parcelas semelhantes em um conjunto de dados, caracterizado pela maior similaridade (SINAGA; YANG, 2020).

A fim de minimizar o efeito do número de *clusters*, foi utilizado o método *Elbow* para inicialização do modelo. Dessa forma a determinação do número de *clusters* para inicialização acontece pelo cálculo do WCSS (*Within-Cluster Sum of Square*) (SAJI, 2021). O WCSS é a soma da distância ao quadrado entre cada ponto e o centroide de cada *cluster*. O princípio do método é a observância da variação do WCSS conforme há aumento da quantidade de *clusters*, dessa forma, é possível identificar a partir de qual quantidade não há melhora no agrupamento. Para a correta escolha da quantidade de clusters, o método WCSS foi comparado ao método *K-means++*, o qual indica a melhor quantidade de *clusters* para o conjunto de dados.

Os agrupamentos ocorrem por meio da definição de centros e cálculo da distância Euclidiana. Na Figura 1 é apresentado o método de funcionamento.

Figura 2 - Esquema funcionamento *K-Means*

Fonte: SAJI (2021).

O algoritmo recebeu a matriz de dados conforme descrito nos tópicos anteriores, todas com valores distribuídos entre 0 e 1, de forma a não ter ordens de grandeza distintas que prejudiquem os dados obtidos. Com base nisso foi calculado o WCSS para número de clusters entre 1 a 25, precedido da execução e geração da matriz de resultado para cada imóvel contido na base de dados.

2.7 Validação dos clusters

A fim de validar os *clusters* formados foi aplicada metodologia de cálculo do preço de terra nua disponibilizado pelo Instituto de Terras de Mato Grosso, INTERMAT, seguindo o decreto 294 de 2019 e a portaria 08 de 2021 (BRASIL, 2019, 2021). Portanto, de acordo com o agrupamento definido pelo INTERMAT, os municípios foram divididos em 3 grupos, sendo o grupo 1 composto pelos municípios apresentados na Tabela 2, os quais possuem valor base de R\$522,37. O grupo 2, os municípios da Tabela 3 com valor base de R\$373,04. Por fim, o grupo 3, composto pelos municípios da Tabela 4, com valor base de R\$158,00.

Tabela 2. Grupo 1 classificação de Valor Base por município

GRUPO 1 - Valor base R\$522,37/ha	
Município	Município
Água Boa	Nova Guarita
Alta Floresta	Nova Maringá
Alto Garças	Nova Monte Verde
Alto Taquari	Nova Mutum
Araguaiana	Nova Olímpia
Araputanga	Nova Olímpia
Barra do Bugres	Nova Xavantina
Barra do Garças	Novo Horizonte do Norte
Brasnorte	Novo São Joaquim
Campinápolis	Paranaita
Campo Novo do Parecis	Pedra Preta
Campo Verde	Porto dos Gaúchos
Campos de Júlio	Primavera do Leste
Canarana	Querencia
Carlinda	Reserva do Cabaçal
Chapada dos Guimarães	Rio Branco
Cláudia	Rondonópolis
Colider	Salto do Céu
Comodoro	Santa Carmem
Denise	Santa Rita do Trivelato
Diamantino	Santo Antonio do Leste
Dom Aquino	São Jose do povo
Feliz Natal	São Jose do Rio Claro
Gaúcha do Norte	São Jose dos Quatros Marcos
General Carneiro	São Pedro da Cipa
Glória d' Oeste	Sapezal
Guiratinga	Sinop
Ipiranga do Norte	Sorriso
Itanhangá	Tabaporã
Itaúba	Tangara da Serra
Itiquira	Tapurah
Jaciara	Terra Nova do Norte
Juara	Vera
Jucimeira	
Lambari d' Oeste	
Lucas do Rio Verde	
Matupá	
Mirassol d' Oeste	
Nova Canaã do Norte	

Fonte: Instituto de Terras do Mato Grosso, INTERMAT (2021).

Tabela 3. Grupo 2 classificação de Valor Base por município

GRUPO 2 - Valor base R\$373,04/ha
Município
Alto Araguaia
Alto Boa Vista
Bom Jesus do Araguaia
Canabrava do Norte
Confresa
Conquista d'Oeste
Figueirópolis d' Oeste
Guarantã do Norte
Indiavaí
Jauru
Luciara
Marcelandia
Nova Brasilandia
Nova Santa Helena
Nova Lacerda
Novo Mundo
Novo Santo Antônio
Peixoto de Azevedo
Planalto Da Serra
Pontes e Lacerda
Porto Alegre do Norte
Porto Esperidião
Ribeirão Cascalheira
Santa Cruz do Xingu
Santa Cruz do Xingu
São Felix do Araguaia
São José do Xingu
Serra Nova Dourada
União do Sul
Vale de São Domingos
Vila Bela Ss. Trindade
Vila Rica

Fonte: Instituto de Terras do Mato Grosso, INTERMAT (2021).

Tabela 4. Grupo 3 classificação de Valor Base por município

GRUPO 3 - Valor base
R\$158,00/ha
Município
Alto Paraguai
Apiacás
Araguaína
Arenápolis
Aripuanã
Barão de Melgaço
Cáceres
Castanheira
Cocalinho
Colniza
Cotriguaçu
Cuiabá
Curvelândia
Jangada
Juína
Juruena
N. Senhora do Livramento
Nobres
Nortelândia
Nortelândia
Nova Marilandia
Nova Nazaré
Paranatinga
Poconé
Pontal do Araguaia
Ponte Branca
Porto Estrela
Poxoreo
Ribeirãozinho
Rondolândia
Rosário Oeste
Santo Afonso
Santo Antonio de Leverger
Tesouro
Torixoréu
Varzea Grande

Fonte: Instituto de Terras do Mato Grosso, INTERMAT (2021).

Os valores base apresentados nas Tabelas 2,3 e 4 correspondem ao mínimo valor pago por hectare, proporcional ao valor pago pelas áreas de preservação permanente. Para as demais áreas, são acrescidos fatores sobre o valor base, em função de aspectos referentes a solo,

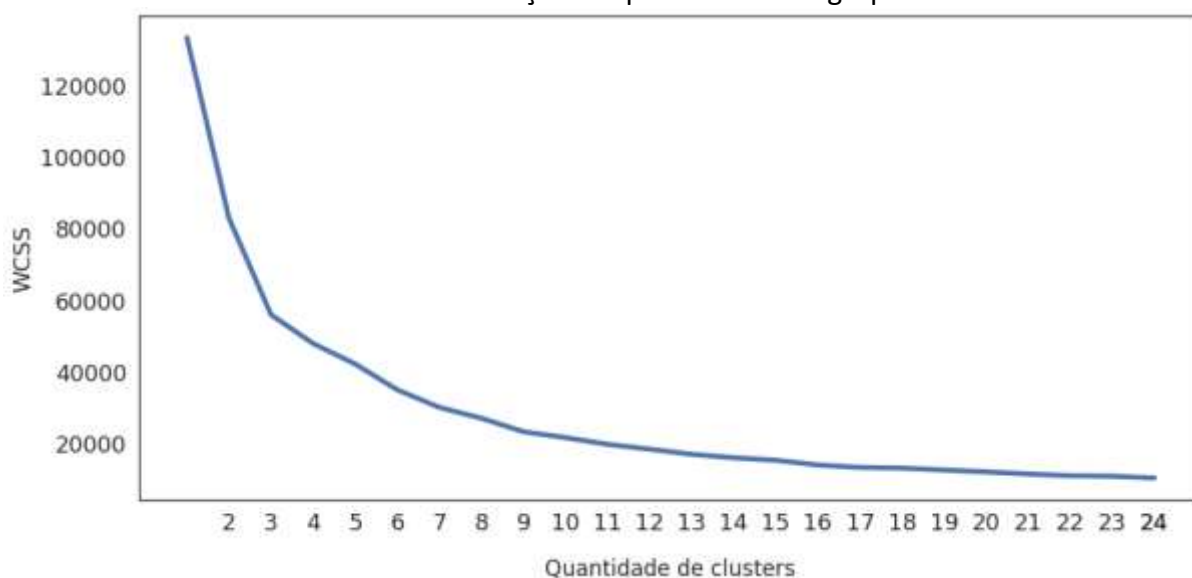
vegetação, bioma e acesso. Para solo e vegetação, o valor acrescido é obtido pela média ponderada das frações de área, enquanto o acesso é pelo simples enquadramento.

Para o tipo de solo, é aplicado fator 1 nas áreas de latossolo ou argissolo; 0,70 em áreas de vertissolo ou cambissolo; e 0,35 para neossolo, plintossolo, gleissolo ou planossolo. Para a vegetação, é aplicado fator 0,50 se o bioma for Cerrado; ou 0,35 para caso seja Pantanal ou Amazônico. Em relação ao acesso, se a via for precária e até 50 km, é aplicado fator 0,20; se for até 30 km e cascalhada é empregue o fator de 0,30; ou 0,50 para até 30 km de via asfaltada.

3 RESULTADOS E DISCUSSÃO

Os mais de 165.000 imóveis rurais únicos contidos no Cadastro Ambiental Rural (CAR) constituíram a base de dados gerada de acordo os aspectos homogeneidade, situação, disponibilidade de recursos e logística. Esta foi submetida no algoritmo *K-Means*, com a quantidade de centros sendo definida pelo WCSS, conforme apresentado no Gráfico 1.

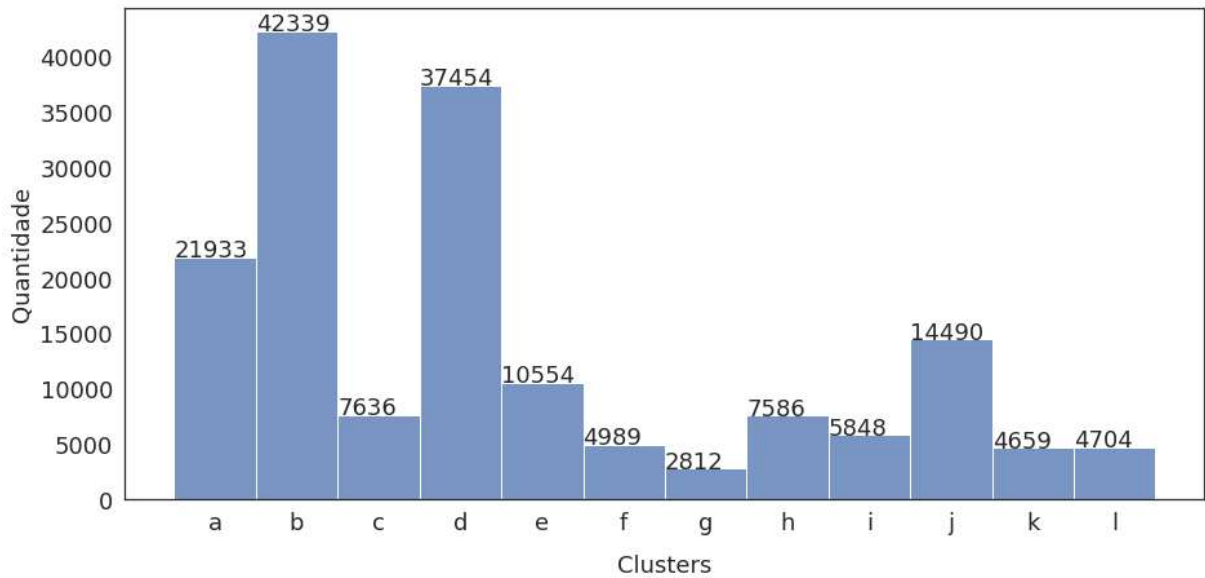
Gráfico 1. WCSS em função da quantidade de agrupamentos



Fonte: O autor.

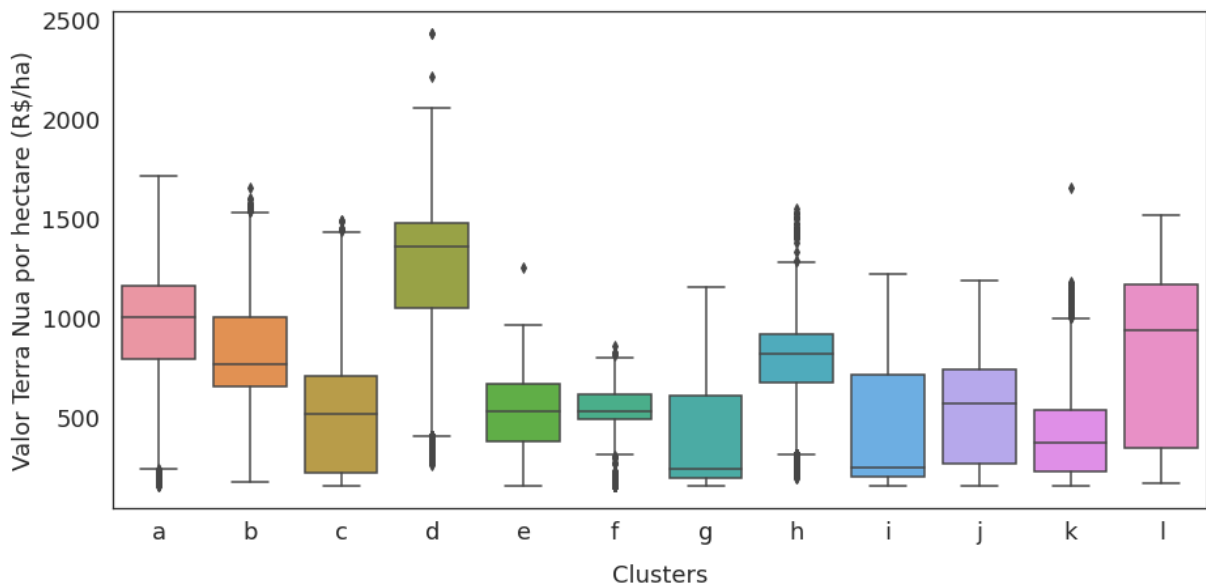
Observa-se pelo Gráfico 1 que a soma da distância entre pontos e centroides diminui à medida que se aumenta o número de clusters, apresentando valores menos expressivos a partir de doze clusters. Dessa forma, foi considerado o número de *clusters* igual a doze para inicialização do algoritmo, uma vez que o método *K-means++* também apresentou melhores resultados para esta quantidade.

Os clusters gerados foram definidos de “a” a “l”, com a quantidade de imóveis por cluster apresentados no Gráfico 2. Em aspectos quantitativos, tem-se o menor cluster com quantidade satisfatória de propriedades, uma vez que a caracterização de, no mínimo 2812 propriedades, ocasiona economia de tempo e mão de obra técnica para levantamento de características em campo de cada propriedade.

Gráfico 2. Histograma de quantidade de imóveis por *cluster*

Fonte: O autor.

A etapa de validação dos clusters foi aplicada e os resultados obtidos são apresentados no Gráfico 3, o qual relaciona cada conjunto de imóveis com o valor de mercado praticado seguindo a metodologia apresentada no tópico 3.

Gráfico 3. Validação dos clusters em relação aos valores de terra nua por hectare

Fonte: O autor.

A partir do gráfico acima se observa que o algoritmo foi bem sucedido quanto aos clusters formados, uma vez que estes apresentaram características distintas quanto a

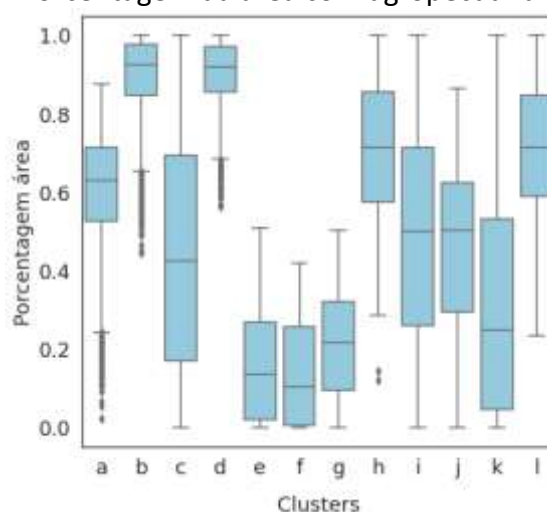
mediana e seus percentis. Além disso, observa-se o pequeno número de *outliers*, se comparado com o tamanho da base de dados; além de caudas pequenas, exceto para os grupos com maior quantidade de propriedades.

É necessário salientar que o algoritmo *K-Means* tem por objetivo gerar clusters quanto as variáveis de entrada, ou seja, é esperado que haja grupos formados somente por característica positivas, outros neutros (os quais possuem algumas características positivas e outras mínimas) e ainda grupos com só características mínimas. Assim, o preço atingido pode ser semelhante entre grupos distintos, uma vez que nos *clusters* com valor de terra nua intermediários ou inferiores os aspectos positivos (que agregam valor ao imóvel) ocorrem somente para algumas variáveis, ao mesmo passo que ocorre desvalorização para outras, gerando assim grupos com características distintas, mas com preços similares.

Para a metodologia de validação adotada, o preço mínimo é de R\$158,00, valor este referente as áreas de preservação permanente em algumas regiões do Mato Grosso, indicando assim que os clusters “g”, “h” e “k” possuem grande representatividade quanto a propriedades com grandes áreas de preservação. Em relação aos valores máximos, o cluster “d” abrangeu a maioria das propriedades com preços acima de R\$1.000,00.

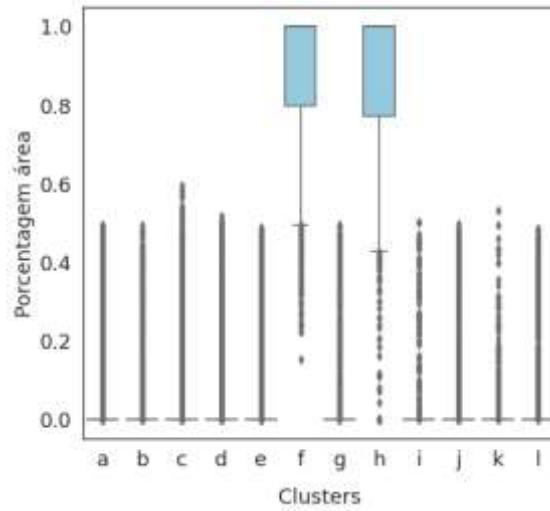
Entre as variáveis do conjunto de dados, as com maior representatividade nos grupos formados foram as áreas de agropecuária; solo areia quartzosa, cambissolo, latossolo, podzólico, plintossolo e concrecionado. Ou seja, quando comparado os clusters à variação se destacam por representarem intervalos distintos por grupo (Gráfico 4 ao 10).

Gráfico 4. Porcentagem da área com agropecuária por *cluster*



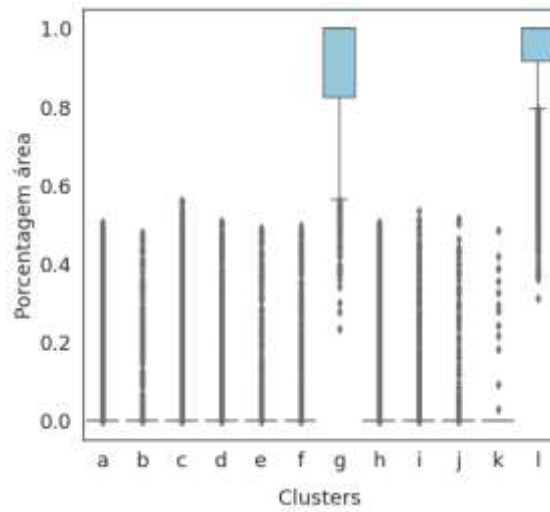
Fonte: O autor.

Gráfico 5. Porcentagem da área com solo areia quartzosa por *cluster*



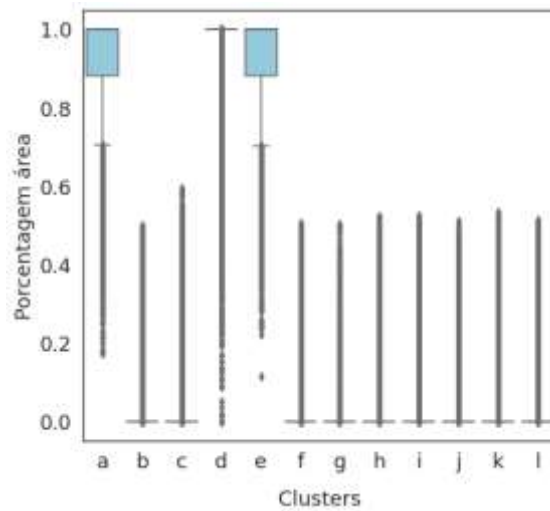
Fonte: O autor.

Gráfico 6. Porcentagem da área com solo cambissolo por *cluster*



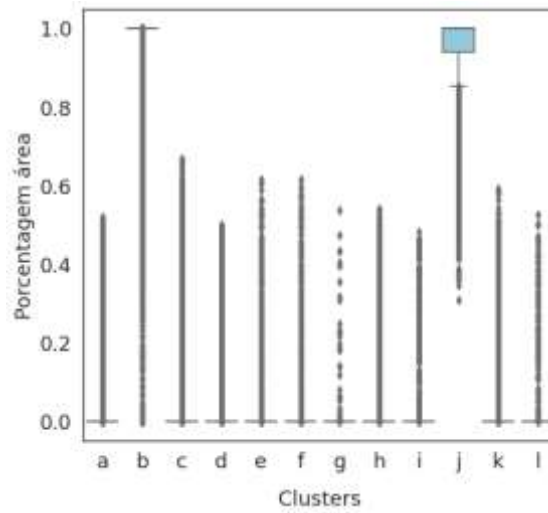
Fonte: O autor.

Gráfico 7. Porcentagem da área com solo latossolo por *cluster*



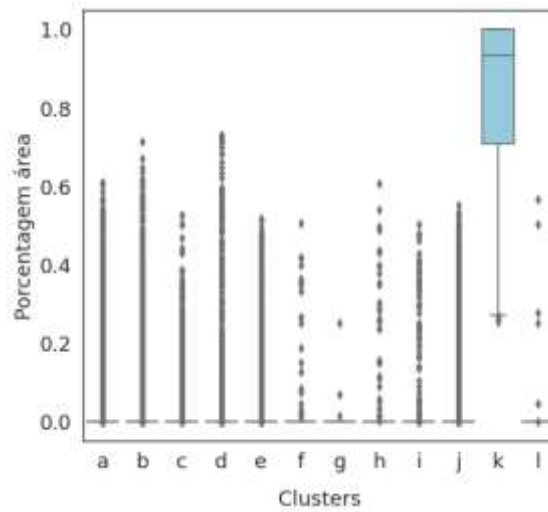
Fonte: O autor.

Gráfico 8. Porcentagem da área com solo podzólico por *cluster*



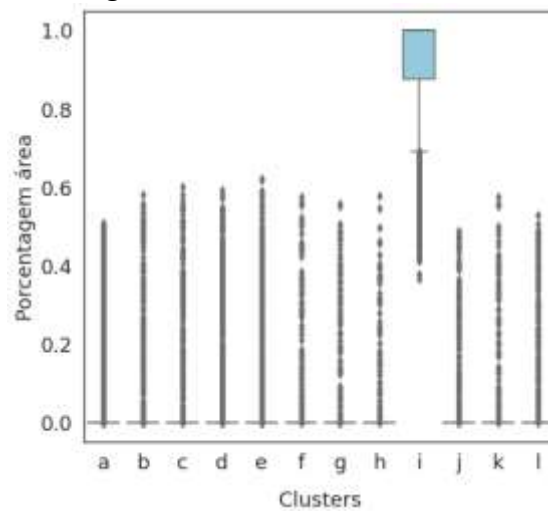
Fonte: O autor.

Gráfico 9. Porcentagem da área com solo plintossolo por *cluster*



Fonte: O autor.

Gráfico 10. Porcentagem da área com solo concrecionado por *cluster*

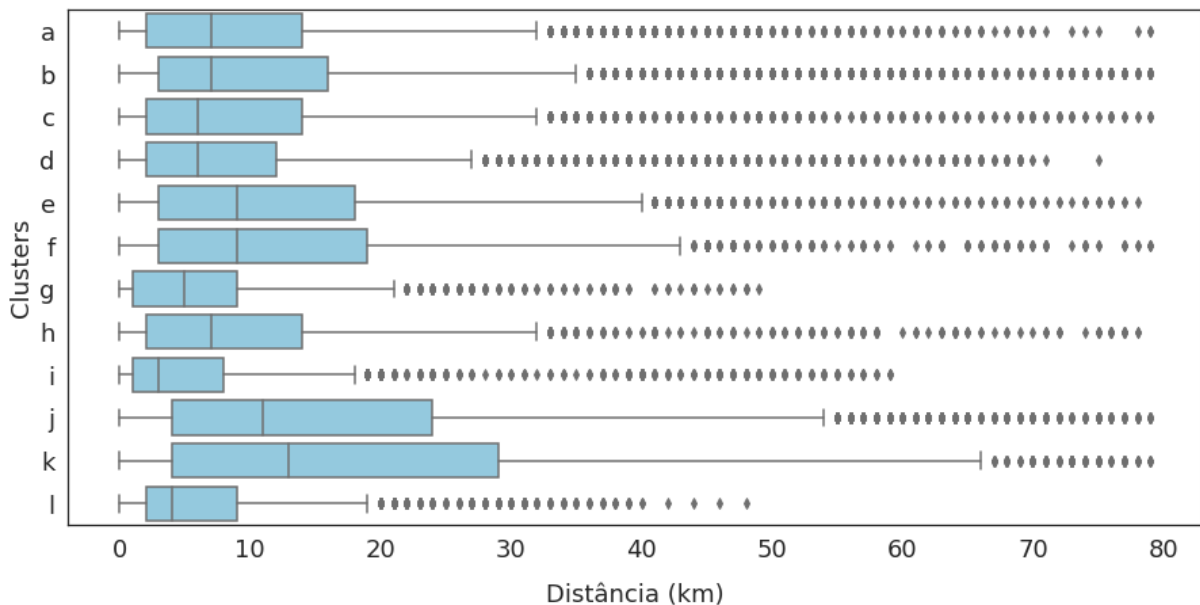


Fonte: O autor.

Para estas variáveis, enquanto os grupos com predomínio apresentam concentração dos dados para porcentagens de áreas superiores a 50%, os demais grupos mantem seus percentis próximos ao zero, com exceção dos *outliers*. Com relação à porcentagem da área dos imóveis com agropecuária, os clusters “a”, “b”, “d”, “h” e “l” são os que contém 50% ou mais das propriedades não respeitando a Lei n 12.651, que define a área mínima de Reserva Legal.

Apesar de algumas variáveis estarem distribuídas em proporções parecidas entre os clusters formados, tal comportamento era esperado, uma vez que a aplicação do algoritmo *K-Means* busca solucionar situações de agrupamento nos quais não se tem os grupos totalmente definidos pelas variáveis de entrada. Em relação as variáveis de porcentagem de área com os tipos de solo areia quartzosa, cambissolo, latossolo, podzólico, plintossolo e concrecionado, os clusters gerados possuem tipos de solo predominante nas propriedades. Com relação a distância dos imóveis às rodovias pavimentadas, os cluster obtidos não concentraram distâncias próximas, conforme apresentado no Gráfico 11.

Gráfico 11. Distância dos imóveis até rodovias pavimentadas por *cluster*

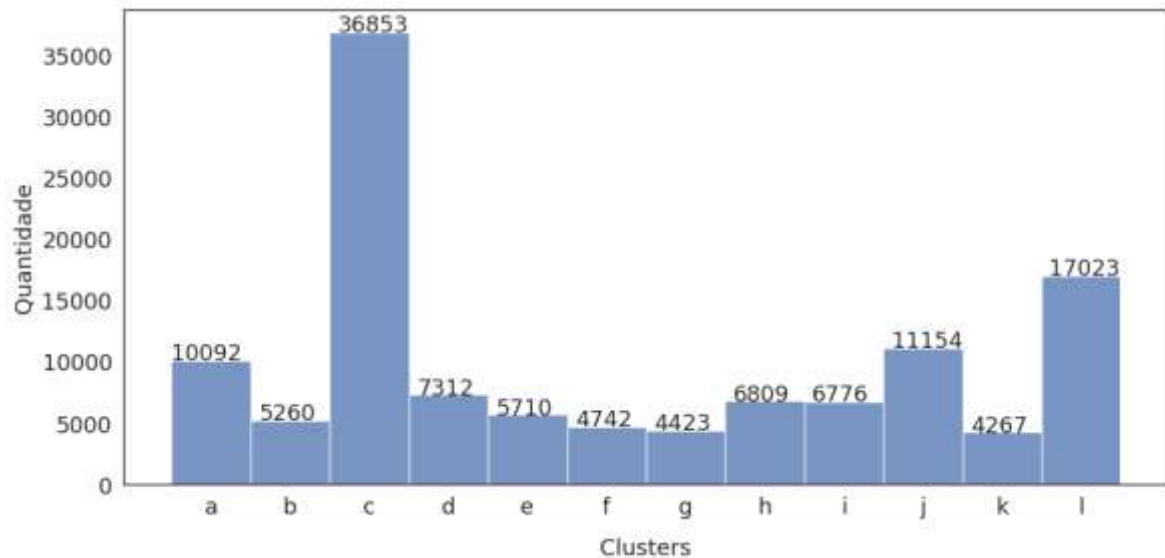


Fonte: O autor.

A distância não foi determinante no agrupamento, sendo observado que para todos os grupos tiveram propriedades com rodovias pavimentadas em sua área, caracterizadas no Gráfico 11 pela distância de zero quilômetros. No entanto, ao aplicar o agrupamento novamente, agora sem os valores de distância, foi constatado alteração nos grupos formados,

significando assim, que esta variável, quando combinada com as demais, tem relevância para o agrupamento (Gráfico 12).

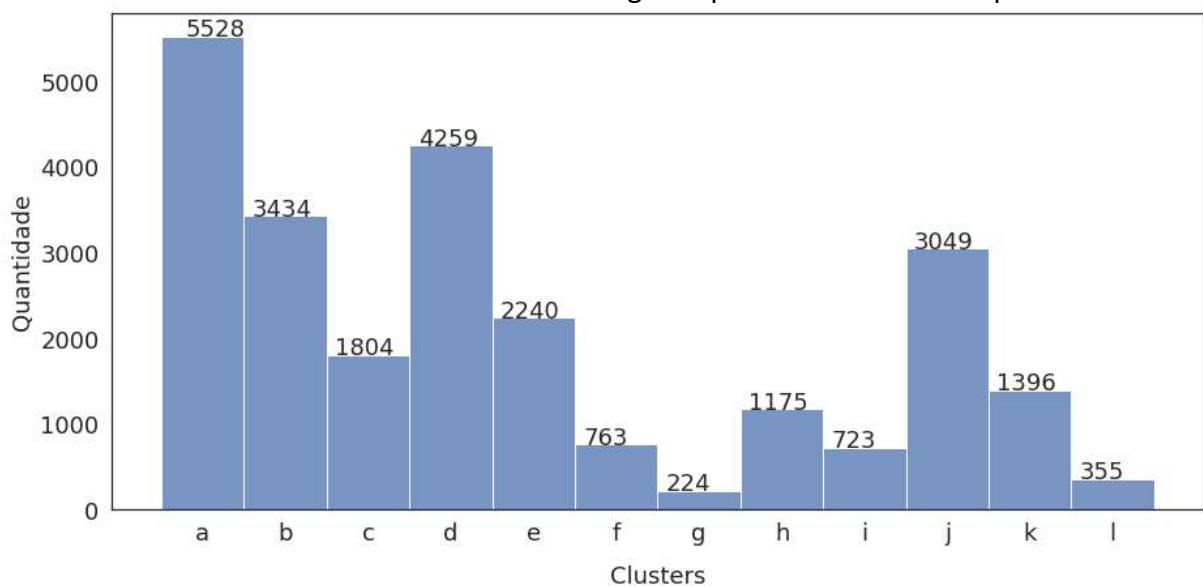
Gráfico 12. Quantidade de imóveis por cluster se desconsiderada a distância de rodovias



Fonte: O autor.

O mesmo ocorreu em relação à presença de recursos hídricos no imóvel, formou-se grupos de maior representatividade, conforme o Gráfico 13. No entanto, essa variável sozinha não é determinante para formação dos grupos, mas gera alterações significativas quando retirada do conjunto de dados submetido ao algoritmo.

Gráfico 13. Quantidade de imóveis com algum tipo de recurso hídrico por *cluster*



Fonte: O autor.

Para as demais variáveis de entrada: ferrovia; hidrovia; solos (aluviais, gleissolos, orgânicos, planossolos, litólicos, solodizado e terra roxa) foi observada pequena variação nos clusters com a retirada de até duas destas variáveis. Isso ocorre devido a pequena quantidade de imóveis com tais características em relação ao total da base. No entanto, tais informações são importantes para a base de dados visando ter caracterização detalhada dos imóveis. Na Tabela 5 são apresentados o total de imóveis para tais variáveis menos representativas.

Tabela 5. Quantidade de imóveis por variável analisada

Tipo	Variável	Quantidade de imóveis
Logística	Ferrovia	108
	Hidrovia	28
Solos	Aluviais	1377
	Gleissolos	9765
	Orgânicos	1618
	Planossolos	1429
	Litólicos	11805
	Solodizados	138
	Terra Roxa	485

Fonte: O autor.

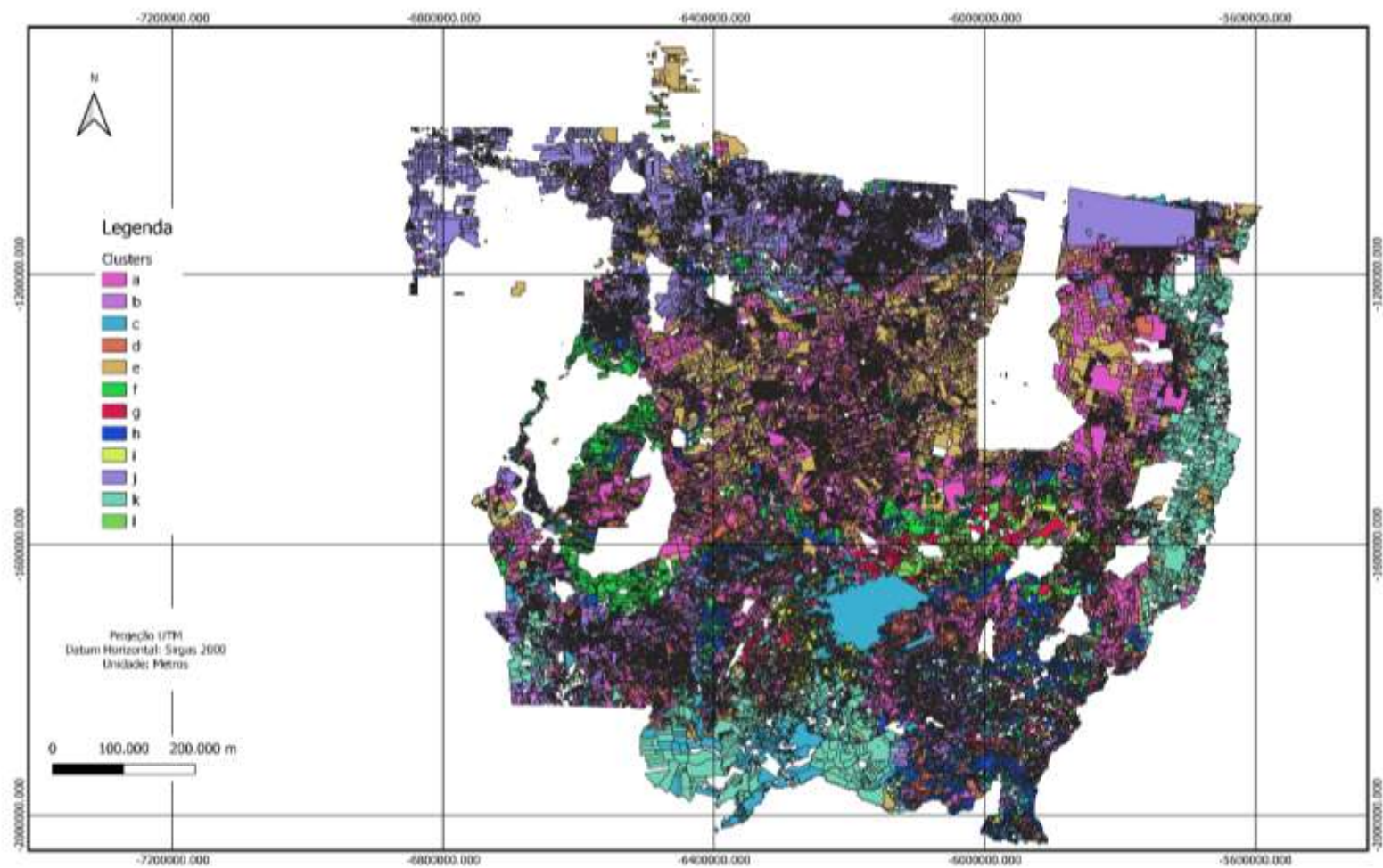
A Figura 3 representa a espacialização dos *clusters* gerados para cada propriedade presente no Cadastro Ambiental Rural do Mato Grosso. As Figuras 4 a 15 apresentam, separadamente, os doze *clusters* selecionados nessa pesquisa, em que é observada que os imóveis podem ser semelhantes mesmo estando em regiões distintas, possibilitando aplicação de valores base semelhantes sem a limitação geográfica.

A partir das Figuras 3 a 15 se observa que os *clusters* formados ultrapassam os limites regionais, apresentando grupos com características semelhantes em diferentes regiões do Estado do Mato Grosso. Tal característica possibilita que regiões de menor especulação determinem seus valores base de acordo com regiões de mercado aquecido, além de possibilitar a adoção de metodologia comum a todo estado.

O *cluster* “a” ficou amplamente distribuído pelas regiões do estado, sendo caracterizado por propriedades entre 50 a 70% de área de agropecuária e com quase sua totalidade de latossolos. O “b” teve predomínio nas regiões Norte e Sul, com imóveis evidenciados por quase sua totalidade de área de agropecuária e solo podzólico. O “c” concentrou-se na região Sul, mas com propriedades especializadas por todo o estado. O “d” abrangeu propriedades de

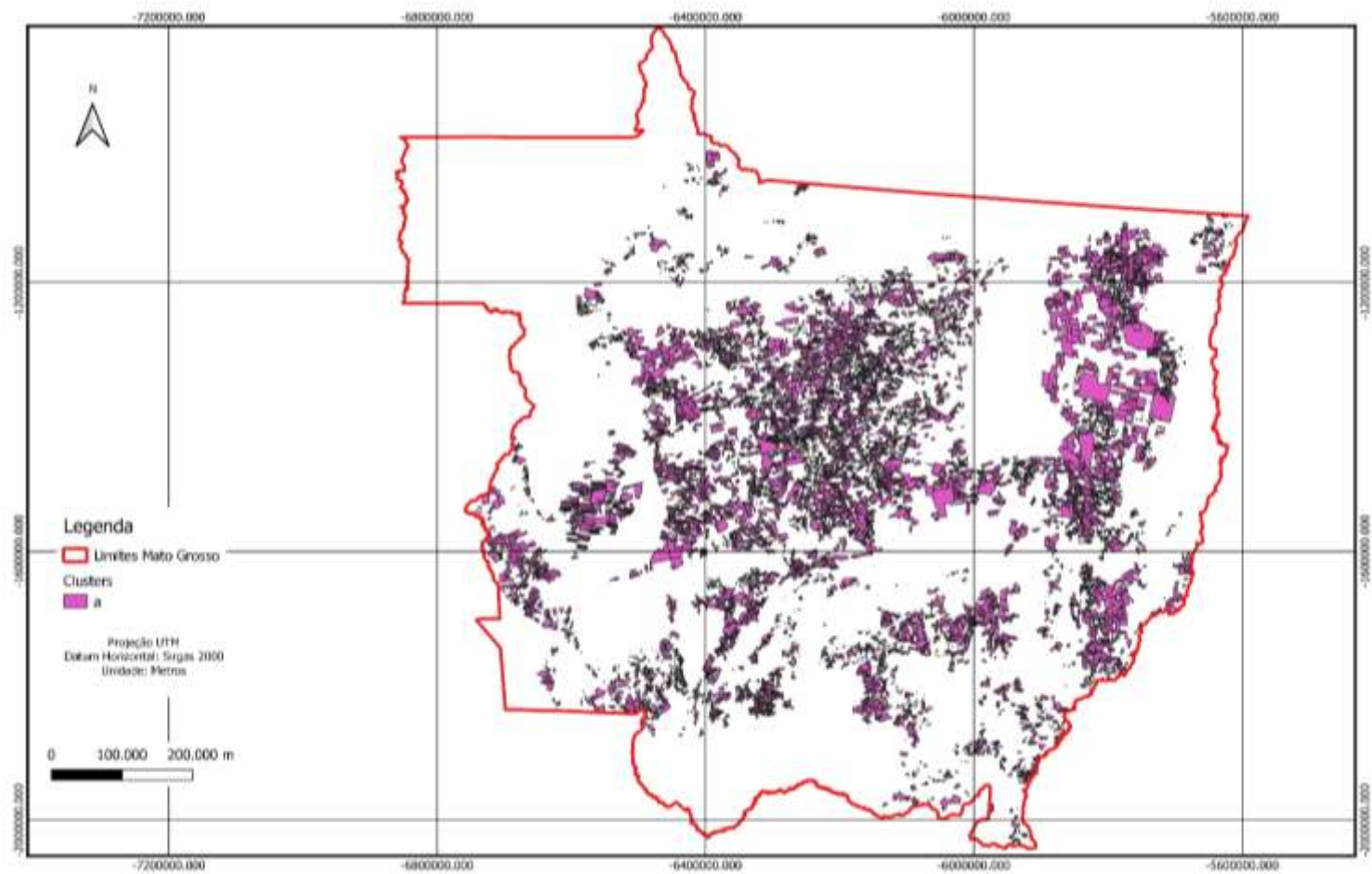
todo o estado, estas com quase a totalidade de área de agropecuária e latossolo. O “e” abrangeu propriedades com até 30% de área de agropecuária, domínio de latossolo e especializada predominantemente na região central do Mato Grosso. O “f” imóveis na faixa central do estado, caracterizados por solo de areia quartzosas. O “h” com maior concentração de imóveis na região Sudeste e com solo predominantemente de areia quartzosas. O “i” teve domínio de solos concrecionados concentrados em pequenas áreas na região Sul e Leste. Para os *clusters* “g” e “L”, destaca-se sua espacialização semelhante, nestes grupos os imóveis se diferenciam pela área de agropecuária e, conseqüentemente, área de preservação permanente; enquanto o “g” representa propriedades com até 50% de área de agropecuária, o “L” tem suas propriedades com mais de 50% de agropecuária, conforme apresentado no Gráfico 4.

Figura 3. Espacialização *clusters* dos imóveis rurais do Mato Grosso



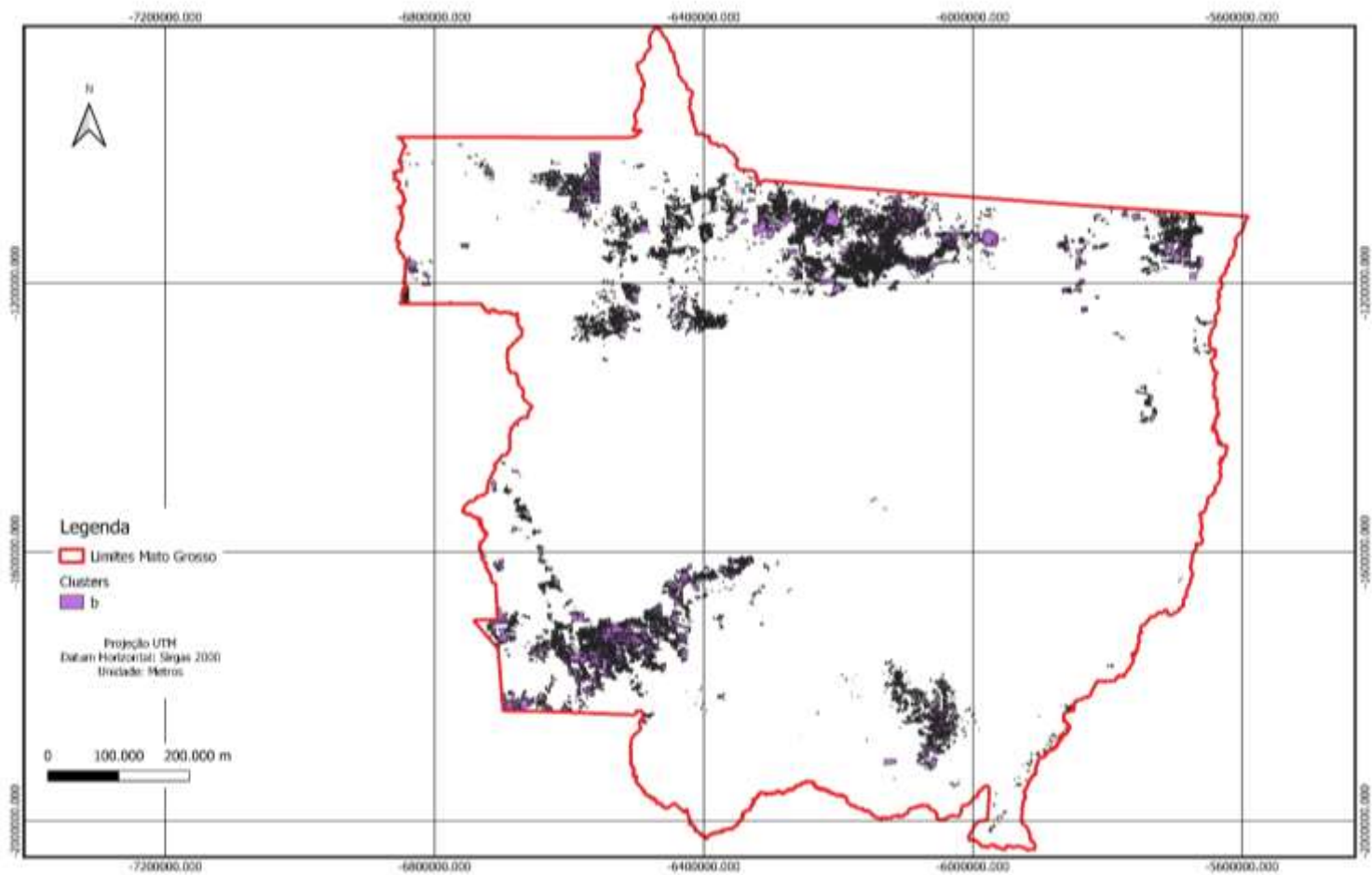
Fonte: O autor.

Figura 4. Espacialização *cluster* A dos imóveis rurais do Mato Grosso



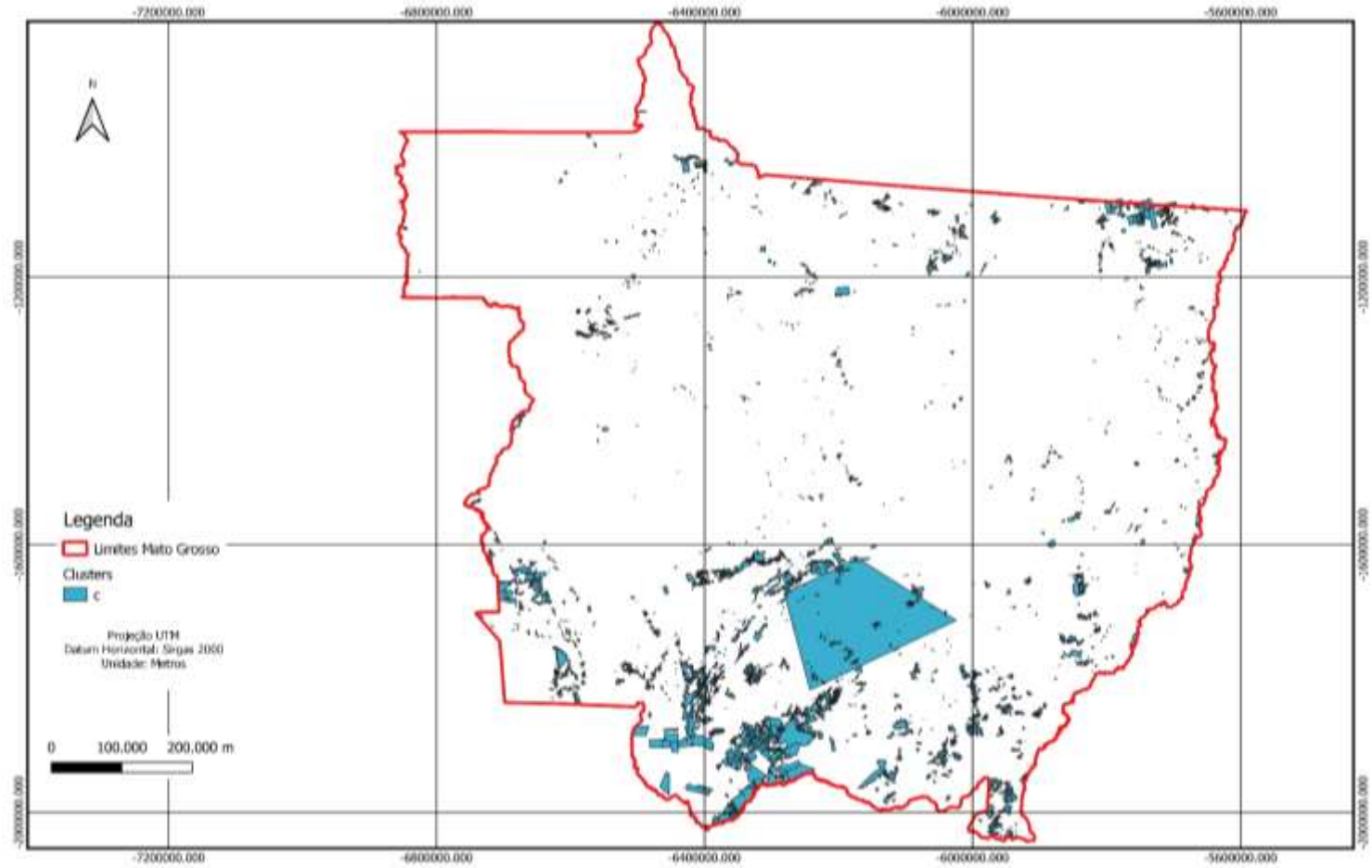
Fonte: O autor.

Figura 5. Espacialização *cluster* B dos imóveis rurais do Mato Grosso



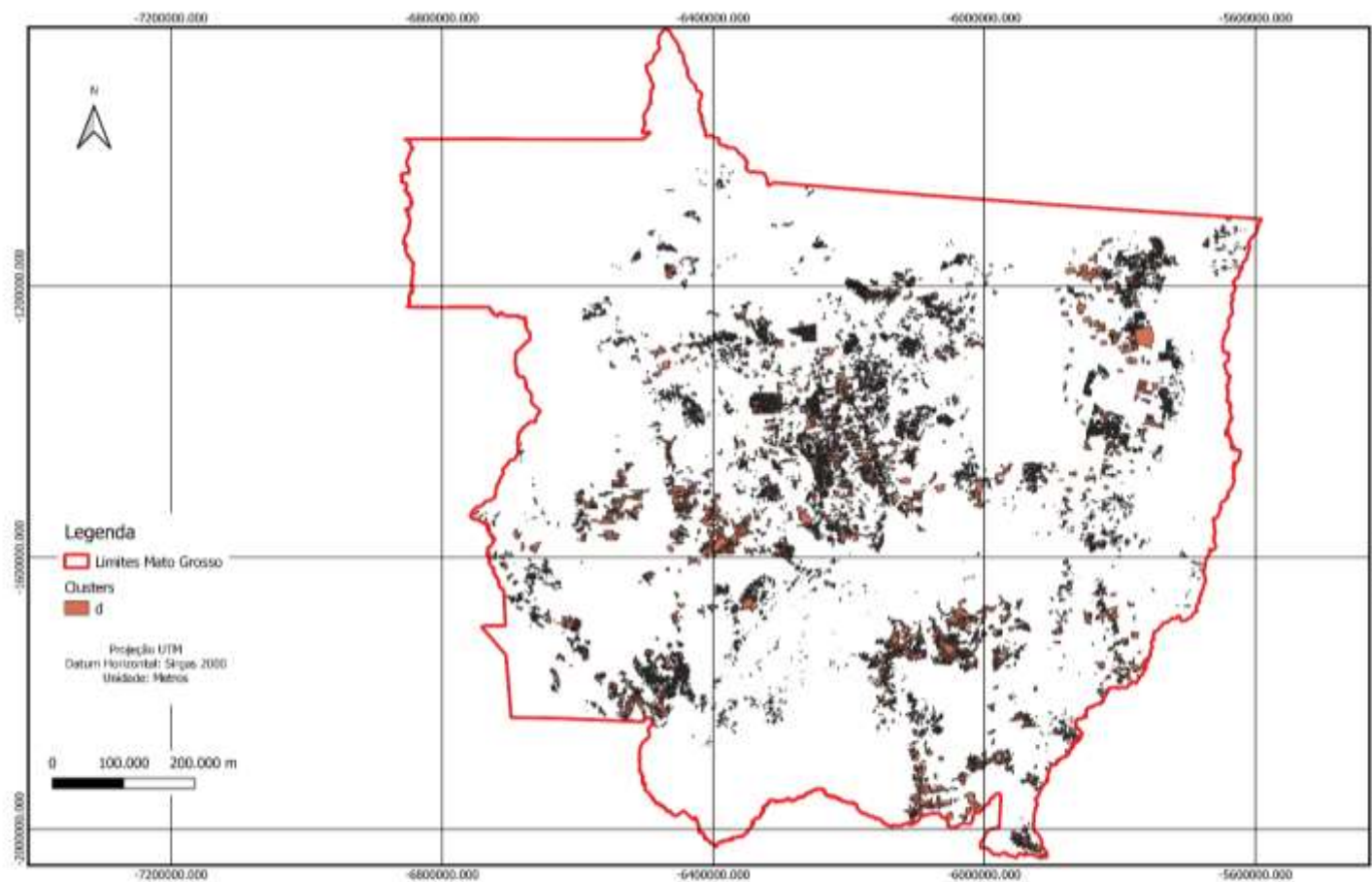
Fonte: O autor.

Figura 6. Espacialização *cluster* C dos imóveis rurais do Mato Grosso



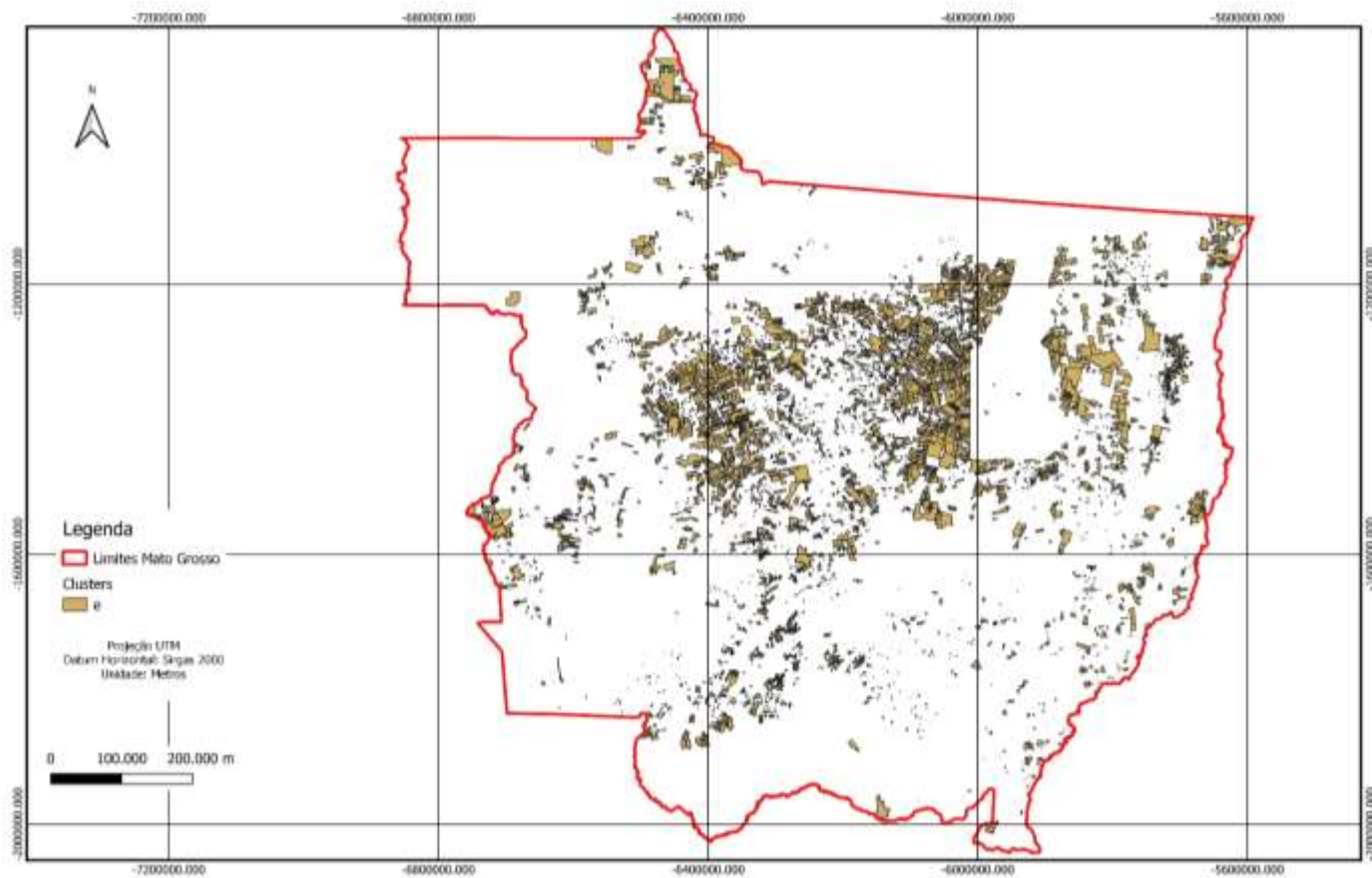
Fonte: O autor.

Figura 7. Espacialização *cluster* D dos imóveis rurais do Mato Grosso



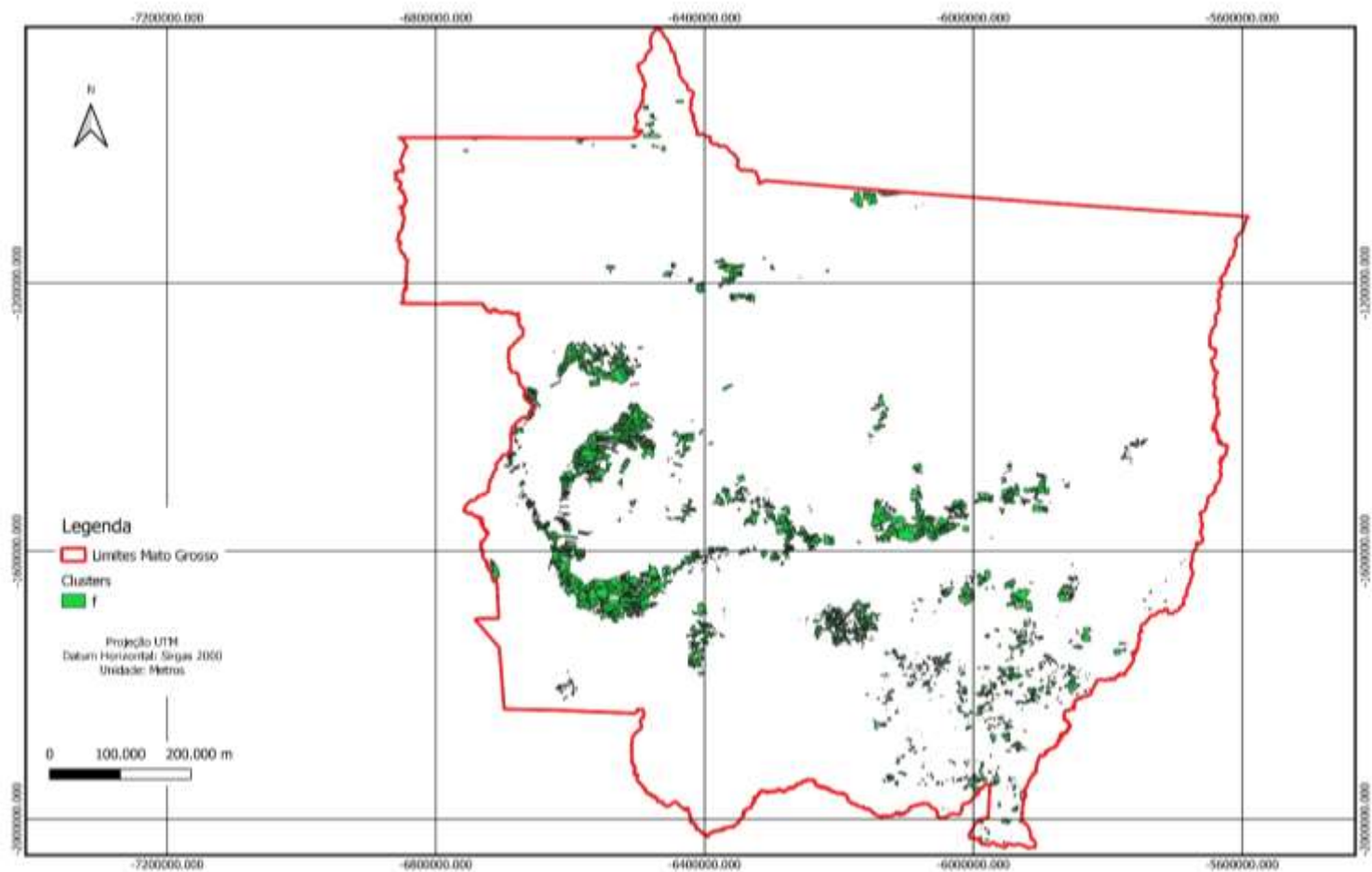
Fonte: O autor.

Figura 8. Espacialização *cluster* E dos imóveis rurais do Mato Grosso.



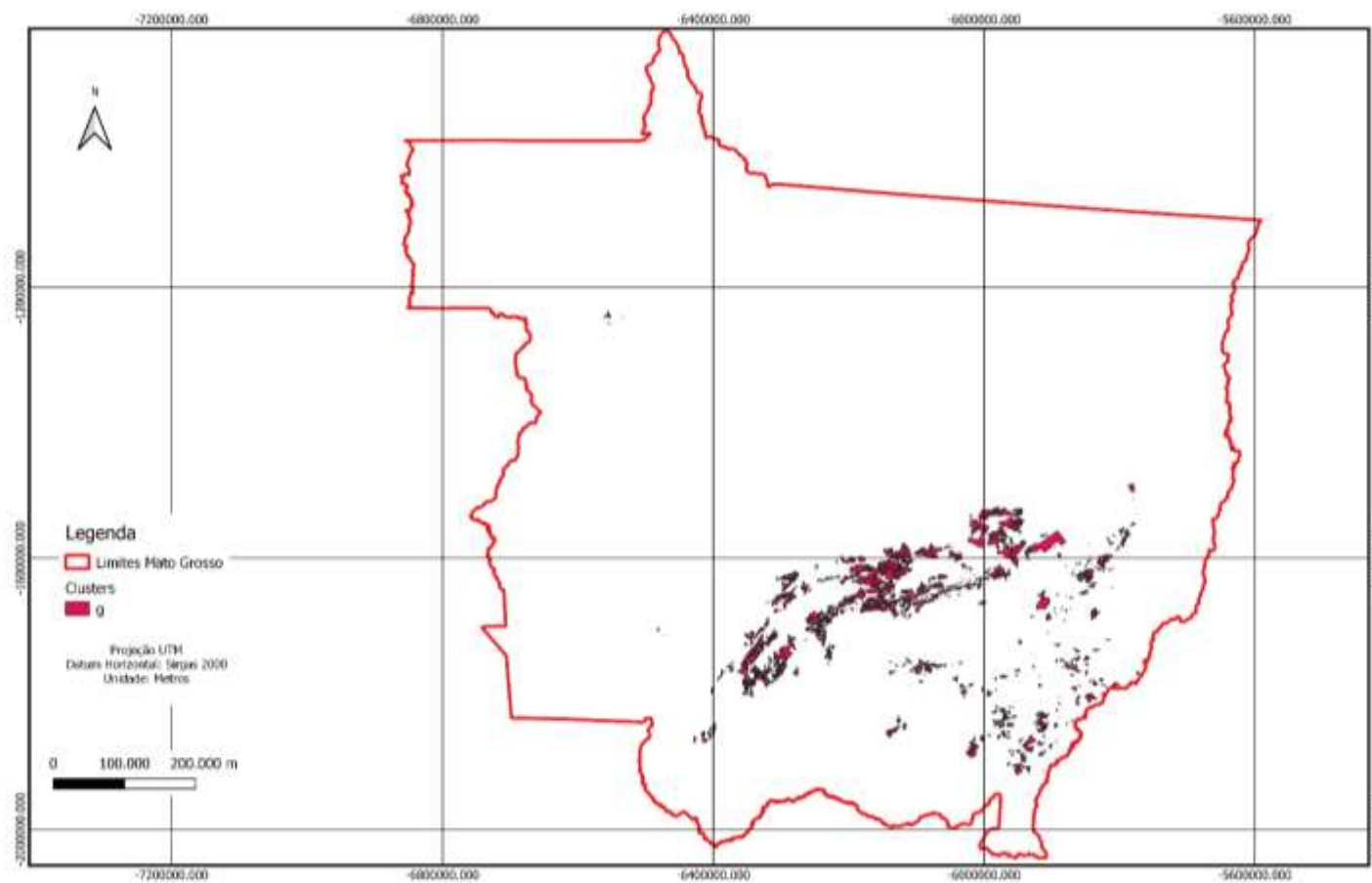
Fonte: O autor.

Figura 9. Espacialização *cluster* F dos imóveis rurais do Mato Grosso



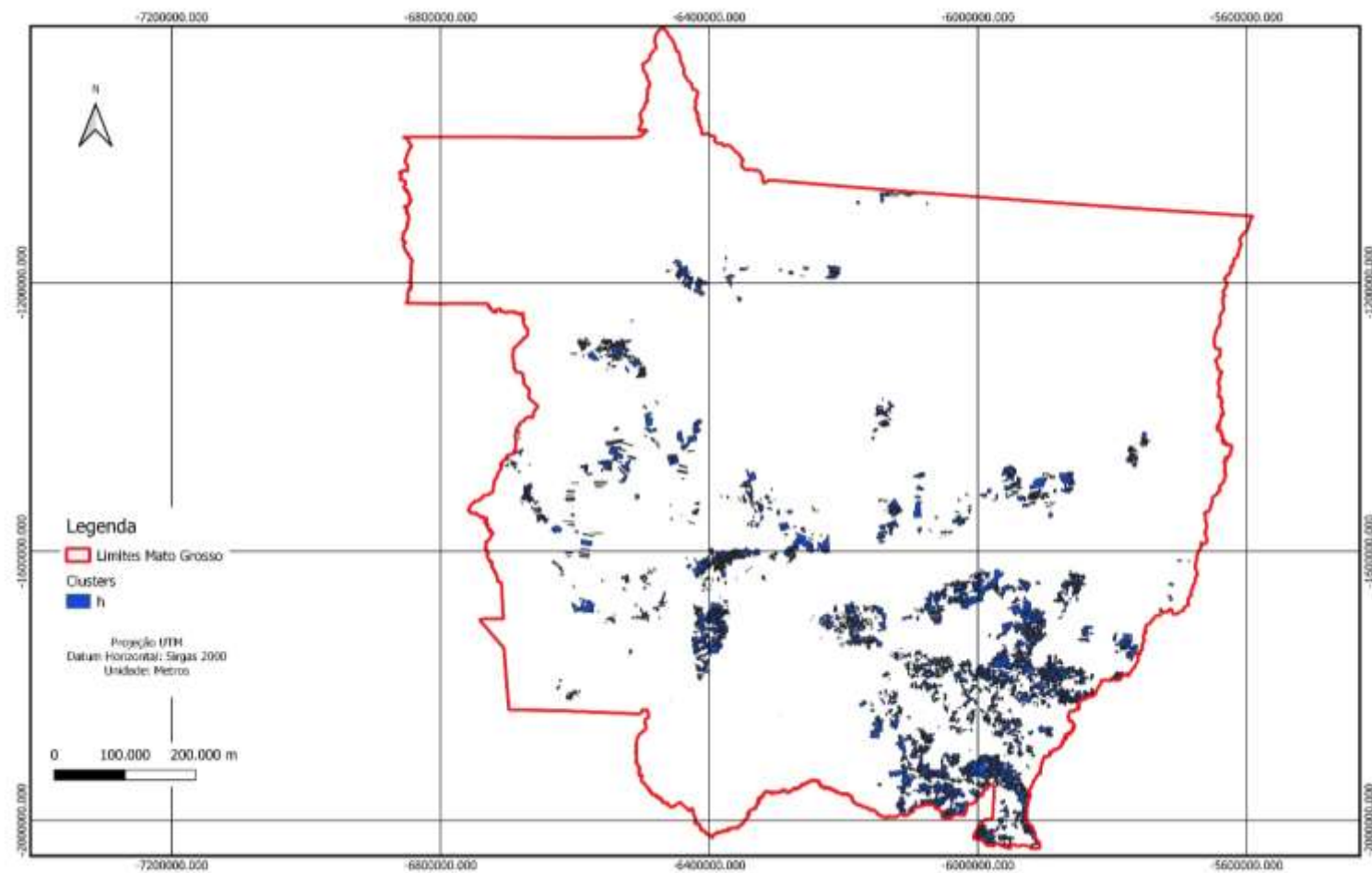
Fonte: O autor.

Figura 10. Espacialização *cluster* G dos imóveis rurais do Mato Grosso



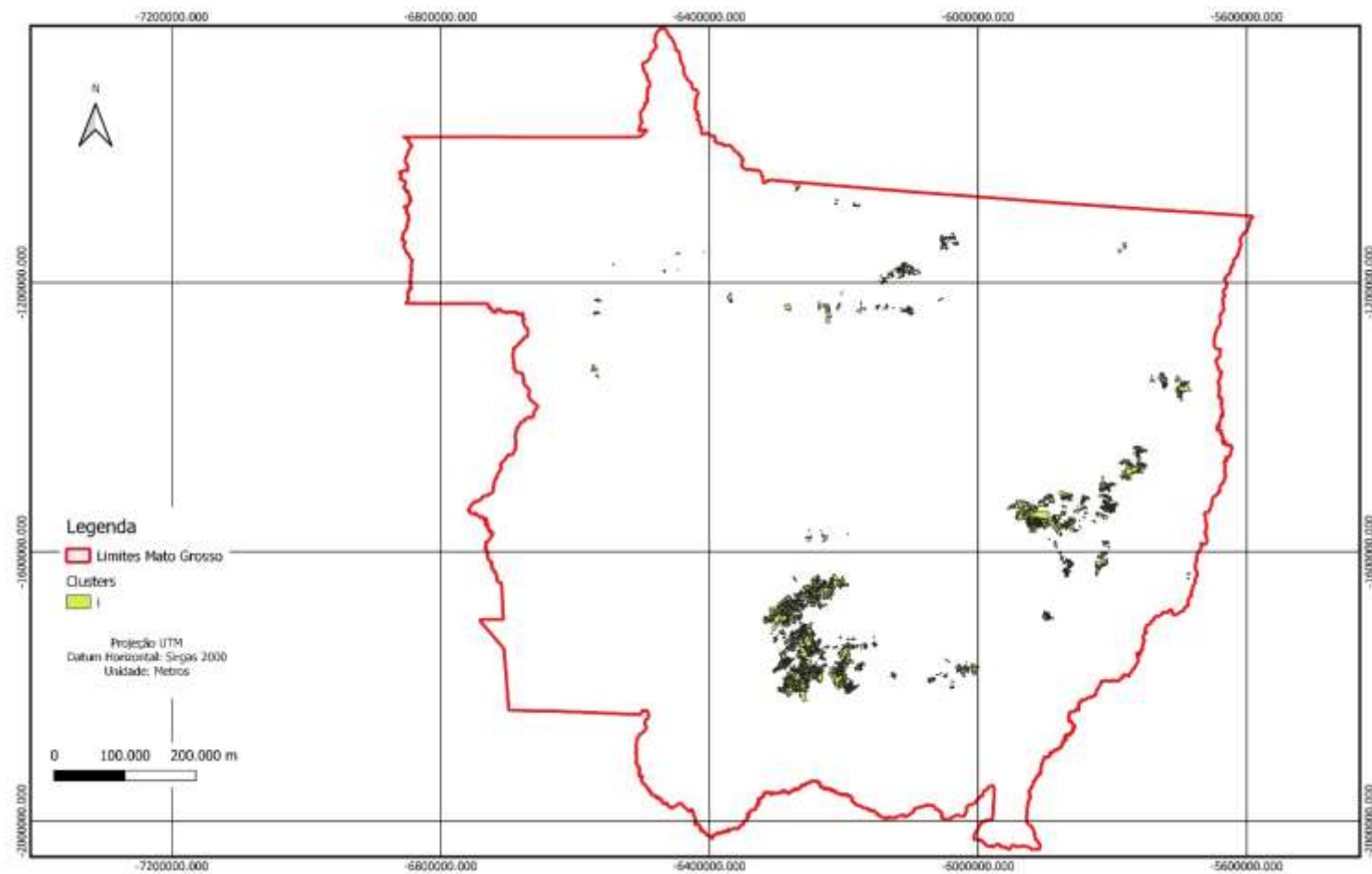
Fonte: O autor.

Figura 11. Espacialização *cluster* H dos imóveis rurais do Mato Grosso



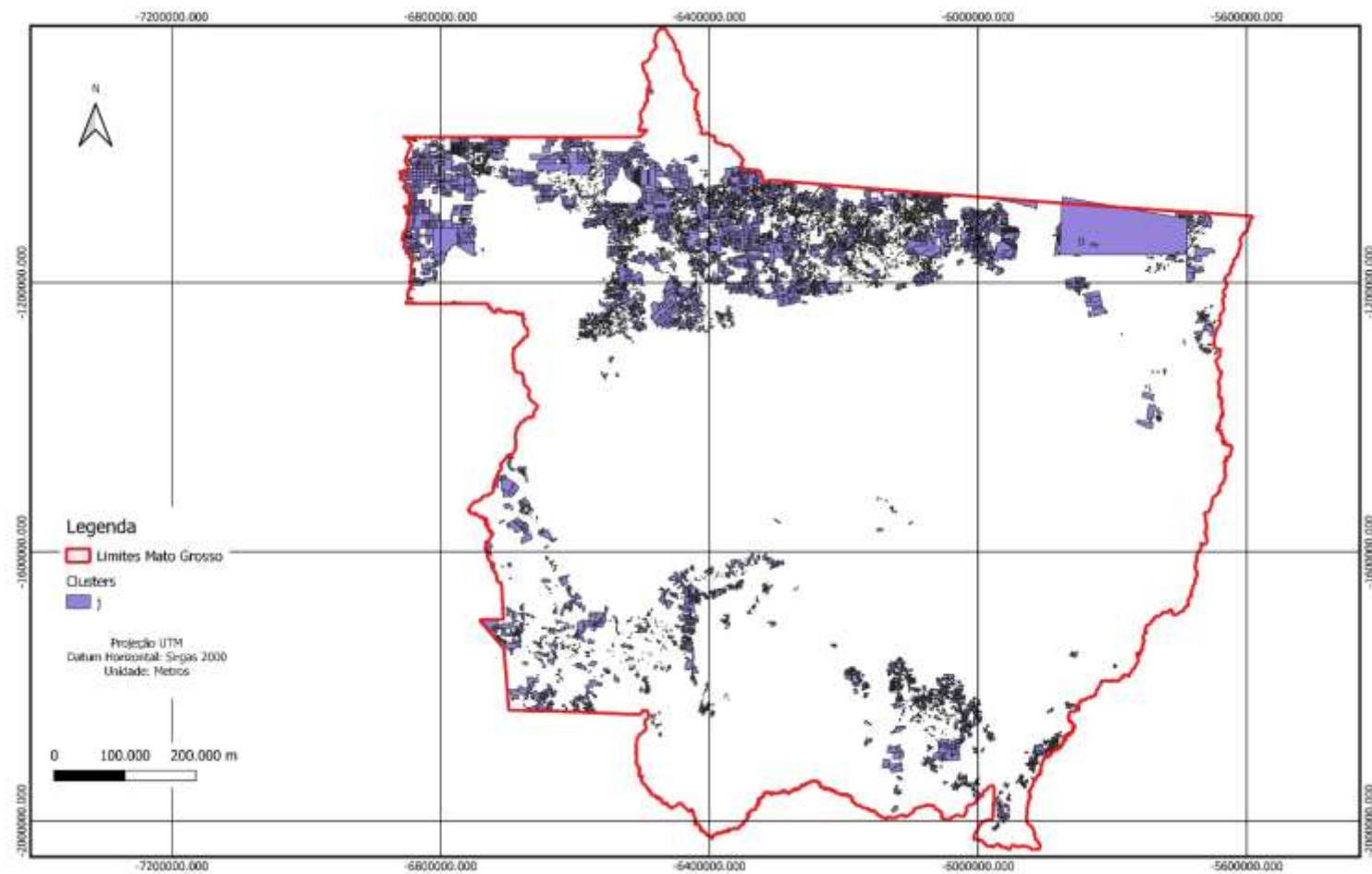
Fonte: O autor.

Figura 12. Espacialização *cluster* I dos imóveis rurais do Mato Grosso



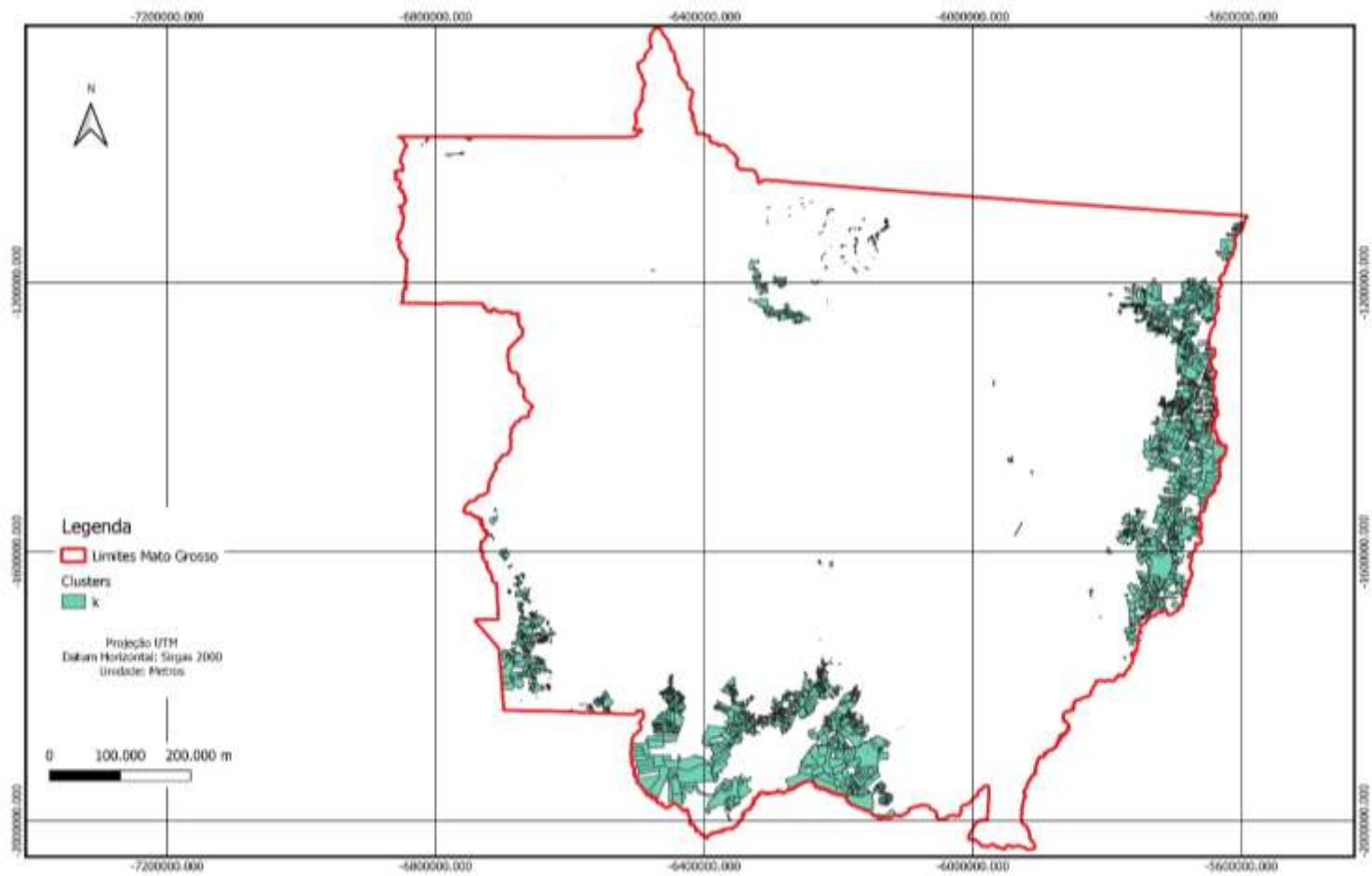
Fonte: O autor

Figura 13. Espacialização *cluster* J dos imóveis rurais do Mato Grosso



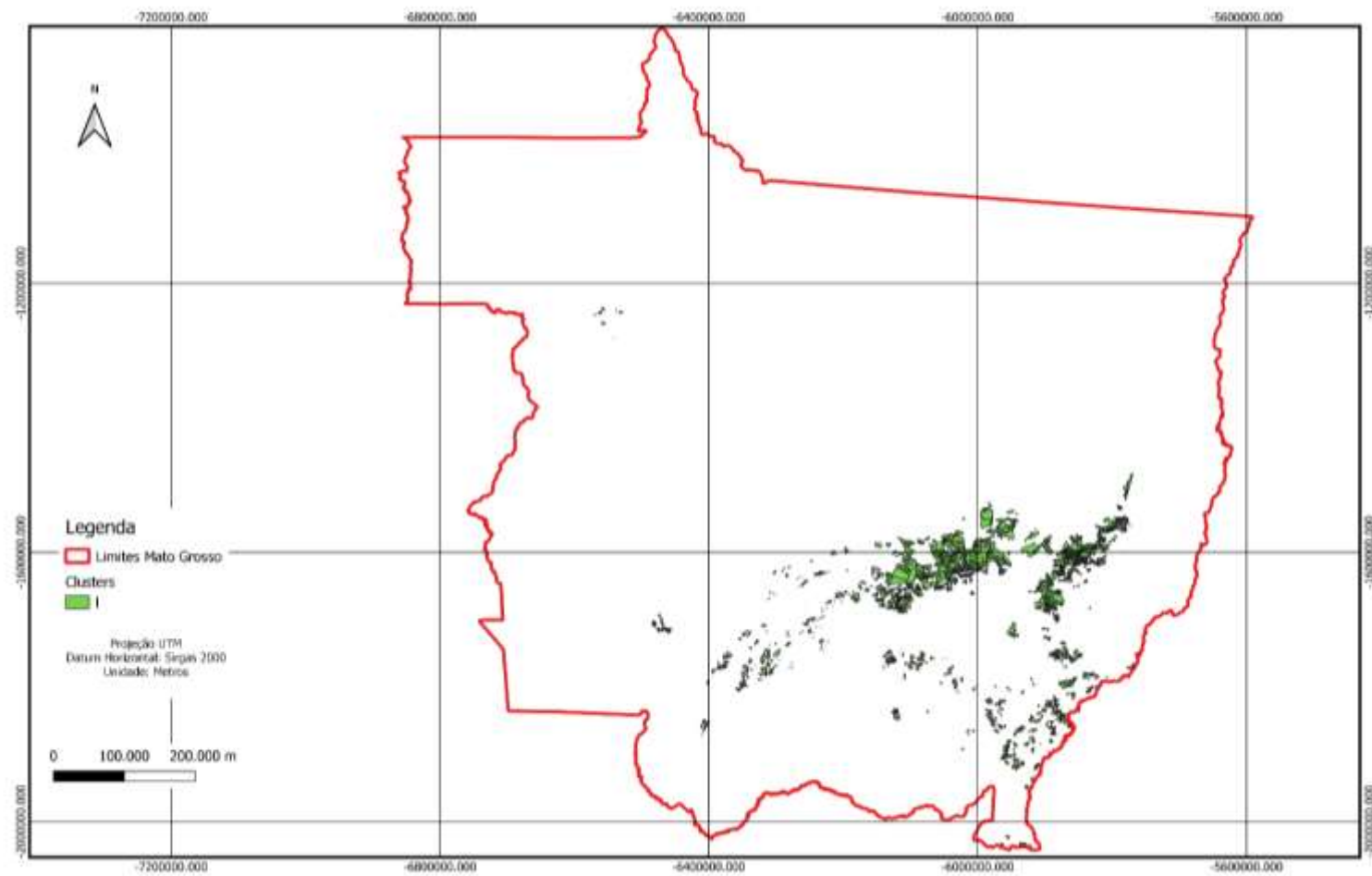
Fonte: O autor.

Figura 14. Espacialização *cluster* K dos imóveis rurais do Mato Grosso



Fonte: O autor.

Figura 15. Espacialização *cluster* L dos imóveis rurais do Mato Grosso



Fonte: O autor.

4 CONCLUSÕES

Os *clusters* gerados pelo algoritmo foram eficientes na distinção e agrupamento de imóveis utilizando dados de rodovias, ferrovias, hidrovias, tipos de solo, disponibilidade hídrica e uso e ocupação do solo. Houve grupos que se destacaram para cada característica, além de outros que foram resultados da combinação de diferentes variáveis.

A variável área de agropecuária, e as referentes aos tipos de solo: areias quartzosas, cambissolo, latossolo, podzólico, plintossolo e concrecionado apresentaram as maiores representatividades no modelo, uma vez que cada cluster formado abrangeu grande parte dos imóveis com tais características. Além disso, os *clusters* transcenderam os limites regionais do estado, possibilitando correlação de imóveis de locais distintos, possibilitando novos estudos sobre o grau de similaridade de imóveis em diferentes regiões.

Convém destacar algumas dificuldades enfrentadas no desenvolvimento do trabalho, que podem ser resolvidas ou abordadas em trabalhos futuros. A princípio, a baixa qualidade dos dados geográficos oriundos do Cadastro Ambiental Rural, que acarretou em esforços para correções topológicas grosseiras, visando não prejudicar as análises realizadas nas fases seguintes do modelo. Além disso, o grande volume de dados processados acarreta em alta disponibilidade computacional e de tempo para realização, prejudicando a inclusão de variáveis relacionadas ao clima ao modelo. Outro fator é a disponibilidade de metodologias para validação dos agrupamentos, ao restringir a validação ao modelo utilizado pelo Instituto de Terras do Mato Grosso há uma sub utilização da base de dados gerada, uma vez que ela apresenta maior quantidade de variáveis, havendo possibilidade de retratar subconjuntos ainda mais específicos e que retratam melhor a realidade.

Como sugestão para trabalhos futuros com este tema, recomenda-se utilização de modelos de *Machine Learning* que façam uso de imagens, de forma a permitir a melhor caracterização e possível precificação pixel a pixel. Por fim, reitera-se a validação em campo dos agrupamentos obtidos, principalmente quanto a dispersão da maioria dos clusters por todas as regiões do Mato Grosso.

5 REFERÊNCIAS

- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-3/NBR14653-3: Avaliação de bens - Parte 3: Imóveis rurais e seus componentes**, 2004. Acesso em: 8 jan. 2022
- AQUINO, J. N. **Uso de técnicas de modelagem e análise espacial na seleção de áreas ideais para construção de aterros sanitários em regiões metropolitanas densamente povoadas**. [s.l.: s.n.].
- BRASIL. INSTRUÇÃO NORMATIVA SRF Nº 84. Brasília. 2001.
- BRASIL. Lei n 12.651, de 25 de maio de 2012. Brasília. 2012.
- BRASIL. Decreto N º 294. 7 nov. Brasília. 2019.
- BRASIL. **Estatísticas e Dados Abertos dos Imóveis Rurais — Português (Brasil)**. Disponível em: <<https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/portal-cnir/estatisticas-e-dados-abertos>>. Acesso em: 21 jan. 2022.
- BRASIL. Portaria 08/2021. 2021.
- CARMEM, A. et al. **RELATÓRIO DE ANÁLISE DE MERCADO DE TERRAS NO ESTADO DE MATO GROSSO/SR (13) – RAMT/MT**. 2019.
- FARIAS, J. R. B. et al. **Caracterização de risco de déficit hídrico nas regiões produtoras de soja no Brasil**. Revista Brasileira de Agrometeorologia, v. 9, n. 3, p. 415–421, 2001.
- GEOPANDAS. **GeoPandas 0.9.0 — GeoPandas 0.9.0 documentation**.
- IBGE. **IBGE | Cidades@ | Mato Grosso | Panorama**. Disponível em: <<https://cidades.ibge.gov.br/brasil/mt/panorama>>. Acesso em: 14 jan. 2022.
- INCRA. **Cadastro de Imóveis Rurais - Situação Jurídica - Titularidade Particular**. Disponível em: <<https://antigo.incra.gov.br/media/docs/estatisticas-imoveis-rurais/brasil.pdf>>. Acesso em: 8 jan. 2022.
- INTERMAT. **Mapa Pedológico do Estado do Mato Grosso**. Cuiabá, 2000. Disponível em: <<http://www.intermat.mt.gov.br/-/11302979-dsee-2000>>. Acesso em: 6 mar. 2022
- MAPBIOMAS. **Mapeamento Anual da Cobertura e Uso do Solo no Brasil (MapBiomias)**.
- MINISTÉRIO DO MEIO AMBIENTE. Sistema de Informações Cartográficas da SEPLAN-MT. Secretaria do Estado do Mato Grosso de Planejamento e Gestão, 2020.
- PYTHON SOFTWARE FOUNDATION. Python Language Reference, version 2.7. Python Software Foundation, 2013.
- QGIS.ORG. QGIS Geographic Information System. QGIS Association.

RASTERIO. Rasterio, 2018. Disponível em: <<https://rasterio.readthedocs.io/en/latest/>>. Acesso em: 6 mar. 2022

SAJI, B. In-depth Intuition of K-Means Clustering Algorithm in Machine Learning. Data Science Blogathon, 20 jan. 2021.

SICAR. Imóveis. Disponível em: <<https://www.car.gov.br/publico/imoveis/index>>. Acesso em: 8 jan. 2022.

SINAGA, K. P.; YANG, M. S. Unsupervised K-means clustering algorithm. **IEEE Access**, v. 8, 2020.

WEBER, E. Avaliação de áreas para instalação de aterro sanitário através de análises em SIG com classificação contínua dos dados. Porto Alegre: UFRS, n. July, 2000.