



**BRUNO WILLIAM DA SILVA SANTOS**

**SELEÇÃO DE VARIÁVEIS PARA A DELIMITAÇÃO DE  
ZONAS DE MANEJO UTILIZANDO MÉTODOS DE  
MINERAÇÃO DE DADOS**

**LAVRAS-MG  
2020**

**BRUNO WILLIAM DA SILVA SANTOS**

**SELEÇÃO DE VARIÁVEIS PARA A DELIMITAÇÃO DE ZONAS DE MANEJO  
UTILIZANDO MÉTODOS DE MINERAÇÃO DE DADOS**

Trabalho de Conclusão de Curso apresentado à  
Universidade Federal de Lavras, como parte  
das exigências do curso de Agronomia para  
obtenção do Título de Bacharel.

Prof. Dra. Michele Duarte de Menezes  
Orientadora

**LAVRAS-MG  
2020**

**BRUNO WILLIAM DA SILVA SANTOS**

**SELEÇÃO DE VARIÁVEIS PARA A DELIMITAÇÃO DE ZONAS DE MANEJO  
UTILIZANDO MÉTODOS DE MINERAÇÃO DE DADOS**

**VARIABLES SELECTION FOR MANAGEMENT ZONES DELIMITATION USING  
DATA MINING METHODS**

Trabalho de Conclusão de Curso apresentado à  
Universidade Federal de Lavras, como parte  
das exigências do curso de Agronomia para  
obtenção do título de Bacharel.

APROVADO em 28 de agosto de 2020.

Prof. Dra. Michele Duarte de Menezes – UFLA

Ms. Alessandro Alvarenga Pereira da Silva – REHAGRO

Ms. Fabio Arnaldo Pomar Avalos – UFLA

Prof. Dra. Michele Duarte de Menezes  
Orientadora

**LAVRAS – MG  
2020**

## AGRADECIMENTOS

Aos meus pais, Claudio e Adriana, pelo incentivo e esforço a mim dedicados, sem os quais esta conquista não seria possível. A minha irmã, por sempre não deixar esquecer o que devo fazer.

À minha avó, por estar acreditando sempre em meu potencial e me incentivar nas minhas lutas.

À minha namorada, Laura, por ser a parceira que é, e sempre está comigo

Aos Núcleos de estudo dos quais participei: NECS e G-Milho, e a todas as pessoas com as quais pude conviver.

Aos amigos que fiz durante minha estadia em Lavras, em especial à República Sucupira e República Pé da Serra. Agradeço por poder chamá-los de família!

À Professora Dra. Michele pela orientação e dedicação. Obrigado por compartilhar comigo conhecimentos que serão, sem dúvidas, de grande valia em minha formação.

Ao amigo Alessandro, pela prontidão e mentorias ao longo de minha graduação.

Ao Fabio, por se dispor prontamente a participar como membro na banca examinadora.

Agradeço o senhor William da Fazenda Retiro, por disponibilizar a sua área de produção e seus dados para estudo.

E principalmente a Deus e Nossa Senhora Aparecida, por conduzirem meus passos, guiando-me e protegendo-me.

## RESUMO

Atualmente, a gestão da variabilidade espacial dos sistemas de produção com técnicas de agricultura de precisão emprega a amostragem em grade regular para a adubação a taxa variável. No entanto, o método de gestão através de zonas de manejo pode apresentar maior eficiência. As zonas de manejo, por sua vez, empregam espacialmente características selecionadas para a sua delimitação que interferem na produtividade de grãos, podendo guiar operações de amostragem e aplicação de insumos, dando maior atenção aos locais onde realmente se necessita. Portanto, é necessária a seleção de variáveis que apresentem significância quando relacionadas à produtividade das culturas de interesse. Diante disso, o presente trabalho tem o objetivo de selecionar variáveis que estão disponíveis no atual cenário de empresas de consultoria em agricultura de precisão que possam ter alta correlação com a produtividade, que são de baixo custo de obtenção e estáveis entre as safras. Os dados foram obtidos na região sul de Minas Gerais, no município de Carrancas, Fazenda Retiro, na segunda safra de 2018, onde foi cultivada a cultura do milho. As variáveis mais significantes quando relacionadas com a produtividade da cultura do milho foram a relação cálcio dividido por potássio, fosforo resina, relação magnésio dividido por potássio, saturação por bases coletadas nas profundidades de 0 a 10 cm e de 10 a 20 cm e o NDVI da cultura do milho.

**Palavras-chave:** Agricultura de precisão. Mineração de dados. Zonas de manejo.

## ABSTRACT

Currently, the management of the spatial variability of production systems using precision Agriculture techniques to regular grid sampling for variable rate fertilization. However, the management method through management zones can be more efficient. The management zones, employ spatially selected characteristics for their delimitation that interfere in yield, being able to guide sampling operations and application of inputs, it can result better attention to places where it is necessary. Therefore, it is necessary to select variables that are significant when related to the corn yield. In view of this, the current research has the objective of selecting variables that are available in the current scenario of precision agriculture consulting companies that may have a high correlation with productivity, which are of low cost to obtain and stable from season to season. The data were obtained in the southern region of Minas Gerais, in the Carrancas city, Retiro Farm, in the second season of 2018 where the corn crop was grown. The most significant variables when related to the corn crop productivity were the calcium ratio divided by potassium, phosphorus resin, magnesium ratio divided by potassium, base saturation at sample's depths and the NDVI of the corn crop.

**Keywords:** Precision agriculture. Data mining. Management Zones.

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>7</b>
<b>2 REFERENCIAL TEÓRICO .....</b>	<b>9</b>
2.1 Agricultura de precisão.....	9
2.2 Geoestatística .....	10
2.3 Variabilidade dos sistemas de produção .....	12
2.4 Mineração de variáveis .....	14
2.5 Zonas de manejo .....	15
<b>3 OBJETIVOS .....</b>	<b>17</b>
3.1 Objetivo geral.....	17
3.2 Objetivos específicos.....	17
<b>4 MATERIAIS E MÉTODOS .....</b>	<b>18</b>
4.1 Área estudada .....	18
4.2 Análise da fertilidade do solo.....	18
4.3 Mapeamento da produtividade .....	21
4.4 Interpolação dos dados de fertilidade do solo e produtividade.....	22
4.5 Atributos obtidos por sensoriamento remoto .....	24
4.5.1 Índice de vegetação.....	24
4.5.2 Atributos do terreno.....	25
4.6 Mineração de dados: seleção das variáveis mais importantes .....	26
4.7 Delimitação de zonas de manejo e o número de zonas.....	26
<b>5 RESULTADOS E DISCUSSÃO .....</b>	<b>27</b>
5.1 Acurácia da interpolação espacial .....	27
5.2 Regressão linear múltipla: variáveis de maior significância .....	29
5.3 Delimitação de zonas de manejo .....	32
<b>6 CONCLUSÃO.....</b>	<b>37</b>
<b>REFERÊNCIAS .....</b>	<b>38</b>

## 1 INTRODUÇÃO

A produção de alimentos é essencial para a existência da humanidade. A comunidade civil e científica, cada vez mais, se dá conta que existe a necessidade de métodos mais sustentáveis de produção com a finalidade de impactar menos o meio ambiente (RIBEIRO; JAIME; VENTURA, 2017). O conhecimento e exploração da variabilidade presente em sistemas de produção podem ser uma alternativa de aumentar a sustentabilidade da produção, proporcionando o uso eficiente de insumos e recursos aplicados ao campo.

Na agricultura, a maioria dos insumos utilizados são de caráter finito, ou seja, um dia eles podem acabar, exemplo disso são os fertilizantes fosfatados que tem uma grande demanda pelos produtores (PANTANO et al., 2020).

Aplicar técnicas de agricultura de precisão, pode auxiliar no uso sustentável de insumos por aplicar doses necessárias, onde realmente necessita. Além disso, pode baixar os custos de produção aumentando a competitividade de mercado dos produtos (MULLA; KHOSLA, 2015).

Dentre os métodos de agricultura de precisão que tem o objetivo de determinar e gerenciar a variabilidade espacial dos sistemas de produção, a utilização de zonas de manejo é uma das estratégias que apresenta maior eficiência. Embora seja um método de baixo custo quando comparado a outros métodos de gestão da variabilidade, a geração de zonas de manejo necessita de muito conhecimento técnico (MULLA; KHOSLA, 2015).

Compreender e gerar zonas de manejo com alta acurácia é de suma importância que os dados utilizados sejam bem coletados, analisados e interpolados. As dificuldades encontradas nesse processo geram uma grande oportunidade de novas pesquisas para a geração de métodos cada vez mais eficientes, fácil execução e de baixo custo que possam ser utilizados no campo (KHOSLA et al., 2010).

A nova tendência da estatística espacial é a utilização de métodos de agrupamento para a delimitação de zonas com características semelhantes, esse método recebe o nome de *Cluster* (NAWAR et al., 2017). O método *Cluster* apresenta bastante aceitação no meio acadêmico, muitos resultados comprovam a eficiência, embora haja diferentes graus de eficiência entre diferentes pesquisas, tem se a existência de muitas pesquisas em andamento.

É notável que anos após anos, acontece a descoberta de novas variáveis que podem auxiliar na explicação de diferentes níveis de produtividade em uma determinada área. O que acontece nesse comportamento se deve a tendências de mercado. Isso pode ser explicado por conta que empresas do setor de agricultura de precisão fazem marketing sensacionalista com o

objetivo de vender seus produtos. Porém, o técnico tem o desafio de compreender as variáveis e identificar a melhor escolha (KHOSLA et al., 2010).

Atualmente, empresas de agricultura de precisão muitas das vezes fazem amostragens em arranjos que não conseguem fazer a verificação da dependência espacial das variáveis analisadas, isso acontece devido ao alto custo de coleta e análise aplicadas no conhecimento da variável em questão. Nesse sentido por meio das zonas de manejo que usa mapas de produtividade é possível realizar relações de causa e efeito, sendo o mapa de produtividade o efeito sempre utilizado para visualização das informações.

O objetivo do presente trabalho é compreender as variáveis disponíveis no atual cenário de empresas de consultoria em agricultura de precisão, de forma a identificar variáveis que apresentam alta correlação com a produtividade de grãos, visando baixo custo econômico de obtenção, estabilidade entre safras e que tenha alta eficiência na geração de zonas de manejo por método estatístico de agrupamentos. Foram obtidos dados de relevo (altitude, altitude acima do nível dos canais, índice topográfico de umidade), atributos químicos do solo coletados nas profundidades de 0-10 e 10-20 cm, mapas de produtividade e índice de vegetação NDVI obtidos por sensoriamento remoto.

## 2 REFERENCIAL TEÓRICO

### 2.1 Agricultura de precisão

A agricultura de precisão ao longo da história se diferencia da agricultura tradicional por levar em consideração a variabilidade espacial e temporal de atributos oriundo dos sistemas de produção agrícola. Com a evolução da agricultura, os agricultores começaram a perceber que necessitavam diferenciar os tratamentos culturais de suas plantações, levando em consideração as peculiaridades distintas em seus campos de produção (MOLIN; AMARAL; COLAÇO, 2015). Dessa maneira as intervenções eram feitas de maneira localizada, podendo assim maximizar o uso dos recursos disponíveis.

Na academia o tema começou a se destacar a partir do trabalho dos pesquisadores Linsley e Bauer (1929), onde os autores mapearam a acidez do solo e produziram uma circular técnica de número 346 da Estação Experimental Agrícola da Universidade de Illinois, onde eles buscavam orientar os agricultores na correta aplicação de calcário. Os autores procuravam mostrar a variabilidade espacial da acidez do solo e a maneira de fazer a intervenção correta na área, levando em consideração o uso consciente e demonstrando que nem toda a área necessitava de aplicação de calcário (LINSLEY; BAUER, 1929). Contudo, o tema voltou a ter maior expressão no ano 1992, quando aconteceu o primeiro workshop abordando o tema com o título de “Soil Specific Crop Management: A workshop on Research and Development Issues” (KHOSLA, 2010), esse evento foi o precursor para dar origem ao principal evento da área, o evento recebe o nome hoje em dia de Congresso Internacional de Agricultura de precisão.

Existem muitos conceitos para definir a agricultura de precisão, a grande maioria deles colocam em evidenciar a variabilidade espacial e temporal dos sistemas de produção agrícola, a aplicação de insumos em taxa variável e a promoção da sustentabilidade ambiental e econômica gerada a partir do uso racional de insumos agrícola.

O Ministério da Agricultura, Pecuária e Abastecimento, define a agricultura de precisão como um sistema de gerenciamento agrícola baseado na variação espacial e temporal da unidade produtiva e visa ao aumento de retorno econômico, à sustentabilidade e à minimização do efeito ao ambiente (MAPA, 2017).

Para BRAMLEY (2009), a agricultura de precisão é um conjunto de tecnologias que promovem melhor gerenciamento da produção agrícola, e reconhece que a produtividade potencial de sistemas agrícolas pode variar consideravelmente em curtas distâncias. O autor também afirma que o uso da agricultura de precisão pode ser considerado um meio de aumentar

a chance de que as estratégias corretas de manejo de culturas sejam implementadas no lugar certo e na hora certa.

Segundo alguns autores que estão envolvidos no âmbito comercial agrícola de tecnologia da informação, máquinas agrícolas e agricultura digital, definem a agricultura de precisão com ênfase em tecnologias, inteligência artificial, internet das coisas, análise de dados por meio de algoritmos utilizando o *Big Data*, geolocalização, automação e robótica. Essas tecnologias são aplicadas em agricultura de precisão, porém são tratadas como ferramentas e não agricultura de precisão. Nesse contexto, atualmente as práticas de agricultura de precisão estão desacreditadas no mercado, por conta de empresas que oferecem soluções milagrosas sem levar em consideração os princípios primordiais da agricultura de precisão.

No Brasil as primeiras atividades relacionadas a agricultura de precisão são datadas no ano de 1995, com a importação de equipamentos e colhedoras equipadas com monitores de produtividade (VIANA et al., 2009). Apesar disso, a agricultura de precisão começou a ganhar maior notoriedade pelos agricultores nos anos 2000, por conta do surgimento das primeiras empresas de consultoria e de serviços (MOLIN; AMARAL; COLAÇO, 2015). O maior enfoque dessa atividade no Brasil, está relacionada a aplicação de fertilizantes e corretivos em taxa variável, mas também em passos menores a dotação de outras práticas como o plantio de culturas em taxa variável, o manejo de pragas utilizando ferramentas de agricultura de precisão e a utilização de sensoriamento remoto para diversos fins vem ganhando maior espaço nas companhias agrícola do Brasil.

Atualmente, a sociedade civil e científica tem dado maior atenção na forma que os agricultores estão cultivando os alimentos, cada vez mais estão cobram formas mais sustentáveis de produção, com menor uso de pesticidas, utilização de produtos menos nocividade ao ambiente entre outras formas de manejo de menor impacto ambiental (RIBEIRO; JAIME; VENTURA, 2017). O uso da agricultura de precisão pode se encaixar nesse contexto, pois emprega várias técnicas de manejo agrícola, que podem contribuir muito para a preservação do meio ambiente e promovendo o uso mais sustentável de insumos agrícola e diminuindo os custos de produção. Além disso, contribui muito no âmbito de segurança alimentar, produzindo mais em menor espaço territorial.

## **2.2 Geoestatística**

A geoestatística diferencia-se das estatísticas clássicas por apresentar as suas variáveis espacialmente distribuídas e o uso de técnicas como krigagem e cokrigagem (KHOSLA et al., 2010). Na agricultura a geoestatística tem o importante papel de auxiliar na caracterização da

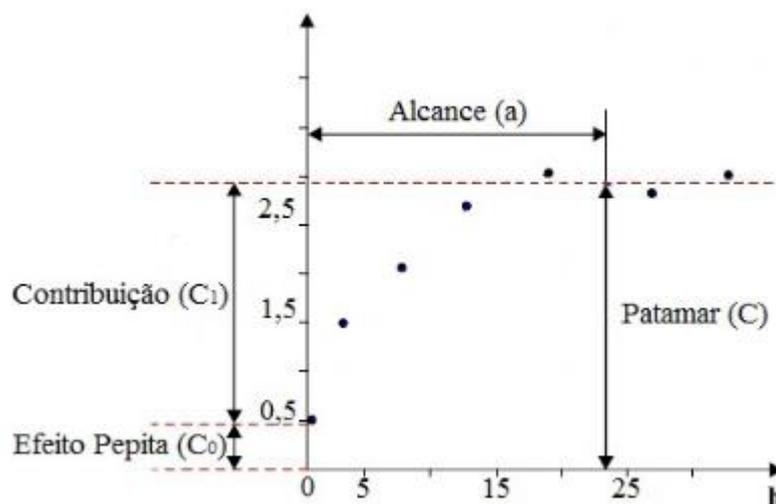
variabilidade espacial existente em sistemas de produção, de tal maneira oferecer a opção de intervenção localizada de manejo. Esse tipo de intervenção pode apresentar maior eficiência agrônômica e uso eficiente de recursos.

O método geostático para a geração de mapas deve ser escolhido com bastante atenção, deve se ter conhecimento da variável estudada, saber se ela apresenta dependência espacial e o tipo de dado que está trabalhando. Os métodos mais utilizados são os determinísticos e os geoestatísticos. O método quadrado do inverso da distância, utilizado por muitas empresas que presta consultoria em agricultura de precisão, consegue interpolar qualquer tipo de dado apresentado, para o método é indiferente o conjunto de dados ter dependência espacial ou não. Por outro lado, métodos como a krigagem, oferece o semivariograma.

A Krigagem é um método geoestatístico que permite interpolar valores desconhecidos, a partir de valores conhecidos. Para se utilizar o método, o recomendado é analisar a dependência espacial da variável em estudo com o auxílio do semivariograma (GREGO C. R., OLIVEIRA R. P., 2014). Existe vários tipos de krigagem, porém, a ordinária é o tipo mais utilizado, por conta de ser um método local de predição que pode estimar um ponto não amostrado, pelo resulta da combinação linear dos valores encontrados na vizinhança próxima.

O semivariograma apresentado na figura 1, é um importante componente a ser analisado no momento de executar a Krigagem. Quando bem ajustado o semivariograma, os valores preditos em locais onde não foi feita a amostragem podem ser o mais próximo dos valores reais. (BASTOS, 2017)

Figura 1 – Semivariograma



Fonte: (BASTOS, 2017)

A validação cruzada também tem que ser observada no processo de krigagem, ela nos mostra a qualidade da predição feita pelo modelo utilizado.

Para o conhecimento detalhado das áreas é necessário planejar um sistema amostral adequado. O método mais utilizado no Brasil, é o de amostragem de solo em grade, onde leva em consideração uma malha amostra pré-estabelecida. A metodologia pode ter muitos arranjos, podendo apresentar pontos equidistantes ou não (MOLIN; AMARAL; COLAÇO, 2015).

A equidistância amostral de solos é bastante discutida por diversos autores. A malha amostral em nível comercialmente mais utilizada varia de  $173 \times 173$  m a  $223 \times 223$  m. Para Mueller et al. (2001), o arranjo comercial é inadequado para a confecção de mapa com alta precisão de representação de nutrientes do solo, porém o uso de arranjo menor de amostragem, implica na geração de maior custo por hectare. Alguns autores defendem que para ter um bom mapa de concentração de nutrientes, é necessário malhas amostrais variando de  $60 \times 60$  m até no máximo de  $76 \times 76$  m (KHOSLA et al., 2010). Outro fator que interfere na qualidade do mapa, é o número de sub amostras e o raio de coleta, para uma boa amostragem é necessário se atentar a esses fatores.

### **2.3 Variabilidade dos sistemas de produção**

Os sistemas de produção além da variabilidade espacial de atributos edafoclimáticos, apresentam também variabilidade temporal e variabilidade antrópica (PIRES et al., 2004). Nos campos de produção esses tipos de variabilidades apresentam alta correlação com a variabilidade da produtividade e podem ajudar a explicar a variabilidade espacial dela.

A variabilidade espacial presente nos atributos dos solos é existente pela influência dos processos de pedogênese. O material de origem, o relevo, a ausência e presença de distúrbios como mudanças climáticas e eventos de deposição e erosão contribuem no processo de pedogênese (BREEMEN; BUURMAN, 2002). Determinados eventos, podem acontecer em diferente intensidade, e variar muito em pequenas escalas de espaço, dando origem em processos pedogenéticos diferentes. Nessa visão de formação dos solos, podemos entender como surge a variabilidade espacial química, física, biológica e topográfica dos solos.

Entre os tipos de variabilidade, existe também a variabilidade temporal. Tal fenômeno advém do fato de ocorrência de eventos naturais e antrópicos. Esses eventos podem contribuir nas mudanças edafoclimáticas e biogeoquímicas dos sistemas de produção. Os fatores que contribuem na variabilidade temporal, podem acontecer em diferentes intensidades nas estações do ano, devido a condições climáticas e estações de cultivo.

O sistema de manejo adotado pelo agricultor pode alterar os atributos químicos, físicos e biológicos dos solos, sendo a camada superficial a que apresenta maior tendência de mudanças provocadas por ações antrópicas (NOETZOLD et al., 2018). A agricultura convencional sempre aplicou técnicas de manejo que consideravam o solo com homogêneo, além de não atender as exigências reais de introdução de insumos no sistema, as intervenções podem apresentar riscos de falhas mecânicas em implementos e condições desfavoráveis. Exemplo disso, são as aplicações de calcário em dias de alta intensidade de vento, na condição apresentada, o coeficiente de variação de calcário tem aumento exorbitante, apresentando assim, alta variabilidade na distribuição de calcário, o que vai gerar alta variabilidade nos atributos químicos provocados por calagem imprópria. Nesse contexto, os sistemas de produção que não são homogêneos, podem apresentar maior variabilidade, esse tipo de variabilidade gerada pelo homem recebe o nome de variabilidade antrópica.

Considerando os aspectos inerente às plantas cultivadas, elas possuem potencial genético para alcançar maior produtividade em situação de campo. Exemplo disso, a soja possui potencial genético para alcançar 339,9 sacas/ha (sc/ha) (NAVARRO JÚNIOR; COSTA, 2002), atualmente a média da região sudeste é de 60,5 sc/ha (CONAB, 2020). Analisando o assunto, podemos afirmar que necessitamos de maior entendimento do funcionamento das relações de interação ambiente, solo e planta, para que possamos alcançar a máxima produtividade potencial.

A fertilidade do solo tem papel indiscutível em relação a obtenção de elevadas produtividade (FURTINI NETO et al., 2001). A atividade de produção agrícola demanda elevado investimento, por conta disso os gestores estão sempre buscando melhores formas de tornar os solos mais férteis.

Segundo SANTI et al. (2012), analisando dados de produtividade e elementos químicos do solo com métodos geostática, o elevado teor de potássio no solo é o que melhor explica a variabilidade espacial da produtividade das culturas de grãos, isso acontece em razão do desbalanço das relações Calcio e potássio e magnésio e potássio. Para (RAIJ, 2011), a relação entre o potássio, cálcio e o magnésio no solo pode acentuar a deficiência de magnésio. Diante disso, podemos resumir que a busca de equilíbrio desses nutrientes pode refletir no aumento de produtividade.

Para GOZDOWSKI et al. (2014), as propriedades do solo com maior relevância quando relacionadas com a variabilidade espacial da produtividade são o carbono orgânico do solo e o teor de areia. Onde o alto conteúdo de areia, os solos apresentam baixo potencial produtivo e zonas com alto carbono orgânico, os solos apresentam alto potencial produtivo. Outras

propriedades que apresentam boa correlação com a variabilidade espacial da produtividade são a capacidade de troca de cátions, magnésio, potássio, condutividade elétrica e pH.

Para MOLIN; AMARAL; COLAÇO, (2015), uma das melhores variáveis para delimitação de zonas de manejo é a classificação de solos. No entanto, esse tipo de levantamento quando feito para conhecer a variabilidade em alta resolução, acaba apresentando alto custo econômico, o que inviabiliza muito agricultores de usar essa metodologia.

Absolutamente, a agricultura de precisão não é a forma mais eficiente de buscar resposta entre as interações planta, solo e ambiente. Na verdade, o papel da agricultura de precisão é utilizar conhecimentos estabelecidos e aplicar as suas ferramentas para buscar soluções que permita conhecer e diminuir a variabilidade alvo, logo, todo o sistema produtivo proporcionara a melhor condição possível para o cultivo.

## **2.4 Mineração de variáveis**

Nos últimos anos as formas de cultivo vêm sofrendo mudanças constantes com o surgimento de novas tecnologias e aplicação dos conceitos de agricultura de precisão. A aplicação de ferramentas de agricultura de precisão produz muitos dados, a interpretação e retiradas de informações se torna muito complexa e de difícil execução. A técnica da mineração de dados, do inglês *Data Mining*, poder ser uma alternativa eficaz para extrair conhecimento a partir de grandes volumes de dados, descobrindo relações ocultas, padrões e gerando regras para prever e correlacionar (GALVÃO; MARIN, 2009).

Para SANTOS (2009), a mineração de dados deve seguir os seguintes passos:

- a) Associações: consiste na seleção de variáveis que podem contribuir na explicação da variável em estudo.
- b) Classificação: descoberta de uma função preditiva que consegue classificar os dados em classes discretas.
- c) Regressão: estabelecer uma função preditiva, com o objetivo de calcular o valor numérico real.
- d) Agrupamento: descoberta de grupos naturais de dados que possivelmente indicam similaridade entre eles.

Na seleção de variáveis para a delimitação de zonas de manejo é essencial a obtenção de dados com alta correlação com a produtividade, fácil obtenção e custo baixo. As variáveis utilizadas podem partir de dados baseados no conhecimento do agricultor, geomorfologia,

análise química do solo, classes de solo, mapas de produtividade, cobertura do solo, sensoriamento remoto e sensores proximal (NAWAR et al., 2017).

No mercado de empresas de agricultura de precisão, existem muitas variáveis que podem auxiliar na explicação de diferentes níveis de produtividade em uma determinada área. No entanto, isso acontece devido a tendências de mercado, onde empresas fazem marketing com o objetivo de vender seus produtos. Porém, nem sempre esses dados são precisos.

A construção de modelos de regressão multivariada é uma boa ferramenta na mineração de dados. Esse tipo de análise permite a seleção de variáveis que apresenta diferentes significâncias na produtividade. Encontrar as melhores variáveis específicas que explique a produtividade de áreas cultivadas com grão, pode ser muito difícil quando a existência de diversas possíveis candidatas.

## **2.5 Zonas de manejo**

A primeira definição dada para zona de manejo foi escrita por Mulla (1991). O autor descreve que cada zona de manejo deve idealmente representar partes do campo que são relativamente semelhantes e homogêneas no status de fertilidade do solo, para que uma recomendação uniforme e diferente de fertilizante possa ser feita para cada zona (MULLA; KHOSLA, 2015).

Para VRINDTS et al (2005), a zona de manejo é como uma sub-região de um campo que possui uma combinação relativamente homogênea de fatores de limitação de produção, para os quais uma taxa única de um insumo específico de uma cultura é apropriada para obter a máxima eficiência dos insumos agrícolas.

Atualmente, o uso de zonas de manejo tem sido empregado para guiar diversas operações de caráter localizado. Em virtude de conhecer a variabilidade existente, a operação localizada busca dar tratamento diferenciado as zonas, levando em consideração o potencial de resposta em qualidade, produtividade, rendimento operacional e diminuição de custos.

Na realidade de muitas fazendas agrícola, a operação de distribuição de corretivos e fertilizantes em taxa fixa varia. Analisando a informação, podemos presumir que quando essas máquinas submetidas a operações de distribuição de corretivos e fertilizantes em taxa variável, a operação não terá a precisão detalhada nos mapas de prescrição. Outro aspecto relevante é o custo por hectare quando se deseja conhecer o campo de produção do detalhamento. Recursos humanos também são um problema enfrentado na área de agricultura de precisão, vez que não

se encontram pessoas capacitadas que tenham conhecimento para tomar decisões importantes relacionadas à utilização de coletas de dados em campo, tratamentos de dados e geoestatísticas. A utilização de zonas de manejo consegue minimizar os problemas apresentados na atual conjuntura de grandes empresas e consultores agrônômicos que utilizam agricultura de precisão, de maneira economicamente viável, precisa e de simples entendimento (KHOSLA et al., 2010).

A determinação de zonas de manejo pode ser feita por diversos métodos, dentre os métodos o que mais chama atenção é o método de agrupamento que recebe o nome de *Cluster*. Por se tratar de um método que permite agrupar variáveis em diferentes agrupamentos de mesma similaridade, esse tipo de análise pode ser bastante explorado.

O método de análise *Multivariate Clustering* (ARCGIS PRO, 2020), é uma importante ferramenta de agrupamento que permite fazer a operação em dois diferentes algoritmos, sendo eles *K means* or *K medoids*. O algoritmo *K medoids*, é mais robusto quando a presença de *outliers*.

Uma grande dificuldade enfrentada na determinação de zonas de manejo é a determinação do número de zonas. O método *Multivariate Clustering*, apresenta uma ferramenta baseada no valor de *pseudo F-statistic values*, que permite melhor interpretação do agrupamento dos dados em estudo.

### 3 OBJETIVOS

Compreender e selecionar variáveis disponíveis no atual cenário de empresas de consultoria em agricultura de precisão, que tenha alta correlação quando contrastadas com a variabilidade espacial da produtividade de grãos em um Cambissolo Háptico, para a utilização na delimitação de zonas de manejo pelo método de Cluster.

#### 3.1 Objetivos específicos

- a) Identificar propriedades do solo e de culturas economicamente viáveis de amostragem com técnicas de agricultura de precisão que tenha correlação com a produtividade de grãos.
- b) Identificar propriedades do solo e culturas que sejam estáveis de safra a safra.
- c) Gerar zonas de manejo através do método de *Cluster*.

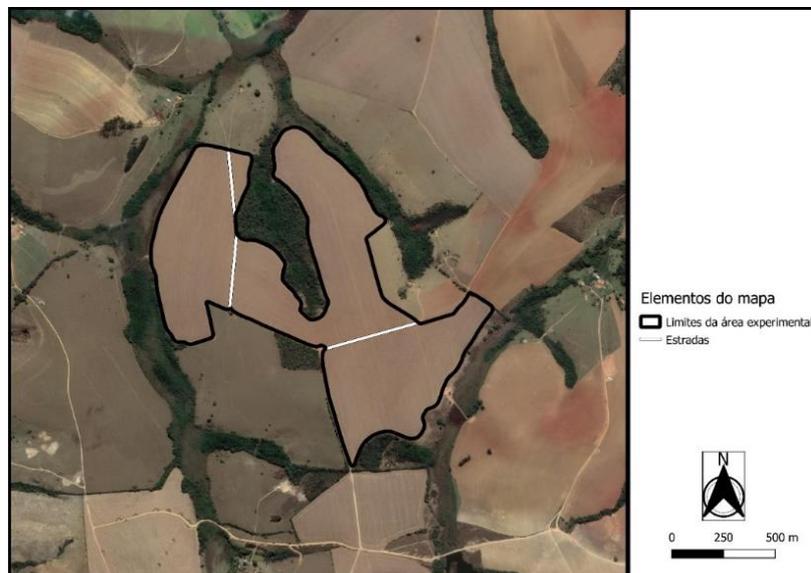
## 4 MATERIAIS E MÉTODOS

### 4.1 Área de estudo

O experimento foi conduzido na segunda safra de milho do ano de 2018 com o híbrido STATUS VIP3 plantado no dia 28 de fevereiro e colhido no dia 13 de julho, na Fazenda Retiro, localizada no município de Carrancas, em Minas Gerais. O campo experimental possui altitude média de 1.059 m e área total de 90,169 ha. Todos os tratos culturais seguiram os protocolos de manejo estabelecidos pelo produtor.

A região de Carrancas apresenta temperatura média anual de 18.8 °C, pluviosidade média anual de 1580 mm. O regime climático das estações de cultivo (safra/safrinha) apresenta estação chuvosa e calorosa no verão, onde são plantadas as culturas de safra e estação seca e temperaturas amenas no inverno, quando são plantadas as culturas de inverno. O campo experimental apresenta predomínio de Cambissolo Háplico e declividade levemente acentuada.

Figura 2 – Área de estudo

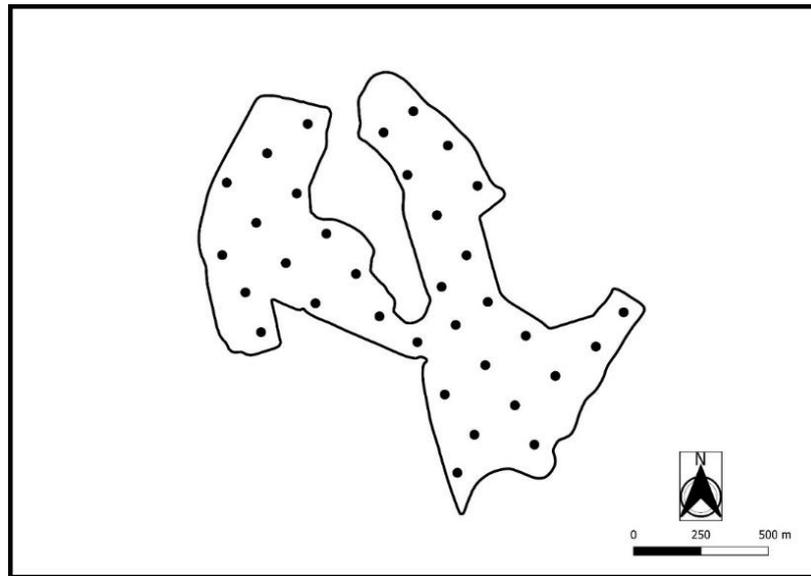


Fonte: do autor (2020)

### 4.2 Análise de fertilidade do solo

As amostras foram coletadas no dia 24/03/2019, no total 34 amostras georreferenciadas, formadas a partir de 15 subamostras (amostragem composta) nas profundidades, 0 a 10 e 10 a 20 cm. Os pontos configuram uma malha amostral de 170 × 170 m (Figura 3).

Figura 3 – Pontos amostrados de solo



Fonte: do autor (2020)

Os solos foram submetidos às análises de pH em água, matéria orgânica do solo % (M.O), cálcio  $\text{cmolc dm}^{-3}$  ( $\text{Ca}^{2+}$ ), magnésio  $\text{cmolc dm}^{-3}$  ( $\text{Mg}^{2+}$ ), potássio  $\text{cmolc dm}^{-3}$  ( $\text{K}^+$ ), Alumínio  $\text{cmolc dm}^{-3}$  ( $\text{Al}^{+3}$ ), P Mehlich-1  $\text{mg dm}^{-3}$  e P Resina  $\text{mg dm}^{-3}$ . Foram também calculados a capacidade de troca de cátions  $\text{cmolc dm}^{-3}$  potencial (T) e efetiva (t), a saturação de bases % (V%), a soma de bases  $\text{cmolc dm}^{-3}$  (SB) e as relações entre os nutrientes Ca/Mg, Ca+Mg/K, Ca/K e Mg/K.

Tabela 1 – Análise descritiva das análises de fertilidade do solo

<b>Amostras de 0 a 10 cm de profundidade</b>						
	Média	Erro padrão	CV%	Desvio padrão	Mínimo	Máximo
M.O	3,4	0,1	10%	0,4	2,5	4,0
T	6,0	0,2	19%	1,1	4,5	9,5
t	3,7	0,1	17%	0,6	2,2	4,8
P Resina	15,0	1,4	56%	8,3	2,7	44,6
P Mehlich	1,4	0,1	50%	0,7	0,3	2,9
$\text{Ca}^{2+}$	2,7	0,1	18%	0,5	1,5	3,6
$\text{Ca}^{2+}$ % na T	45,0	1,4	19%	8,4	23,7	58,1
Relação Ca+Mg/K	16,6	0,9	33%	5,5	7,6	30,7
Relação Ca/Mg	3,6	0,1	12%	0,4	2,7	4,4
$\text{Mg}^{2+}$	0,8	0,0	25%	0,2	0,4	1,1
$\text{Mg}^{2+}$ % na T	12,9	0,5	24%	3,1	5,4	17,5
$\text{K}^+$	0,2	0,0	21%	0,0	0,2	0,3
$\text{K}^+$ % na T	3,7	0,2	27%	1,0	2,0	6,1
Relação Ca/K	12,9	0,7	33%	4,2	6,0	24,0
Relação Mg/K	3,7	0,2	37%	1,4	1,6	0,9
pH	6,0	0,1	12%	0,7	4,1	7,0

V%	62,0	1,9	18%	11,2	31,2	76,7
SB	3,7	0,1	17%	0,6	2,2	4,8
S	3,3	0,4	70%	2,3	0,8	13,5
<b>Amostras de 10 a 20 cm de profundidade</b>						
	Média	Erro padrão	CV%	Desvio padrão	Mínimo	Máximo
M.O	2,98	0,08	15%	0,45	2,30	4,00
T	5,12	0,17	20%	1,01	3,96	8,81
P Resina	24,36	2,67	64%	15,55	0,10	60,10
P Mehlich	2,12	0,22	61%	1,28	0,10	5,70
Ca <sup>2+</sup>	1,36	0,06	27%	0,37	0,70	2,30
Relação Mg/K	2,15	0,14	38%	0,81	1,00	3,89
pH	5,25	0,09	10%	0,51	3,70	5,90
V%	38,29	1,56	24%	9,09	22,27	55,09
SB	1,93	0,08	25%	0,49	1,06	3,19
S	12,82	0,81	37%	4,70	2,60	23,20
Relação Mg/K	2,15	0,14	38%	0,81	1,00	3,89
pH	5,25	0,09	10%	0,51	3,70	5,90
V%	38,29	1,56	24%	9,09	22,27	55,09
SB	1,93	0,08	25%	0,49	1,06	3,19
S	12,82	0,81	37%	4,70	2,60	23,20

CV%, coeficiente de variação em %

Fonte: do autor (2020)

A estatística descritiva das análises realizadas encontra-se na Tabela 1 para as profundidades de 0 a 10 e 10 a 20 cm de profundidade, respectivamente. O objetivo desta é análise consiste em observar o comportamento dos dados para entender melhor sua amplitude além de buscar possíveis erros de execução de amostragem e laboratorial que podem gerar resultados discrepantes. No presente trabalho, as análises descritivas dos dados das análises de solo, não foram observadas nenhum tipo de erro amostral e laboratorial.

### 4.3 Mapeamento da produtividade

O mapa de produtividade da cultura do milho foi gerado a partir da colhedora Case IH 5230, equipada com o monitor de produtividade CASE IH Advanced Farming System (AFS®). O equipamento utilizado na colheita é capaz de gerar planilhas que atrelam dados de produtividade e coordenadas geográficas.

Tabela 2 - Análise descritiva dos dados de colheita do milho.

Número de pontos	Mínimo (kg ha <sup>-1</sup> )	Máximo (kg ha <sup>-1</sup> )	Coefficiente de variação (%)
<b>Antes do Tratamento</b>			
105148	0	50000	45,39
<b>Depois do Tratamento</b>			
19847	6611	13055	10,77

Fonte: do autor (2020)

Os dados de colheita foram tratados com o auxílio do programa *MAP FILTER 2.0*, que permite identificar e excluir dados que podem ser ocasionados por falhas do equipamento de colheita ou erros do operador. A análise descritiva está disponível na tabela 2.

#### 4.4 Interpolação dos dados de fertilidade do solo e produtividade

Os mapas de fertilidade do solo e produtividade, foram gerados pelo método geoestatístico Krigagem Ordinária, seguindo uma metodologia de análises e ajuste do variograma que procuraram ao máximo valores preditos os mais próximos dos reais.

A priori das análises geostatísticas, foram gerados os gráficos e índices estatísticos:

- a) Histograma, observando a assimetria dos dados e procura identificar *outliers*.
- b) Gráfico de QQPLOT, nessa análise procura ajustar os pontos sobre a linha guia que é representada por uma linha na posição de 45°. Dessa forma, é possível a escolha do método de transformação mais adequado a natureza dos dados.
- c) Análise de tendência, essa análise é feita para detectar se a existência de tendências, podendo ser elas, constante, primeiro, segundo e terceiro ordem.
- d) Determinação do tamanho do *Lag*, essa análise busca calcular os intervalos entre as lags ideais para o variogramas.

O cálculo da variância foi dado pelo modelo:

$$\gamma^*(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

onde  $N(h)$  é o número de pares de valores medidos  $Z(x_i)$ ,  $Z(x_i+h)$ , separados por um vetor  $h$  (BASTOS, 2017).

Os modelos de ajuste do variograma adotados foram o exponencial, gaussiano e esférico. Sempre escolhendo o que melhor ajustasse no variograma experimental de forma visual e na visualização dos índices oferecidos pela validação cruzada.

### Modelo exponencial

Sua equação é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0, \\ C_0 + C_1 \left(1 - e^{-\frac{h}{a}}\right), & \text{se } 0 < h < \infty. \end{cases}$$

### Modelo Esférico

Sua equação é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0, \\ C_0 + C_1 \left[ \frac{3}{2} \left(\frac{h}{a}\right) - \frac{1}{2} \left(\frac{h}{a}\right)^3 \right], & \text{se } 0 < h \leq a, \\ C_0 + C_1, & \text{se } h > a. \end{cases}$$

### Modelo Gaussiano

Sua equação é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0, \\ C_0 + C_1 \left[ 1 - e^{-\left(\frac{h}{a}\right)^2} \right], & \text{se } 0 < h < \infty. \end{cases}$$

Onde  $a$  Alcance,  $C_0$  efeito pepita,  $C_1$  contribuição e  $h$  distância entre os pares separados.

Uma alternativa para os dados que não apresentam dependência espacial foi o cálculo do inverso do quadrado da distância.

$$x_p = \frac{\sum_{i=1}^n \left(\frac{1}{d_i^2} \times x_i\right)}{\sum_{i=1}^n \left(\frac{1}{d_i^2}\right)}$$

Onde  $x_p$  valor interpolado,  $x_i$  valor do  $i$ -ésimo ponto vizinho,  $d_i$  distância entre o  $i$ -ésimo ponto vizinho e o ponto de interesse.

A validação cruzada foi utilizada no auxílio da escolha do melhor modelo a ser utilizado. Nas operações, foi observada a inclinação da curva sobre o intercepto, além do cálculo dos valores de raiz quadrada do erro médio (RQM) e erro médio.

## 4.5 Atributos obtidos por sensoriamento Remoto

### 4.5.1 Índice de vegetação

O Índice de Vegetação da Diferença Normalizada, foi calculado a partir das imagens dos satélites Landsat 8 e Sentinel 2.

Tabela 3 – Observações das imagens de satélite.

<b>Sentinel</b>		
<b>Data da Coleta</b>	<b>Dias após o plantio</b>	<b>Ciclo da cultura</b>
01/06/2018	93	135
21/06/2018	113	135
<b>Landsat8</b>		
20/06/2018	112	135

Fonte: do autor (2020)

As imagens foram obtidas nas condições apresentadas na Tabela 3. Na seleção de imagens, o objetivo foi coletar imagens datadas em dias em que a cultura já se apresentava estabelecidas, nos estádios V8 e VT.

Os índices de vegetação obtidos do satélite Sentinel 2, foram calculadas no Google Earth Engine, disponível no site, [www.earthengine.google.com](http://www.earthengine.google.com). A resolução espacial das imagens é de 10 metros e com valores de reflectância ao topo da atmosfera.<sup>1</sup>

As imagens do satélite Landsat 8 foram baixadas do site *Earth Explorer*<sup>2</sup>. Para melhorar a qualidade da refletância, elas foram tratadas com uma ferramenta do programa *Quantum GIS* chamada *Semi-Automatic Classification Plugin*. O cálculo do índice de vegetação para as imagens obtidas dos dois satélites, foram calculadas utilizando a fórmula:

<sup>1</sup> A descrição das imagens pode ser encontrada no link: <[https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS\\_S2#description](https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2#description)>.

<sup>2</sup> Consulte na íntegra: <<https://earthexplorer.usgs.gov/>>

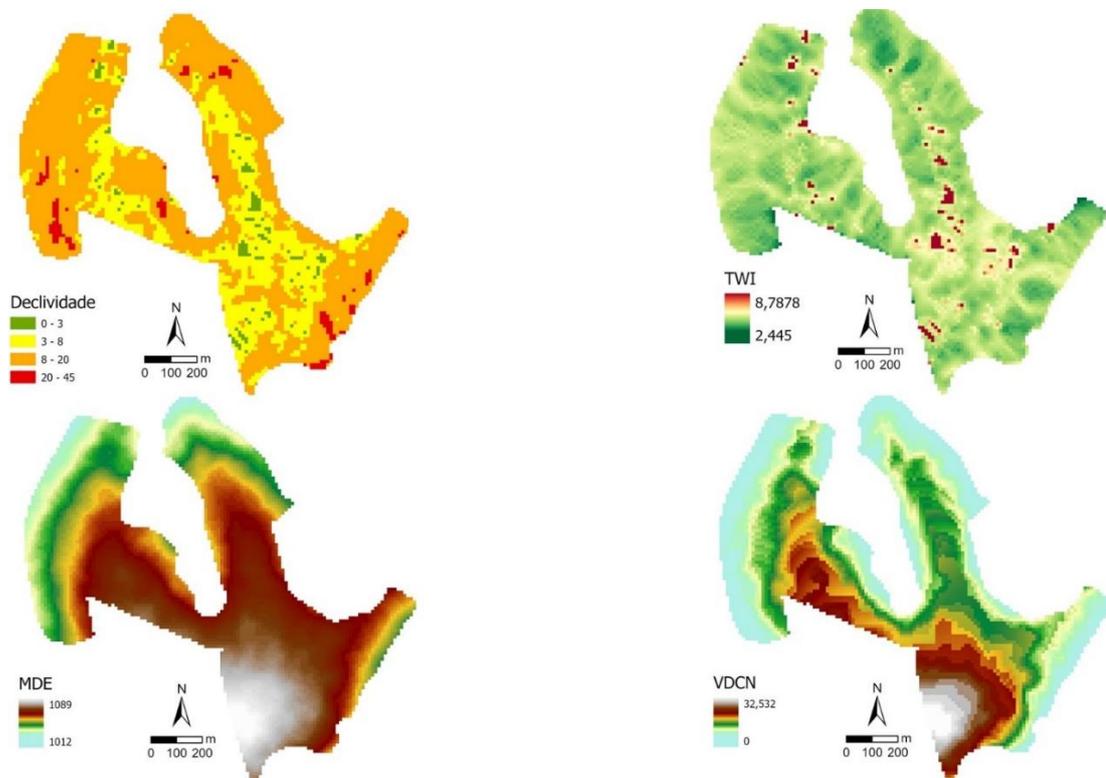
$$NDVI = \frac{NIR - RED}{NIR + RED}$$

onde NIR é a reflectância do espectro infravermelho próximo e RED a reflectância do espectro vermelho.

#### 4.5.2 Atributos do terreno

Foi utilizado o modelo digital de elevação (MDE), Alos-Palsar com resolução de 12,5 m. A partir deste, foram calculados outros índices de terreno: índice topográfico de umidade e a distância vertical a partir do nível dos canais. Para tanto, foi utilizado o software SAGA GIS (Conrad et al., 2015). Tais índices foram calculados para aumentar o número de informações a respeito do relevo, uma vez que localização dos solos em uma paisagem, ou mesmo locais de maior ou menor probabilidade de acúmulo de água, são mais bem identificados a partir destes dois índices, respectivamente.

Figura 4 – Mapas dos atributos do terreno



TWI – índice topográfico de umidade, MDE – Modelo digital topográfico, VDCN: distância vertical para a rede do canal

Fonte: do autor (2020)

#### **4.6 Mineração de dados: seleção de variáveis mais importantes**

As análises foram executadas com o objetivo de selecionar variáveis que apresentem alta correlação com a produtividade da cultura do milho, além de sua fácil obtenção e sem custo. Na seleção de variáveis para a geração de zonas de manejo, foi utilizada as análises de regressão linear múltipla e método de grupamento. Esse processo respeitou princípios utilizados na mineração de dados.

Para as associações de seleção de dados que consiste na fase de escolha de variáveis que apresentam alta significância com a produtividade das culturas estudadas, foram aplicados conceitos agronômicos, com o intuito de selecionar as variáveis disponíveis que melhor pode explicar a produtividade de grãos.

As classificações, aconteceu com a interpolação dos dados utilizando a krigagem ordinária. A classificação dos produtos gerados, utilizou se quatro classes de intervalos simétricos, de acordo os resultados da variável analisada. Esse número de intervalos foi escolhido para entender melhor os dados visualmente.

Já os dados de terreno, a declividade foi classificada de acordo com as classes de potencialidade de erosão e o modelo digital de elevação de acordo com a convenção de declividade do programa Arc GIS PRO decli1. Para os outros dados de terreno e os índices de vegetação, foi aplicada a mesma lógica de classificação dos mapas de fertilidade do solo.

O método de regressão utilizado, busca estabelecer uma função preditiva com a produtividade de grãos na área estudada. A regressão exploratória teve o objetivo de calcular um modelo de regressão e a significância de cada variável contrastada com o milho. O processo, foi executado no Arc GIS PRO 2.4.0, utilizando a ferramenta *Exploratory Regression*.

#### **4.7 Delimitação de zonas de manejo e o número de zonas**

A delimitação das zonas de manejo foi feita através do método de agrupamento, utilizando a ferramenta *Multivariate Clustering*, disponível no Arc GIS PRO 2.4.0. As variáveis utilizadas foram selecionadas através do processo de mineração de dados. Após a análise, foram selecionadas as cinco variáveis que apresentaram maior significância para a delimitação das zonas de manejo juntamente com a produtividade da cultura.

O algoritmo *K Medoids*, se deu a sua escolha por apresenta alta robustez quando os dados apresentam *outlier*, podemos levar esse fator em consideração, pois dados obtidos de fenômenos naturais podem ter a existência de amostragem imprecisa.

O número de zonas, aconteceu respeitando o índice *Pseudo F-statistic values*, que é expresso em um gráfico resultante da análise de *Multivariate Clustering*. Para a interpretação do gráfico, é necessário observar o maior valor de *Pseudo F-statistic* no eixo y do gráfico, e o valor de número de zonas no eixo x, associando os valores dos eixos x e y, podemos concluir sobre o número mais adequado de zonas de manejo.

## 5 RESULTADOS E DISCUSSÃO

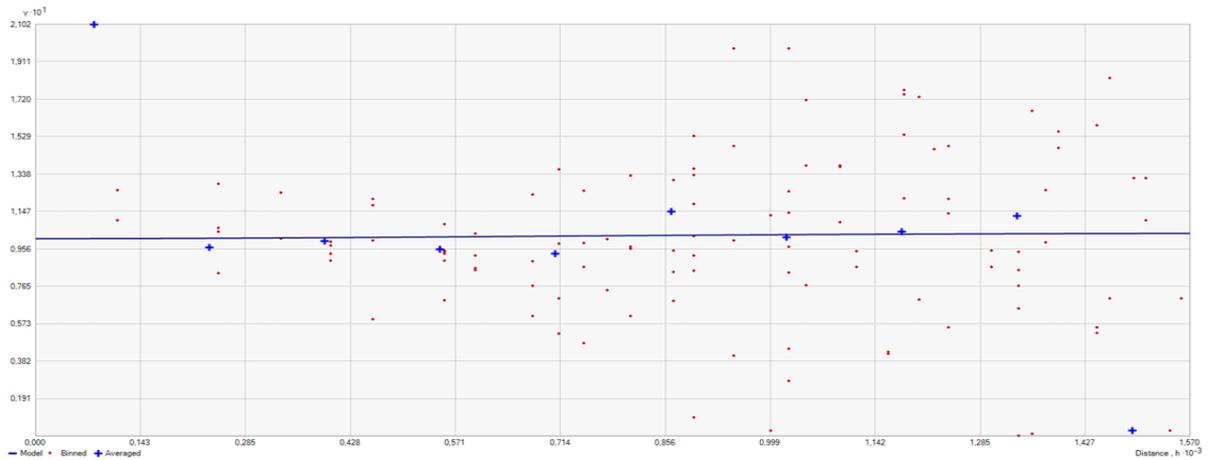
### 5.1 Acurácia da interpolação espacial

Para o delineamento das zonas de manejo, foi necessária a interpolação de dados pontuais (fertilidade e produtividade). O tamanho do *lag* foi calculado com vistas a não mascarar efeitos de correlação a curto alcance, e deste modo, 157,744676 e 2,8877 foram os valores calculados para os dados de fertilidade e produtividade, respectivamente. Na Tabela 4 é possível observar os parâmetros utilizados a priori ao ajuste dos semivariogramas, onde: a transformação do tipo Box Cox foi a mais empregada para a normalização dos dados; foram analisadas e retiradas as tendências dos dados, cujos polinomiais ajustados variaram de 1 a 3.

Tabela 4 – Análise exploratória dos dados de fertilidade do solo e produtividade.

<b>Sumário de variáveis significantes</b>		
<b>Variável</b>	<b>Transformação / Parâmetro</b>	<b>Análises de tendência</b>
P resina 0-10 cm	-	2
Relação Ca/K 0-10 cm	Box Cox / 1	1
Relação Ca+Mg/K 0-10 cm	-	2
T 0-10 cm	-	3
V% 10-20 cm	Box Cox / 2	3
Relação Ca/Mg 0-10 cm	Box Cox / 3	3
S 0-10 cm	Box Cox / 3	2
K 0-10 cm	Box Cox / 0,1	2
K% na T 10-20 cm	-	1
M.O 0-10 cm	Box Cox / 1	2
T 0-10 cm	Box Cox / 0,1	3
P resina 10-20 cm	-	1
Ca% na T 0-10 cm	Box Cox / 3	3
Relação Mg/K 0-10 cm	Box Cox / 0,1	2
Ca 0-10 cm	-	3
V% 0-10 cm	Box Cox / 5	2
P melich 0-10 cm	-	3
pH 0-10	Box Cox / 4	2
pH 10-20	-	3
Mg 0-10 cm	Box Cox / 2	3
K% na T 0-10 cm	Box Cox / 5	1
M.O 10-20 cm	Box Cox / 0,1	1
K 10-20 cm	Box Cox / 0,7	1
Produtividade milho	Box Cox / 4	2

Figura 5 – Semivariograma experimental cálcio



Fonte: do autor (2020)

Após a modelagem do semivariograma experimental, foi executada a operação de validação cruzada, onde índices estatísticos ou de validação foram gerados analisando os valores de interpolação executada *versus* a interpolação de referência (dados analisados). A Tabela 6 apresenta os índices de acurácia para os dados de fertilidade na profundidade de 0 a 10 cm, de 10 a 20 cm e dos dados de produtividade. Ao contrário dos dados de solo, os dados de produtividade apresentaram bons ajustes na validação cruzada, podem ser vistos na tabela 5. A densidade de pontos gerados pelos monitores de colheita apresenta alta densidade amostral, a validação cruzada apresentou bons resultados.

Tabela 5 - Parâmetros da validação cruzada

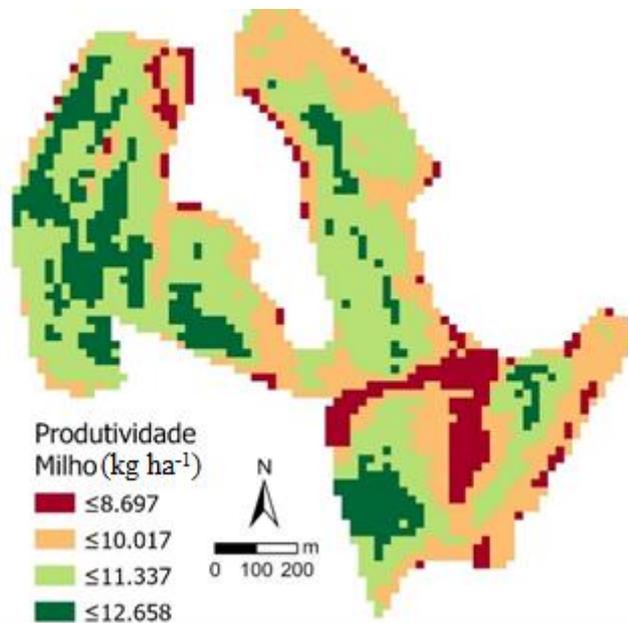
Validação Cruzada					
Variável	Erro médio	RQM	Variável	Erro médio	RQM
<b>P resina 0-10 cm</b>	0,3681	97,108	<b>Ca% na T 0-10 cm</b>	-0,6725	116,712
<b>Relação Ca/K 0-10 cm</b>	0,0915	38,690	<b>Relação Mg/K 0-10 cm</b>	0,0218	12,547
<b>Relação Ca+Mg/K 0-10 cm</b>	-0,0079	51,680	<b>Ca 0-10 cm</b>	-10,440	152,625
<b>T 0-10 cm</b>	-0,0559	12,661	<b>V% 0-10 cm</b>	-50,953	186,625
<b>V% 10-20 cm</b>	-0,4485	82,600	<b>P melich 0-10 cm</b>	0,0206	0,8187
<b>Relação Ca/Mg 0-10 cm</b>	-0,0243	0,5455	<b>pH 0-10</b>	0,0091	0,8406
<b>S 0-10 cm</b>	-0,0204	0,6991	<b>pH 10-20</b>	0,0304	10,993
<b>K 0-10 cm</b>	55,362	0,0429	<b>Mg 0-10 cm</b>	-0,0057	0,1936
<b>K% na T 10-20 cm</b>	0,1084	10,655	<b>K% na T 0-10 cm</b>	0,1084	10,655
<b>M.O 0-10 cm</b>	0,0029	0,3892	<b>M.O 10-20 cm</b>	0,0283	0,38
<b>T 0-10 cm</b>	-0,0559	12,661	<b>K 10-20 cm</b>	-0,0015	0,0428
<b>P resina 10-20 cm</b>	-10,440	152,625	<b>Produtividade milho</b>	-8,41	328,27

RQM - raiz quadrada do erro médio

Fonte: do autor (2020).

Na Figura 6 é possível observar a distribuição espacial da produtividade da cultura do milho. Levando em consideração a produtividade da região Sul de Minas, os níveis de produtividade estão de acordo com os níveis de produtividade da região, porém percebe-se alguns pontos limitantes de produção e outros com produtividade bem elevada. Em geral esses pontos limitantes estão atrelados a bordaduras. A distribuição espacial da produtividade é um importante fator para se comparar com as zonas de manejo delimitadas pelo método de agrupamento.

Figura 6 – Mapa de produtividade da cultura do milho



Fonte: do autor (2020)

## 5.2 Regressão linear múltipla: variáveis de maior significância

Para o maior entendimento dos atributos que melhor se correlacionam com a produtividade de milho, foi calculada a regressão linear múltipla, levando em consideração como variável dependente a produtividade do milho *versus* os dados de fertilidade do solo, relevo e NDVI. A Tabela 6 apresenta a ordem de significância das variáveis e seus respectivos valores positivos e negativos de contribuição. Na sequência, foi gerado um modelo de regressão, cujo coeficiente de determinação explicou 57% da variância dos dados de produtividade do milho. Por apresentarem maior significância na correlação com os dados de

produtividade, a relação Ca/K, P resina, Mg/K e o NDVI e o V% nas duas profundidades incluindo os dados de produtividade, foram utilizados para delineamento das zonas de manejo.

Tabela 6 - Variáveis analisadas para o milho (continua)

<b>Sumário de variáveis significantes</b>			
<b>Variável</b>	<b>% Significância</b>	<b>% Negativa</b>	<b>% Positiva</b>
NDVI	100,00	0,00	100,00
P resina (0-10 cm)	99,16	99,42	0,58
Relação Ca/K (0-10 cm)	94,75	96,37	3,63
Relação Ca+Mg/K (0-10 cm)	94,69	97,78	2,22
T (0-10 cm)	93,61	96,62	3,38
VDCN	91,19	4,92	95,08
V% 10-20 cm	90,51	5,83	94,17
Relação Ca/Mg 0-10 cm	89,15	6,90	93,10
S 0-10 cm	88,91	88,03	11,97
K 0-10 cm	88,77	87,72	12,28
K% na T 10-20 cm	86,16	10,30	89,70
M.O 0-10 cm	84,62	79,31	20,69
T 0-10 cm	82,63	26,85	73,15
P resina 10-20 cm	82,37	27,25	72,75
Ca% na T 0-10 cm	81,33	17,50	82,50
Relação Mg/K 0-10 cm	81,05	49,45	50,55
Ca 0-10 cm	80,97	42,02	57,98
V% 0-10 cm	79,25	56,23	43,77
P melich 0-10 cm	78,81	82,28	17,72
pH 0-10	78,30	81,33	18,67
pH 10-20	78,30	81,33	18,67
Mg 0-10 cm	78,25	50,57	49,43
K% na T 0-10 cm	78,10	42,95	57,05
M.O 10-20 cm	77,62	56,73	43,27
K 10-20 cm	75,59	22,58	77,42

Fonte: do autor (2020)

Por meio do NDVI é possível inferir o vigor das plantas de milho, onde maiores valores tendem a ocorrer onde as plantas se apresentam mais saudáveis, com menos ataque de doenças e pragas e maior níveis de clorofila em suas folhas (DEMATTE et al., 2020), conseqüentemente, tal índice se correlaciona à produtividade. Foram coletadas imagens de NDVI nos estádios de V8 e VT, sendo estes estádios correspondentes a fase de desenvolvimento vegetativo. Segundo Vian et al., (2018), valores de NDVI abaixo de 0,66 para o milho no estádio V8 apresentam baixo potencial de produção e valores entre 0,66 e 0,75 apresentam médio

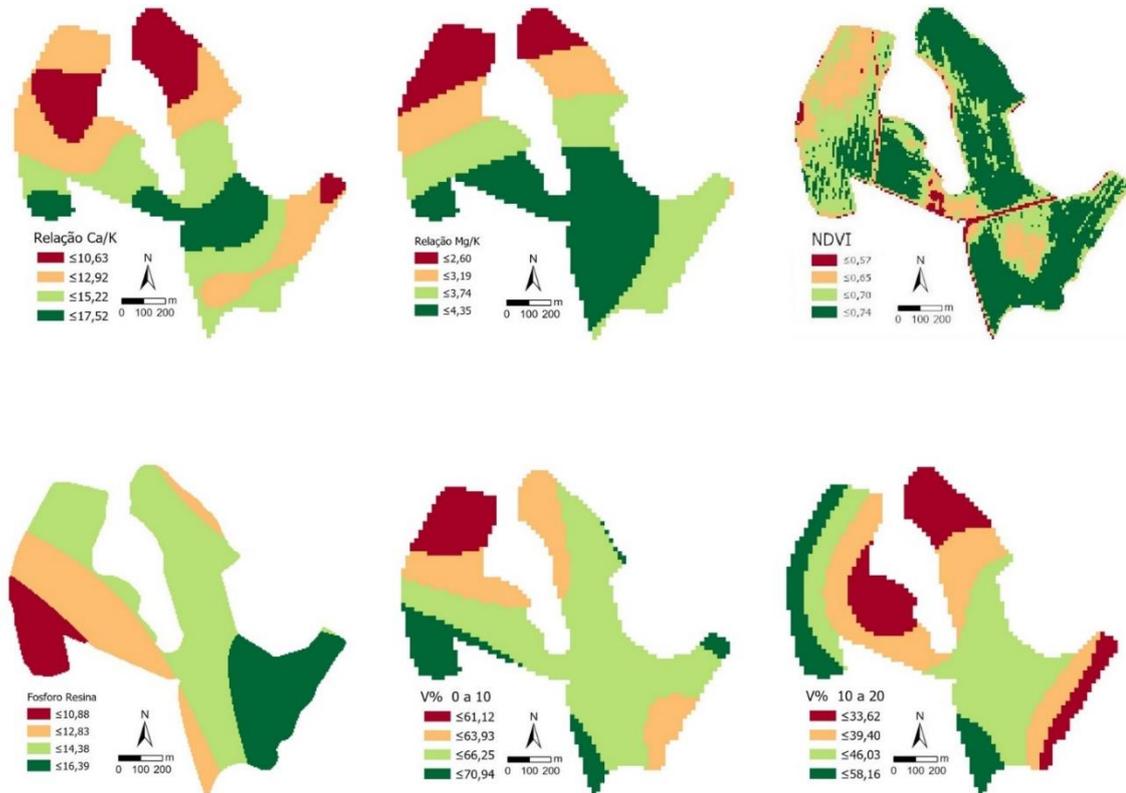
potencial de produção. Sendo assim com base nas informações do NDVI coletadas na área experimental apresenta as classes de médio e baixo potencial de produção.

A Relação Ca/K e Mg/K, representa o equilíbrio da distribuição das bases do solo na CTC potencial, segundo (SANTI et al., 2012), analisando dados de produtividade e elementos químicos do solo com métodos geostática, o elevado teor de potássio no solo é o que melhor explica a variabilidade espacial da produtividade das culturas de grãos. No presente trabalho, o K apresentou 88,77 de significância quando confrontado com a produtividade, mas a relação Ca/K e Mg/K, apresentou níveis maiores significância. Nesse sentido, com os resultados do trabalho, podemos concluir que além do K apresentar interferência na produtividade, a relação do mesmo com o Mg e Ca, também interfere na produtividade. A Relação Ca/K, cujo sua faixa considerada como adequada é de 15 a 25 (Sousa e Lobato, 2004). Observado os mapas, podemos notar em alguns locais áreas onde os teores se encontram abaixo de 12,62. Os teores no solo estão representados nas cores laranja para teores médios e vermelhas maior atenção é necessário em função de maior desequilíbrio dos nutrientes do solo. Já para a Relação Mg/K, apresenta o seu nível adequado em valores entre 5 e 15 (Sousa e Lobato, 2004). A área apresentou teores abaixo de 4,35, ou seja, toda área está fora do adequado, quando nos referimos relação Mg/K. No presente trabalho em média foi possível observar valores de 12,9 e 7,88 nas camadas de 0-10 e 10-20 cm respectivamente.

Na figura 7, podemos notar a distribuição espacial do atributo do solo P resina, O P apresenta alta capacidade de adsorção nas argilas nos solos do cerrado brasileiro, apresentando alto grau de complexidade quanto a sua disponibilidade para as plantas (RAIJ, 2011). Os níveis adequados de P resina, são de 15 a 20 em áreas de sequeiro (Sousa e Lobato, 2004). Na área de estudo, os níveis de P resina apresentam maior valor de significância negativa, os níveis se encontram fora da faixa do adequado em praticamente toda a área (Sousa e Lobato, 2004). Entretanto ao avaliar os valores médios nas camadas de 0-10 e 10-20 cm apresentam valores de 15 e 24,36 mg/dm<sup>3</sup> respectivamente.

Para o V (%) é considerado como ideal para a cultura do milho é de 70% (Sousa e Lobato, 2004), na área de estudo é observa-se que a maior parte da área se encontra abaixo do indicado para as duas profundidades analisadas, sendo na profundidade de 0 a 10 cm os valores mais próximos do adequado, isso pode ser explicado pelo fato de aplicação de calcário sem incorporação em profundidade e as raízes das culturas apresentarem maiores efeitos nas camadas superficiais do solo.

Figura 7 – Mapas dos atributos significativos na análise exploratória da cultura do milho



Fonte: do autor (2020)

### 5.3 Delineamento de zonas de manejo

A escolha do número de zonas de manejo para a cultura do milho se deu através da interpretação da Figura 7. Foi levado em consideração o valor calculado de *Pseudo F-statistic* que, quando contrastado com o eixo x, podemos observar o número ideal de 5 zonas de manejo (valor mais elevado do pseudo-F). Este método consiste em uma vantagem com relação à escolha comumente subjetiva a respeito do número de zonas de manejo a ser delimitada.

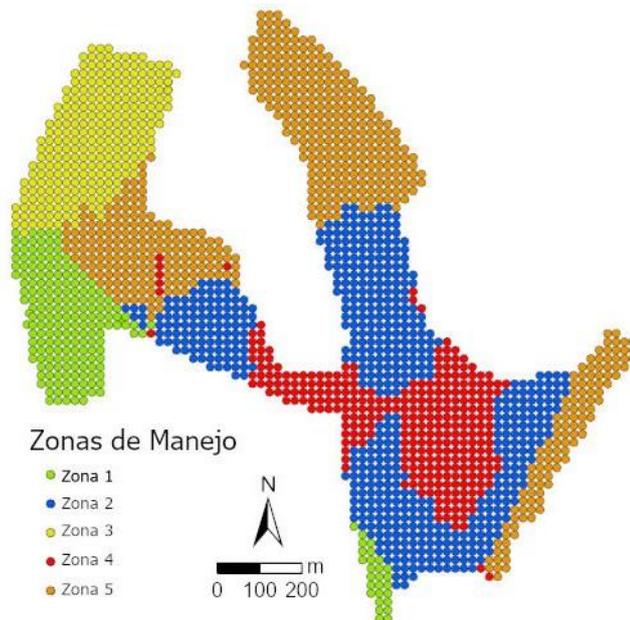
Figura 7 – Número de zonas de manejo da cultura do milho



Fonte: do autor (2020)

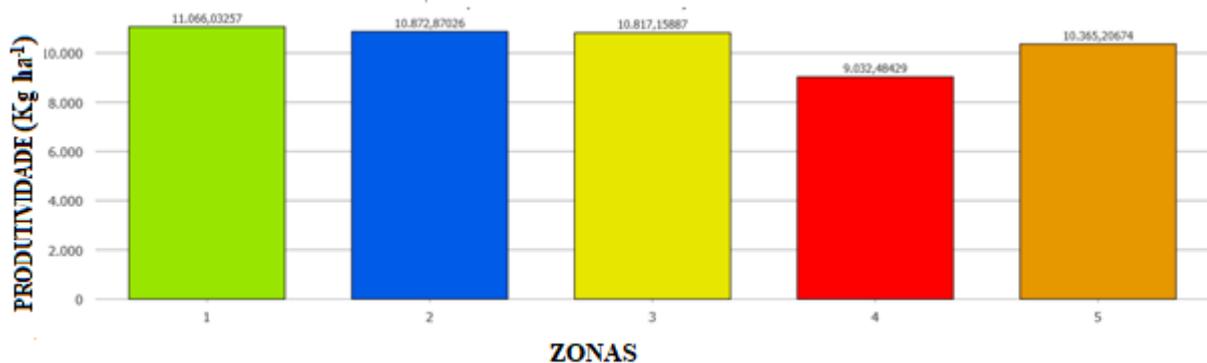
Na Figura 8 é possível observar as zonas de manejo estabelecidas a partir do algoritmo *K medoids*, as zonas de manejo apresentaram tamanhos possíveis de se manejar com a adubação a taxa variável. Além disso, podemos identificar as zonas com maior potencial de produtividade em verde e a zona com menor potencial em vermelha.

Figura 8 – Zonas de manejo da cultura do milho



Fonte: do autor (2020)

Figura 9 – Produtividade das zonas de manejo da cultura do milho



Fonte: do autor (2020)

Na figura 9 podemos observar os dados de produtividade média para cada zona de manejo estabelecida (topo das barras da Figura). Para maior eficiência do método de gestão utilizando zonas de manejo, é necessária que haja diferenças entre as médias de produtividade em cada zona. A área em estudo apresentou algumas zonas com medias distintas, porém as

zonas 2, 3 e 5 apresentaram médias bem próximas, que podem ser explicadas devido à presença de outras variáveis, onde podemos observar a distribuição distintas entre as zonas.

Através da análise descritiva (Tabela 7), é possível melhor compreender a dispersão dos dados em cada zona de manejo. O coeficiente de variação de cada zona se apresenta baixo de modo geral, ou seja, a variabilidade dos dados em relação à média é baixa para todos os dados utilizados para delimitação das zonas de manejo, incluindo a produtividade.

Quando nos referimos à zona 1, ela foi agrupada com os valores mais altos de produtividade, podemos então inferir que ela apresenta o maior potencial de produção. A zona 2 está bem próxima da zona 3 em relação a produtividade. No entanto, a zona 2 está localizada em manchas melhores dos outros atributos analisados, devido a isso ela foi gerada. Levando em consideração a comparação entre a zona 2 e 3, a zona 3 está localizada em zonas de baixo NDVI, baixo valor de V% e a baixa relação Ca/K. O local com menor potencial de produtividade está representado pela zona 4 que pode ser explicado pela presença de estradas afetando a produtividade nas bordaduras da lavoura, devido ao fato de que nas bordaduras há a possibilidade remonte durante o plantio dobrando a população de plantas, possibilidade de retiradas de milho em época de milho verde, compactação ou até mesmo a poeira da estrada são fatores que podem causar consequências na diminuição da produtividade. A zona 5, após a zona 4, é a de menor potencial de produtividade. Essa zona está localizada em mancha que apresenta baixa V% e a baixa relação Mg/K de 10 a 20 cm.

Tabela 7 – Análise descritiva das zonas de manejo (continuação)

<b>Zona 1</b>							
	<b>Produtividade do Milho (kg ha<sup>-1</sup>)</b>	<b>Relação Ca/K</b>	<b>Relação Ca/Mg</b>	<b>P Resina</b>	<b>V% 0 a 10 cm</b>	<b>V% 10 a 20 cm</b>	<b>NDVI</b>
Média	11066	13,56	3,58	10,16	67,20	46,21	0,69
Desvio padrão	548,32	1,47	0,35	1,11	1,54	5,10	0,02
Coeficiente de variação	0,05	0,11	0,10	0,11	0,02	0,11	0,04
Mínimo	9221,74	11,00	2,78	8,02	63,99	34,71	0,59
Máximo	12.330	16,18	4,19	12,63	70,73	56,66	0,74

<b>Zona 2</b>							
	<b>Produtividade do Milho (kg ha<sup>-1</sup>)</b>	<b>Relação Ca/K</b>	<b>Relação Ca/Mg</b>	<b>P Resina</b>	<b>V% 0 a 10 cm</b>	<b>V% 10 a 20 cm</b>	<b>NDVI</b>
Média	10872	13,81	3,86	13,75	64,81	40,1	0,72
Desvio padrão	697,03	1,09	0,22	1,15	0,69	4,19	0,01
Coefficiente de variação	0,06	0,08	0,06	0,08	0,01	0,1	0,02
Mínimo	9125,93	11,33	3,41	11	63,5	31,09	0,66
Máximo	12658	16,87	4,3	15,88	66,53	50,14	0,75
<b>Zona 3</b>							
	<b>Produtividade do Milho (kg ha<sup>-1</sup>)</b>	<b>Relação Ca/K</b>	<b>Relação Ca/Mg</b>	<b>P Resina</b>	<b>V% 0 a 10 cm</b>	<b>V% 10 a 20 cm</b>	<b>NDVI</b>
Média	10817	10,75	2,43	13,07	60,29	42,29	0,65
Desvio padrão	917,47	0,47	0,25	0,83	1,76	4,53	0,02
Coefficiente de variação	0,08	0,04	0,1	0,06	0,03	0,11	0,04
Mínimo	7824	9,9	1,93	10,85	56,72	34,43	0,56
Máximo	12355	11,69	2,96	14,19	64,45	50,75	0,7
<b>Zona 4</b>							
	<b>Produtividade do Milho (kg ha<sup>-1</sup>)</b>	<b>Relação Ca/K</b>	<b>Relação Ca/Mg</b>	<b>P Resina</b>	<b>V% 0 a 10 cm</b>	<b>V% 10 a 20 cm</b>	<b>NDVI</b>
Média	9900	12,41	3,42	13,62	64,29	36,02	0,69
Desvio padrão	1031,26	2,59	0,55	1,21	0,95	4,91	0,04
Coefficiente de variação	0,1	0,21	0,16	0,09	0,01	0,14	0,06

Mínimo	7418	8,33	2,08	11,02	61,06	25,47	0,42
Máximo	12306	17,53	4,33	16,24	67,1	44,87	0,75
<b>Zona 5</b>							
	<b>Produtividade do Milho (kg ha<sup>-1</sup>)</b>	<b>Relação Ca/K</b>	<b>Relação Ca/Mg</b>	<b>P Resina</b>	<b>V% 0 a 10 cm</b>	<b>V% 10 a 20 cm</b>	<b>NDVI</b>
Média	10365	10,78	3,1	13,38	64,08	33,13	0,71
Desvio padrão	833,4	1,34	0,39	1,28	1,07	3,12	0,02
Coefficiente de variação	0,08	0,12	0,12	0,1	0,02	0,09	0,03
Mínimo	8.489,28	8,33	2,08	11,02	61,06	25,47	0,61
Máximo	12306	14,38	3,79	16,24	67,1	40,88	0,75 (conclusão)

Fonte: do autor (2020)

## 6 CONCLUSÃO

As melhores variáveis para delimitação de zonas de manejo para a cultura do milho foram Relação Ca/K, Relação Ca/Mg 0 a 10 cm, P Resina 0 a 10 cm, V% 0 a 10 cm, V% 10 a 20 cm e o NDVI.

As zonas de manejo conseguiram identificar áreas com alto potencial de produtividade e áreas com baixo potencial de produtividade, também conseguiu identificar a interferência da presença de estradas afetando a produtividade nas bordaduras da lavoura, seja ela por remonte de plantio, retiradas de milho em época de milho verde, compactação causadas pelas manobras de cabeceira na operações de colheita, tratos culturais, plantio, ou até mesmo a poeira da estrada.

O agrupamento utilizando o algoritmo *K medoids*, possibilitou zonas bem definidas e de possível manejo com adubação a taxa variada.

Em questão ao tratamento de dados de produtividade, devem ser realizados estudos futuros para correções de dados de produtividade, pois mesmo utilizando o software de tratamento dos dados, alguns erros amostrais ainda continuaram presentes nos dados.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ARCGIS PRO. **How Multivariate Clustering works**. 2020. Disponível em: <[https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-multivariate-clustering-works.htm#ESRI\\_SECTION1\\_103DAE5B22924FEFA7F8B9A79D9614F7](https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-multivariate-clustering-works.htm#ESRI_SECTION1_103DAE5B22924FEFA7F8B9A79D9614F7)>. Acesso em: 7 ago. 2020.
- BASTOS, Rafael Lemos. **Métodos para seleção de modelos de semivariograma em campos aleatórios gaussianos**. 2017. Universidade Federal de Lavras, 2017.
- BRAMLEY, R. G. V. Lessons from nearly 20 years of Precision Agriculture research, development, and adoption as a guide to its appropriate application. **Crop and Pasture Science**, v. 60, n. 3, p. 197–217, 2009. DOI: 10.1071/CP08304.
- BREEMEN, Nico Van; BUURMAN, Peter. **Soil formation**. 2. ed. v. 53. Springer, 2002.
- CONAB. **Acompanhamento da Safra Brasileira de Grãos (2019/2020)**. [online]. Disponível em: <<https://www.conab.gov.br/info-agro/safra-graos/boletim-da-safra-de-graos>>. Acesso em 14 de agosto de 2020.
- DEMATTÊ, José Alexandre M.; POPPIEL, Raul Roberto; ELIZABET, Nélide. **A importância das geotecnologias na caracterização do meio físico com vistas ao aumento da produtividade**. [S. l.], 2020.
- FURTINI NETO, A. E.; VALE, F. R.; RESENDE, A. V.; GUILHERME, L. R.; GUEDES, G. A. A. FURTINI NETO, A. E. et al. **Fertilidade do solo**. Lavras: FAEPE, 2001.
- GALVÃO, Noemi Dreyer; MARIN, Heimar De Fátima. Data mining: A literature review. **ACTA Paulista de Enfermagem**, v. 22, n. 5, p. 686–690, 2009.
- GOZDOWSKI, Dariusz et al. Determination of the Most Relevant Soil Properties for the Delineation of Management Zones in Production Fields. **Communications in Soil Science and Plant Analysis**, v. 45, n. 17, p. 2289–2304, 2014.
- GREGO C. R., OLIVEIRA R. P., Vieira S. R. Geoestatística aplicada a Agricultura de Precisão. **Agricultura de precisão: Resultados de um novo olhar**, p. 74–83, 2014.
- KHOSLA, R. et al. Spatial Variation and Site-Specific Management Zones. **Geostatistical Applications for Precision Agriculture**. p. 195–219.
- KHOSLA, Rajiv. The 10th International Conference on Precision Agriculture. 1–5, 2010.
- LINSLEY, Clyde Maurice; BAUER, Frederick Charles. Test your soil for acidity. **University of Illinois, Agricultural Experiment Station**, Chicago, Illinois, n. 346, p. 2–4, 1929.
- MAPA. **Principais conceitos e expressões utilizados na agricultura de precisão** - ap., n. 61, p. 1, 2017. Disponível em: <<http://www.agricultura.gov.br/assuntos/sustentabilidade/tecnologia-agropecuaria/agricultura-de-precisao-1/arquivos-de-agricultura-de-precisao/conceitos-relacionados-a-agricultura-de-precisao>>. Acesso em 14 de agosto de 2020.
- MOLIN, José Paulo; AMARAL, Lucas Rios; COLAÇO, André Freitas. **Agricultura de precisão**. [s.l. s.n.]. v. 53
- MULLA, David; KHOSLA, Raj. **Historical Evolution and Recent Advances in Precision Farming**. [S. l.], p. 1–36, 2015. DOI: 10.1201/b18759-2.

NAVARRO JÚNIOR, Hugo Motta; COSTA, José Antonio. Expressão do potencial de rendimento de cultivares de soja. [*S. l.*], n. 1, p. 275–279, 2002.

NAWAR, Said; CORSTANJE, Ronald; HALCRO, Graham; MULLA, David; MOUAZEN, Abdul M. **Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review**. 1. ed. [s.l.] : Elsevier Inc., 2017. v. 143 DOI: 10.1016/bs.agron.2017.01.003. Disponível em: <http://dx.doi.org/10.1016/bs.agron.2017.01.003>.

NOETZOLD, Rafael et al. Variabilidade espacial e temporal de atributos químicos do solo durante cinco safras. **Revista Brasileira de Geomática**, v. 6, n. 4, p. 328, 2018.

PANTANO, Glaucia et al. SUSTENTABILIDADE NO USO DO FÓSFORO: UMA QUESTÃO DE SEGURANÇA HÍDRICA E ALIMENTAR. **Quím. Nova**, São Paulo, v. 39, n. 6, p. 732-740, July 2016. Available from <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-40422016000600732&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422016000600732&lng=en&nrm=iso)>. access on 17 Aug. 2020. <http://dx.doi.org/10.5935/0100-4042.20160086>.

PIRES, João Leonardo Fernandes et al. Discutindo Agricultura de Precisão - Aspectos Gerais. **Documentos Online** 42, p. 18, 2004. Disponível em: <<https://www.embrapa.br/trigo/busca-de-publicacoes/-/publicacao/852495/discutindo-agricultura-de-precisao---aspectos-gerais>>. Acesso em 14 de agosto 2020.

RAIJ, B. Van. **Fertilidade do solo e manejo de nutrientes**. **International Plant Nutrition Institute**, 2011.

RIBEIRO, Helena; JAIME, Patrícia Constante; VENTURA, Deisy. Alimentação e sustentabilidade. **Estudos Avancados**, v. 31, n. 89, p. 185–198, 2017.

SANTI, Antônio Luis et al. Análise de componentes principais de atributos químicos e físicos do solo limitantes à produtividade de grãos. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, p. 1346–1357, 2012.

SANTOS, Rafael. Conceitos de Mineração de Dados na Web. Sociedade Brasileira de Computação (SBC): XV SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, p. 1–40, 2009. Disponível em: <<http://www.lac.inpe.br/~rafael.santos/Docs/WebMedia/2009/webmedia2009.pdf>>.

VIANA, A. A. N. et al. **Agricultura de Precisão - Boletim Técnico**. [online]. Disponível em: <[http://www.agricultura.gov.br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/agric\\_precisao.pdf](http://www.agricultura.gov.br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/agric_precisao.pdf)>. Acesso em 14 de agosto de 2020.

VRINDTS, E. et al. Management zones based on correlation between soil compaction, yield and crop data. **Biosystems Engineering**, v. 92, n. 4, p. 419–428, 2005.