

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

ANTÔNIO DAVID VINISKI

**AVALIAÇÃO DA EFICIÊNCIA DO USO DA MINERAÇÃO DE DADOS CLÁSSICA
E ESPACIAL NA ESTIMATIVA DE PRODUTIVIDADE DE GRÃOS EM IMAGENS
OBTIDAS POR MEIO DE AERONAVE REMOTAMENTE PILOTADA**

PONTA GROSSA
2018

ANTÔNIO DAVID VINISKI

**AVALIAÇÃO DA EFICIÊNCIA DO USO DA MINERAÇÃO DE DADOS CLÁSSICA
E ESPACIAL NA ESTIMATIVA DE PRODUTIVIDADE DE GRÃOS EM IMAGENS
OBTIDAS POR MEIO DE AERONAVE REMOTAMENTE PILOTADA**

Dissertação apresentada ao Programa de Pós-Graduação
em Computação Aplicada da Universidade Estadual de
Ponta Grossa, como requisito parcial para obtenção do
título de mestre.

Orientação: Prof^a. Dr.^a Alaine Margarete Guimarães

Coorientador: Prof. Dr. Eduardo Fávero Caires

PONTA GROSSA
2018

Ficha Catalográfica
Elaborada pelo Setor de Tratamento da Informação BICEN/UEPG

V785 Viniski, Antônio David
Avaliação da eficiência do uso da mineração de dados clássica e espacial na estimativa de produtividade de grãos em imagens obtidas por meio de aeronave remotamente pilotada/ Antônio David Viniski. Ponta Grossa, 2018.
64f.

Dissertação (Mestrado em Computação Aplicada - Área de Concentração: Computação para Tecnologias em Agricultura), Universidade Estadual de Ponta Grossa.

Orientadora: Prof^a Dr^a Elaine Margarete Guimarães.

Coorientador: Prof. Dr. Eduardo Fávero Caires.

1.Mineração de dados espaciais.
2.Sensoriamento remoto. 3.Aprendizado de máquina. 4.Resolução espacial. 5.Drone.
I.Guimarães, Elaine Margarete. II. Caires, Eduardo Fávero. III. Universidade Estadual de Ponta Grossa. Mestrado em Computação Aplicada. IV. T.

CDD: 006.312

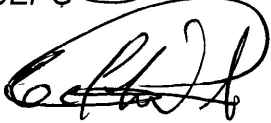
TERMO DE APROVAÇÃO

Antônio David Viniski

**“AVALIAÇÃO DA EFICIÊNCIA DO USO DA MINERAÇÃO DE
DADOS CLÁSSICA E ESPACIAL NA ESTIMATIVA DE PRODUTIVIDADE
DE GRÃOS EM IMAGENS OBTIDAS POR MEIO DE AERONAVE
REMOTAMENTE PILOTADA”**

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:


Prof. Dr.ª Elaine Margarete Guimarães
UEPG


Dr. Helio Antonio Wood Joris
FUNDAÇÃO ABC/Castro-PR


Prof. Dr. José Carlos Ferreira da Rocha
UEPG

Ponta Grossa, 16 de março de 2018.

DEDICATÓRIA

Dedico este trabalho aos meus pais, minha irmã, amigos e familiares

AGRADECIMENTOS

Agradeço primeiramente a minha família, em especial, meus pais, que estiveram sempre me apoiando em todas as minhas decisões, me incentivando a ser uma pessoa melhor e lutar pelos meus objetivos, sempre com muito amor, carinho e afeto. A minha irmã que compartilhou de muitos momentos comigo, não me deixando desistir, me fazendo sempre persistente e batalhador. Aos demais parentes e amigos que sempre torceram por mim.

Agradeço imensamente a minha orientadora Elaine Margarete Guimarães, por acreditar em mim e no meu potencial, sempre me incentivando a estudar, buscar soluções e continuar acreditando perante todas as dificuldades, sou muito grato por ter a oportunidade e honra de ser seu orientado.

Aos meus colegas de laboratório Alessandro Stachak e David Zampier, que não mediam esforços em me ajudar nos experimentos, saídas de campo, enfrentando as dificuldades da pesquisa junto comigo. E também aos demais amigos, Luiz Rodolfo, Giancarlo Rodrigues, Fábio Santos, dentre outros colegas de mestrado e graduação, que compartilharam, não só dos momentos de dificuldades, mas também dos cafés da tarde, aniversários, churrascos dentre outras atividades que faziam dos dias mais divertidos e menos desgastantes.

Aos proprietários e colaboradores da Fazenda Paiquerê, em especial ao gerente Allison Fornari, que permitiu o acompanhamento das áreas experimentais, sempre disposto a sanar as dúvidas e informações necessárias sobre as mesmas.

Aos meus vizinhos, ou melhor, amigos de condomínio, Regina, Joice, Rafael, Anderson, Alan, e todos os demais, que foram minha segunda família durante esses dois anos de mestrado.

A Capes, pelo financiamento da pesquisa por meio da bolsa de estudos concedida.

EPÍGRAFE

“Determinação, coragem e autoconfiança são fatores decisivos para o sucesso. Se estamos possuídos por uma inabalável determinação, conseguiremos superá-los. Independentemente das circunstâncias, devemos ser sempre humildes, recatados e despidos de orgulho”

(Dalai Lama)

RESUMO

O sensoriamento remoto agrícola tem fornecido um volumoso conjunto de dados espaciais, os quais podem ser utilizados em diferentes segmentos, como na estimativa de produtividade de grãos. Dentre as tecnologias empregadas no SR, a utilização de aeronaves remotamente pilotadas (RPA) na agricultura vêm crescendo, sendo uma alternativa na obtenção de dados para a estimativa de produtividade. Porém, esses conjuntos de dados gerados demandam métodos e técnicas capazes de extrair informações úteis e relevantes dos mesmos. Algumas técnicas de geoestatística, como a krigagem, têm sido empregadas, mas a utilização da mineração de dados (MD), assim como da mineração de dados espaciais (MDE), podem ser alternativas viáveis para suprir essa demanda. Este trabalho teve como objetivo avaliar o uso de técnicas de MD e MDE na estimativa da produtividade de grãos de soja e trigo, utilizando dados de imagens obtidas por meio de RPA. A área de estudo localiza-se no município de Piraí do Sul, Paraná. Foi utilizada uma RPA de asa fixa para o acompanhamento das culturas de soja e trigo. No imageamento do trigo foram utilizadas duas câmeras, uma com a captura de imagens no espectro visível (RGB), e outra no infravermelho próximo (NIR), tendo sendo analisadas também as resoluções espaciais de 10 e 20 cm/pixel para cada câmera. Para a soja apenas a câmera RGB foi utilizada e as resoluções espaciais sobrevoadas foram 10, 20 e 26 cm/pixel. Os dados do atributo meta, a produtividade das culturas, foram obtidos por meio de colhedoras de precisão. Os atributos de predição, correspondendo aos valores das bandas espectrais e altitude do terreno, foram submetidos aos algoritmos de MD empregando as técnicas de regressão linear múltipla (RLM), redes neurais artificiais (RNA) e máquina de vetores de suporte para regressão (SVR). Para a MDE, foi utilizado o modelo aditivo generalizado (GAM). Para fins de comparação, os dados foram também analisados pelo método tradicional de krigagem. As técnicas foram testadas considerando duas abordagens principais: (i) utilizando apenas as bandas espectrais para estimativa e, (ii) utilizando as bandas espectrais e os valores de altitude do terreno. Para a MD clássica, os melhores resultados foram obtidos com a técnica SVR, utilizando o kernel Laplacian. Na MDE, o método GAM com a função de ajuste gaussiana apresentou os melhores resultados. Tanto para as técnicas clássicas de MD como para a MDE, a incorporação da altitude nos modelos de regressão possibilitou aumento considerável nos coeficientes de correlação e determinação, com consequente diminuição no erro (RMSE). Os valores de correlação obtidos com a MDE foram semelhantes aos obtidos com o método de krigagem, porém a MDE foi mais eficiente em avaliar o impacto dos atributos de predição (valores das bandas espectrais e altitude) na estimativa do atributo meta. Com isso, conclui-se que a MDE mostra-se viável de ser utilizada como ferramenta na geração de modelos para estimativa de produtividade de grãos com base em dados de imagens de RPA.

Palavras-chave: Mineração de dados espaciais, sensoriamento remoto, aprendizado de máquina, resolução espacial, drone, krigagem.

ABSTRACT

Agricultural remote sensing (RS) has provided a massive set of spatial data which can be used in different segments, such as in grain yield estimation. Among the technologies applied in RS, the use of remotely piloted aircraft (RPA) in agriculture is growing as an alternative to obtain data for estimating productivity. However, these generated data sets require methods and techniques capable of extracting useful and relevant information from them. Some geostatistics techniques have been applied, such as kriging, but the use of data mining (DM) as well as spatial data mining (SDM) can be viable alternatives to meet that demand. The goal of this work was to evaluate the use of DM and SDM techniques for estimating soybean and wheat grain yield using image data obtained by RPA. The study area is located in Pirai do Sul, Paraná State. A fixed wing RPA was used to monitor soybean and wheat crops. In wheat crop imaging two cameras were used, one to capture images in the visible spectrum (RGB), and the other one using the near infrared (NIR) spectrum. Also, it was analyzed the spatial resolutions of 10 and 20 cm / pixel for each camera. For soybean only the RGB camera was used and the overhead spatial resolutions were 10, 20 and 26 cm / pixel. The goal attribute data (crop yield), was obtained by precision harvester. The prediction attributes, corresponding to the values of spectral bands and terrain altitude, were submitted to DM algorithms using the multiple linear regression (MLR), artificial neural networks (ANN) and support vector regression (SVR) techniques. For SDM, the generalized additive model (GAM) was used. For comparison purposes, data were also analyzed by the traditional kriging method. The techniques were tested using two main approaches: (i) using only spectral bands for estimation and, (ii) using spectral bands and terrain altitude values. For classical DM, the best results were obtained with SVR technique, using the Laplacian kernel. The GAM method with the Gaussian fit function presented the best results for SDM. For both classical DM and SDM techniques, adding altitude in the regression models allowed a considerable increase in correlation and determination coefficients, with consequent decrease in error (RMSE). The correlation values obtained with SDM were similar to those obtained with kriging method, but SDM was more efficient in evaluating the impact of the prediction attributes (spectral bands and altitude) in the estimation of the goal attribute. Thus, it is concluded that SDM can be useful as a tool for estimating grain yield based on RPA image data.

Keywords: spatial data mining, remote sensing, machine learning, spatial resolution, drone, kriging.

LISTA DE FIGURAS

Figura 1 – Categorias de RPAs disponíveis no mercado para imageamento aéreo.	26
Figura 2 - Área sobrevoada nas safras de trigo 2016 e soja 2016-2017.....	31
Figura 3 - Sequência da metodologia empregada na obtenção dos modelos para a estimativa da produtividade.....	32
Figura 4 – RPA eBee da empresa SenseFly utilizada no imageamento da área de estudo	33
Figura 5 - Câmeras utilizadas no imageamento da área de estudo.....	33
Figura 6 - Software <i>Emotion 2</i> , no qual é realizada a configuração e o acompanhamento do voo da RPA.....	34
Figura 7 - GCP posicionado antes do imageamento da área de estudo.....	35
Figura 8 – Locais de posicionamentos dos GCPs nas áreas de trigo 2016 (a) e soja 2016-2017 (b).	35
Figura 9 - Receptores GNSS utilizados na leitura das coordenadas geográficas a) GNSS Trimble R4 RTK e coletor de dados Juno 5D b) GNSS Trimble ProXT e o coletor de dados Juno SB.	36
Figura 10 – Pontos de coleta das imagens pela PRA, nos voos realizados na safra de soja 2016-2016, no dia 20/02/2017, nas resoluções espaciais de 10 (a), 20 (b) e 26 (c) cm/pixel.	37
Figura 11 – Nuvem de pontos gerada pelo software Pix4D.....	38
Figura 12 - Algoritmo utilizado para o georreferenciamento, o qual está disponível no software QGIS.....	39
Figura 13 - Mapas de produtividade da safra de trigo 2016 (a) e de soja 2016-2017 (b), com valores em t/ha.....	40
Figura 14 – Sequência de etapas para extração dos dados de refletância espectral, altitude e produtividade. Todas as etapas são realizadas no software QGIS.	41
Figura 15 - Sequência de etapas para obtenção de dados de produtividade, bandas espectrais e altitude, geração da base de dados e execução dos algoritmos de mineração de dados.....	42

LISTA DE QUADROS

Quadro 1 – Alguns dos kernels para a SVR disponíveis no pacote Kernlab do software R	18
Quadro 2 - Modelos de regressão espaciais mais comuns na literatura, nos quais a autocorrelação espacial é permitida e considerada na descoberta de padrões presentes nos dados.....	21
Quadro 3 - Resumo de trabalhos com objetivo de estimar a produtividade de diferentes culturas e com técnicas distintas para geração de modelos de estimativa.	23
Quadro 4 – Algumas aplicações de RPAs na agricultura, elencando os métodos tradicionais e como podem ser realizados com esta plataforma (adaptado de Sankaran et al. (2015).	27
Quadro 5 - Principais SIGs utilizados em pesquisas relacionadas a análise da cobertura terrestre	30
Quadro 6 - Especificações das câmeras RGB e NIR utilizadas para obtenção das imagens em campo	33
Quadro 7 - Script para definição da pasta de trabalho contendo as bases de dados e para a divisão da base em conjuntos treinamento e teste.	43
Quadro 8 - Fórmulas utilizadas na MD clássica para a estimativa da produtividade com os dados das câmeras RGB e NIR, com e sem a utilização da altitude	44
Quadro 9 - Comandos e parâmetros para execução das técnicas de MD clássicas no software R	44
Quadro 10 - Fórmulas utilizadas no método GAM	46
Quadro 11 - Script para execução do algoritmo GAM.....	46
Quadro 12 - Script para execução do algoritmo de Krigagem com método Gstat.....	47

LISTA DE TABELAS

Tabela 1 – Informações sobre os voos realizados nas safras de trigo (2016) e soja (2016-2017).	36
Tabela 2 – Estatística descritiva dos dados da cultura de trigo safra 2016 e dos valores espectrais para as câmeras e resoluções espaciais utilizadas	49
Tabela 3 - Resultados da execução das técnicas de MD clássicas para a estimativa da produtividade do trigo por meio dos dados obtidos pela câmera RGB e NIR, nas resoluções espaciais de 10 e 20 cm/pixel	50
Tabela 4 - Resultados obtidos com a execução das técnicas de MDE e krigagem para a estimativa da produtividade do trigo por meio dos dados obtidos com as câmeras RGB e NIR	52
Tabela 5 - Estatística descritiva dos dados da cultura de soja safra 2016-2017 e dos valores espectrais para as câmeras e resoluções espaciais utilizadas.....	53
Tabela 6 - Resultados da execução das técnicas de mineração de dados clássicas para a estimativa de produtividade da soja utilizando a câmera RGB	53
Tabela 7 - Resultados obtidos pelos métodos de MDE e de krigagem para a estimativa da produtividade da soja utilizando a câmera RGB	54
Tabela 8 - Resumo dos melhores resultados obtidos para a estimativa de produtividade das culturas de trigo e soja.	55

LISTA DE SIGLAS

AP	Agricultura de precisão
CAR	Modelo autoregressivo condicional (do inglês <i>Conditional Autoregressive Models</i>)
GAM	Modelo aditivo generalizado (do inglês <i>Generalized Additive Model</i>)
GCP	Ponto de controle no terreno (do inglês <i>Ground Control Point</i>)
GNSS	Sistemas globais de navegação por satélite (do inglês <i>Global Navigation Satellite Systems</i>)
GPS	Sistema de Posicionamento Global (do inglês <i>Global Positioning System</i>)
IV	Índice de vegetação
KDD	Descoberta de conhecimento em bases de dados (do inglês <i>Knowledge Discovery in Databases</i>)
MD	Mineração de dados
MDE	Mineração de dados espaciais
NIR	Infravermelho próximo (do inglês <i>Near Infrared</i>)
RBF	Função de base radial (do inglês <i>Radial Basis Function</i>)
RGB	Vermelho, verde e azul (do inglês <i>Red, Green e Blue</i>)
RLM	Regressão linear múltipla
RMSE	Raiz do erro quadrático médio (do inglês <i>Root Mean Square Error</i>).
RNA	Rede neural artificial
RPA	Aeronave pilotada remotamente (do inglês <i>Remotely Piloted Aircraft</i>)
RPAS	Sistema aéreo pilotado remotamente (do inglês <i>Remote Pilot Aircraft System</i>)
SAR	Modelo autoregressivo espacial (do inglês <i>Spatial Autorregressive Model</i>)
SIG	Sistema de informações geográficas
SR	Sensoriamento Remoto
SVM	Máquina de vetores de suporte (do inglês <i>Support Vector Machine</i>)
SVR	Vetores de suporte para regressão (do inglês <i>Support Vector Regression</i>)
UAV	Veículo aéreo não tripulado (do inglês <i>Unmanned aerial vehicle</i>).
VANT	Veículo aéreo não tripulado

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS	16
2.1	OBJETIVO GERAL	16
2.2	OBJETIVOS ESPECÍFICOS	16
3	REVISÃO BIBLIOGRÁFICA	17
3.1	MINERAÇÃO DE DADOS	17
3.2	MINERAÇÃO DE DADOS ESPACIAIS	19
3.3	ESTIMATIVA DE PRODUTIVIDADE DE GRÃOS	23
3.4	SENSORIAMENTO REMOTO NA AGRICULTURA	24
3.4.1	<i>RPAS na agricultura</i>	25
3.4.2	<i>Pré-Processamento das imagens obtidas por meio sensoriamento remoto</i>	28
3.4.3	<i>Sistemas de Informação Geográfica</i>	29
4	MATERIAL E MÉTODOS	31
4.1	ÁREA EXPERIMENTAL	31
4.2	AQUISIÇÃO DAS IMAGENS EM CAMPO	32
4.2.1	<i>Geração do plano de voo</i>	34
4.2.2	<i>Posicionamento dos GCPs</i>	34
4.2.3	<i>Imageamento</i>	36
4.3	GERAÇÃO DOS MAPAS DE REFLETÂNCIA	37
4.3.1	<i>Correção do georreferenciamento dos mapas de refletância</i>	38
4.4	OBTENÇÃO DOS DADOS DE PRODUTIVIDADE EM CAMPO	39
4.5	EXTRAÇÃO DOS DADOS DOS MAPAS DE REFLETÂNCIA E MAPAS DE PRODUTIVIDADE	40
4.6	MINERAÇÃO DE DADOS	43
4.6.1	<i>Execução da Mineração de dados Clássica</i>	43
4.6.2	<i>Execução da Mineração de dados Espaciais</i>	45
4.7	ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS	47
5	RESULTADOS	49
5.1	RESULTADOS OBTIDOS NA ESTIMATIVA DA PRODUTIVIDADE DA CULTURA DE TRIGO	49
5.2	RESULTADOS OBTIDOS NA ESTIMATIVA DA PRODUTIVIDADE DA CULTURA DA SOJA	52
5.3	DISCUSSÃO DOS RESULTADOS	55
6	CONCLUSÃO	58
	REFERÊNCIAS	60

1 INTRODUÇÃO

Estimar safras agrícolas com precisão e antecipação é uma atividade complexa, em função da diversidade de cultivares encontradas no mercado, diferentes tipos de manejo, tipos de solos ou clima encontrados. Vários fatores referentes ao solo, à planta e à atmosfera interagem entre si determinando a produtividade das culturas agrícolas (ARAUJO; SOUZA; TSUKAHARA, 2011). Diferentes conjuntos de dados podem ser utilizados para a estimativa de produtividade agrícola, sendo os dados fornecidos pelo sensoriamento remoto frequentemente utilizados, por possibilitarem observações repetidas das culturas e permitirem a análise em diferentes estádios fenológicos. Os modelos de estimativa gerados pelos dados de sensoriamento remoto têm sido apontados como técnica alternativa à atual metodologia para elaboração de mapas de produtividade, por meio de colhedoras instrumentadas (ARAUJO; VETTORAZZI; MOLIN, 2005).

As atuais técnicas de aquisição dados, como o sensoriamento remoto de alta resolução, os sistemas globais de navegação por satélite (GNSS – *Global Navigation Satellite Systems*) e de informação geográfica (SIGs) e as aeronaves remotamente pilotadas (RPAs – *Remote Pilot Aircraft System*), tem permitido a aquisição de volumosos conjuntos de dados espaciais. Existe uma demanda por métodos eficazes e eficientes para extrair informações relevantes e potencialmente úteis desses conjuntos de dados. Alguns métodos tradicionais da geoestatística, como a krigagem (frequentemente presente nos SIGs), possibilitam a análise de dados espaciais de forma rápida, no entanto, apresentam limitações dependendo do conjunto de dados sendo estudado.

A mineração de dados (MD), cujo objetivo é a obtenção de padrões novos, válidos, úteis e compreensíveis, em grandes conjuntos de dados (FAYYAD; UTHURUSAMY, 1996), se torna uma alternativa para exploração desses volumosos conjuntos de dados espaciais. No entanto, segundo Vatsavai et al. (2012), muitos algoritmos clássicos de MD, ou seja, aqueles que não consideram a dimensão espacial dos dados na descoberta de padrões, supõem que as amostras de aprendizagem são independentes e identicamente distribuídas. Esta suposição é violada no caso de dados espaciais, devido à autocorrelação espacial (influência de vizinhança), para os quais, algoritmos clássicos de MD produzem modelos com baixa acurácia na previsão. Para enfrentar esse desafio, a mineração de dados espaciais (MDE) emergiu como um campo de pesquisa ativa, com foco no desenvolvimento da teoria, metodologia e prática, para extração de conhecimento e informações úteis de dados espaciais maciços e complexos. A MDE é uma

extensão da MD, cujos algoritmos analisam a influência da dependência espacial dos dados na descoberta de padrões (MENNIS; GUO, 2009).

Com isso, a utilização da MDE se mostra adequada em contextos agrícolas, devido ao enorme conjunto de dados espaciais coletados das culturas periodicamente, podendo auxiliar na geração de diferentes modelos, como de estimativa da produtividade. Essa estimativa geralmente é realizada por meio de algoritmos de regressão clássicos, os quais não consideram a dimensão espacial dos atributos utilizados na estimativa (ALGANCI et al., 2014; ARAÚJO; VETTORAZZI; MOLIN, 2005; HOLZMAN; RIVAS; PICCOLO, 2014; PICOLI et al., 2009). A produtividade, por sua vez, varia conforme se altera a posição/localização geográfica da área cultivada, pois as características físico-químicas do terreno, a incidência de doenças, plantas daninhas e pragas não se distribuem de maneira uniforme. Essa variabilidade ocorre nas diferentes culturas, entre elas o trigo e a soja, as quais se destacam regionalmente, sendo uma cultura de inverno e outra de verão, respectivamente. Diferentes metodologias são aplicadas para a estimativa de produtividade de culturas agrícolas, no entanto, a maioria delas fazem uso de imagens de satélite, sendo realizadas estimativas em grandes extensões territoriais, tanto regionais, quanto nacionais. As estimativas utilizando RPAs vêm se desenvolvendo nos últimos anos, mas ainda existem poucos trabalhos publicados com esta finalidade.

A busca por métodos mais simples e econômicos para estimativa de produtividade dessas culturas levou a pesquisas sobre o uso de RPAs para tal finalidade. Essas aeronaves proporcionam baixo custo de aquisição de dados, com alta resolução espacial e temporal. O uso de RPAs para estimativa de produtividades envolve um conjunto de etapas nas quais tecnologias de sensoriamento remoto, processamento e análise de imagens, mineração de dados, gerenciamento de dados geográficos e detecção de objetos são necessárias. Essas tecnologias ainda precisam ser adaptadas e combinadas, de modo a permitir a geração de modelos de estimativas mais precisos.

As RPAs permitem a utilização de diferentes tipos de câmeras, as quais podem fornecer dados de várias faixas espectrais, como do espectro visível (RGB – do inglês *Red*, *Green* e *Blue*) e infravermelho próximo (NIR – do inglês *Near Infrared*). No entanto, não se sabe quais regiões do espectro fornecem dados com capacidade de estimativa superior. Da mesma forma, como não existem muitos estudos envolvendo RPAs para estimativa de produtividade agrícola, os dados são coletados com diferentes resoluções espaciais, assim, não se tem um protocolo definido da utilização desta tecnologia para esta finalidade. Com isso, testar diferentes câmeras e resoluções espaciais torna-se importante para avaliar configurações mais adequadas para a estimativa de produtividade.

Além das informações espectrais disponibilizadas pelas câmeras, as RPAs também possibilitam a geração de modelos digitais de superfície (DSM – do inglês *Digital Surface Model*) e modelos digitais do terreno (DTM – do inglês *Digital Terrain Model*). Estes modelos fornecem novas informações sobre as culturas, como a altura da planta e altitude do terreno, as quais, juntamente com as informações espectrais extraídas das imagens, podem ser utilizadas na estimativa de produtividade de grãos, melhorando a capacidade de estimativa dos modelos.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Avaliar a eficiência da mineração de dados espaciais na estimativa de produtividade de trigo e soja a partir de imagens obtidas por meio de RPA.

2.2 OBJETIVOS ESPECÍFICOS

- Mensurar o desempenho de técnicas de MD clássica e espacial para estimativa de produtividade de trigo e soja em imagens de RPA
- Avaliar o desempenho do uso de câmeras RGB e NIR na estimativa de produtividade de trigo e soja utilizando imagens de RPA
- Analisar a influência da resolução espacial na estimativa de produtividade de trigo e soja utilizando imagens de RPA, considerando resoluções entre 10 e 26 cm/px.
- Verificar o impacto da incorporação da altitude do terreno na estimativa de produtividade de trigo e soja utilizando imagens de RPA
- Comparar, utilizando medidas estatísticas, como coeficiente de correlação (r), coeficiente de determinação (R^2) e raiz do erro quadrático médio (RMSE), os resultados obtidos com a MD e MDE com os resultados obtidos pela técnica krigagem.

3 REVISÃO BIBLIOGRÁFICA

3.1 MINERAÇÃO DE DADOS

A MD é o processo de exploração de grandes quantidades de dados, com o objetivo de identificação de anomalias, correlações e novos padrões válidos, úteis e compreensíveis, para suportar e auxiliar a tomada de decisões e proporcionar vantagens estratégicas (FAYYAD; UTHURUSAMY, 1996). A MD, para alguns autores, é considerada uma etapa do processo de descoberta de conhecimento em bases de dados (KDD – do inglês *Knowledge Discovery in Databases*). Alguns algoritmos de MD, no entanto, englobam todas as etapas da KDD, neste contexto, o termo MD considera todo o processo de descoberta de conhecimento.

Na agricultura, existem várias aplicações das técnicas de MD, dentre elas, a classificação de solos, plantas e culturas, bem como a previsão de condições climáticas (MUCHERINO; PAPAJOJGI; PARDALOS, 2009). Diferentes técnicas de MD são aplicadas na resolução de problemas agrícolas variados, sendo a regressão linear múltipla (RLM) uma das técnicas clássicas mais utilizadas (BAILEY; GATRELL, 1995; VATSAVAI et al., 2012).

A vantagem da RLM está no fato de gerar um modelo simples de ser aplicado, enquanto a SVR e RNA geram modelos complexos, geralmente incorporados em softwares para poderem ser reproduzidos posteriormente. No entanto, a RLM traz bons resultados apenas quando a variação dos conjuntos de dados ocorre de forma linear, diferente da SVR e RNA, que permitem a descoberta de padrões em dados com diferentes tipos de comportamentos.

A máquina de vetores de suporte (SVM do inglês *Support Vector Machine*), assim como a máquina de vetores de suporte para regressão (SVR – do inglês *Support Vector Regression*) e as redes neurais artificiais (RNAs) tem sido fortemente utilizadas em diferentes estimativas agrícolas, como na estimativa de produtividade (KUWATA; SHIBASAKI, 2016), previsão e mapeamento do carbono orgânico no solo (WERE et al., 2015) e em previsões climáticas (BELAYNEH et al., 2014).

A máquina de vetores de suporte é um tipo de técnica de aprendizado de máquina com aplicações bem-sucedidas tanto na classificação de dados (SVM) quanto na geração de modelos de regressão (SVR). Uma das principais características da SVR é que, ao invés de minimizar o erro de treinamento observado, a técnica tenta minimizar o erro generalizado, de forma a obter um desempenho generalizado. Este erro de generalização é a combinação do erro de treinamento e um termo de regularização que controla a complexidade do espaço de hipóteses (BASAK; PAL; PATRANABIS, 2007).

A produtividade agrícola é influenciada por diversos fatores, como o solo, clima, tipo de cultivar, tipo manejo da cultura, dentre outros. Com isso, os modelos de regressão para

estimativa de produtividade devem ter forte capacidade de generalização, podendo ser aplicados em culturas com diferentes condições de manejo, mantendo a capacidade preditiva. Neste contexto, a SVR pode trazer bons resultados, visto que seu foco principal é gerar modelos com maior capacidade de generalização. A SVR é utilizada especificamente para obter modelos preditivos por meio de uma função de kernel não linear e um número de vetores de suporte identificados (BASAK; PAL; PATRANABIS, 2007), podendo ser expressa pela fórmula:

$$Y = \sum_{i=1}^N (a_i^*, a) K(x_i, x) + b, \quad (1)$$

onde N é o número de vetores de suporte, a_i^*, a são multiplicadores de Lagrange obtidos por meio da resolução do problema de otimização quadrática (VAPNIK; GOLOWICH; SMOLA, 1997), $K(x_i, x)$ representa a função do kernel e b é uma constante. No Quadro 1 são apresentadas as funções de alguns dos principais kernels disponíveis para a SVR no pacote kernlab (KARATZOGLOU et al., 2004) do software R.

Quadro 1 – Alguns dos kernels para a SVR disponíveis no pacote Kernlab do software R

Kernel	Fórmulas
Gaussian RBF	$K(x_i, x) = \exp(-\sigma \ x_i - x\ ^2)$ (2)
Polinomial	$K(x_i, x) = (scale \cdot \langle x_i, x \rangle + offset)^{degree}$ (3)
Linear	$K(x_i, x) = \langle x_i, x \rangle$ (4)
Laplace RBF	$K(x_i, x) = \exp(-\sigma \ x_i - x\)$ (5)
Bessel	$K(x_i, x) = \frac{J_{v+1}(\sigma \ x_i - x\)}{\ x_i - x\ ^{-n(v+1)}}$ (6)

onde x_i, x representam dois vetores em um mapeamento de características determinado, σ é um parâmetro livre, $scale$ e $offset$ são constantes, $degree$ representa o grau do polinômio e J_{v+1} indica a função de Bessel, RBF (do inglês *Radial Basis Function*).

Os algoritmos de RNA, por sua vez, simulam os processos de aprendizagem humana, por meio do estabelecimento e reforço de ligações entre dados de entrada e saída. Os links, em seguida, conectam dados de entrada e saída na ausência de dados de treinamento (CAMPBELL; WYNNE, 2011). As RNAs são formadas por elementos de processamento simples, conhecidos como neurônios artificiais, sendo estes ativados por uma função (função de ativação), obtendo uma única resposta. O neurônio artificial é a unidade de processamento de informação de uma RNA, podendo ser representado matematicamente como:

$$Y_k = \varphi(V_k), \quad (7)$$

em que: Y_k = saída do neurônio artificial; φ = função de ativação; V_k = resultado do combinador linear, ou seja:

$$V_k = \sum_0^m x_m w_m, \quad (8)$$

sendo x_m o número de entradas e w_m o peso para cada entrada de m . A RNA Perceptron multicamadas é a mais popularmente utilizada. A arquitetura da rede multicamadas consiste em camada de entrada, camada escondida e camada de saída, cada uma com um conjunto de nós interconectados (neurônios) trabalhando em paralelo para transformar os dados de entrada em valores de saída.

A regressão linear é uma das mais conhecidas técnicas clássicas de MD. No entanto, assim como a SVR e a RNA, parte da hipótese de uma distribuição idêntica e independente nas amostras de dados de aprendizagem, o que na maioria das vezes não ocorre com os dados espaciais, que são frequentemente caracterizados pela autocorrelação espacial (BAILEY; GATRELL, 1995; VATSAVAI et al., 2012). Assim, surge uma nova vertente para pesquisas, que busca a criação de métodos e técnicas nas quais a dimensão espacial dos dados é considerada durante a descoberta de conhecimento.

3.2 MINERAÇÃO DE DADOS ESPACIAIS

A mineração de dados espaciais, também conhecida como mineração de dados geográficos, é a extensão da MD voltada para domínios de aplicação onde a consideração da dimensão espacial é essencial na extração de conhecimento. A importância da consideração dos relacionamentos de vizinhança, está no fato de que, para processos espaciais, a medida associada a um local ou objeto tende a ser similar às medidas em locais próximos (RODDICK; HORNSBY; SPILIOPOULOU, 2001). Portanto, a MDE surgiu como uma área de pesquisa ativa para extrair conhecimento implícito e relevante de grandes bancos de dados espaciais (ANSELIN; SYABRI; KHO, 2006).

A MDE é um campo de pesquisa crescente que ainda está em uma fase muito precoce. Desde o início do século XXI, devido às aplicações generalizadas do sistema de posicionamento global (GPS – do inglês *Global Positioning System*), compartilhamento e mapeamento de dados espaciais na web, detecção remota de alta resolução e serviços baseados em localização, cada vez mais domínios de pesquisa criaram ou obtiveram acesso a dados geográficos de alta qualidade (MENNIS; GUO, 2009).

A MDE engloba várias tarefas e, para cada tarefa, muitas vezes estão disponíveis diversos métodos, sejam eles computacionais, estatísticos, visuais ou alguma combinação deles. O objetivo da MDE é integrar e desenvolver métodos em vários campos para a análise de dados

espaciais grandes e complexos (MENNIS; GUO, 2009). Os dados espaciais, ou objetos, são compostos não apenas de atributos gerais representados como texto, mas também de atributos espaciais, como ponto, linha e superfície de espaço bidimensional ou tridimensional. Portanto, usar os métodos tradicionais de estatísticas clássicas para analisar dados espaciais pode gerar resultados incoerentes, dado o papel importante que a localização desempenha na compreensão dos fenômenos observados no espaço (ALKATHIRI; ABDUL; POTDAR, 2016).

Segundo Hamdad et al. (2015), as ferramentas de MDE funcionam de acordo com duas abordagens principais. A primeira abordagem, que é a mais intuitiva, porém mais demorada, consiste em usar algoritmos de MD clássicos em dados espaciais pré-processados. O pré-processamento consiste em extrair a tabela de dados suavizados da matriz entre objetos vizinhos ou representar as relações espaciais como novas características (atributos). A segunda abordagem consiste em desenvolver técnicas de MDE específicas, que levem em consideração dinamicamente as relações espaciais dentro do processo de mineração. Portanto, o processo exploratório desta abordagem é mais rápido que o primeiro, porém mais complicado de implementar.

Existe um conjunto de ferramentas de MDE propostas na literatura, dentre elas: o GeoMiner (HAN; KOPERSKI; STEFANOVIC, 1997), o primeiro software de extração de conhecimento de bases de dados espaciais; Weka-gdpm (BOGORNÝ et al., 2006), uma API espacial de pré-processamento que pode ser adicionada ao software Weka para tratar dados espaciais; a Ingens (APPICE; LANZA; MALERBA, 2007), uma plataforma integrada para MDE dentro de um ambiente SIG; o SD-Miner (BAE et al., 2009), uma ferramenta composta por interface gráfica de usuário, MD clássica e espacial e armazenamento de dados; o GeoDMA (KÖRTING; GARCIA FONSECA; CÂMARA, 2013), que funciona como um *plugin* para o SIG TerraView, contendo algoritmos de segmentação, extração e seleção de características e classificação; e por fim o EasySDM (HAMDAD et al., 2015), um software que integra as duas abordagens principais da MDE (realizada no pré-processamento dos dados e com algoritmos específicos) e dispõe de um módulo de visualização de dados.

Para Bae et al. (2009), existe uma carência de softwares comerciais para MDE, sendo a maioria deles softwares acadêmicos e de domínio público. Segundo os autores, para impulsionar a utilização da MDE em pesquisas relacionadas com dados espaciais reais, o desenvolvimento de ferramentas é indispensável. A maioria dessas ferramentas propostas na literatura até o momento para trabalhar com dados espaciais, dispõem de algoritmos de agrupamento, classificação e mineração de regras de associação. Com isso, existem poucas

opções de métodos e técnicas de regressão, sendo estas ferramentas fundamentais para obtenção de modelos de estimativa.

Os modelos de regressão espacial mais utilizados possuem um coeficiente de auto regressão espacial (ρ), que mede a força da autocorrelação espacial, e uma matriz de peso espacial (w_{ij}), a qual corresponde a uma estrutura de vizinhança. Os modelos auto regressivos condicionais (CAR – do inglês *Conditional Autoregressive Models*) e auto regressivos espaciais (SAR – do inglês *Spatial Autoregressive Models*) incorporam a autocorrelação espacial usando a matriz de vizinhança, que especifica o relacionamento entre os valores de resposta (no caso do CAR) ou resíduos (no caso do SAR) em cada local (i) e aqueles locais vizinhos (j) (CRESSIE, 1993). Assim como nos modelos CAR e SAR, no modelo de aditivos generalizados (GAM – do inglês *Generalized Additive Models*) (WOOD, 2006), as dependências espaciais presentes nos dados são atendidas pelo termo de autocorrelação e melhoram a classificação geral e a precisão nas estimativas (TASSINARI; LORENZON; PEIXOTO, 2013). O Quadro 2 apresenta as equações dos modelos de regressão espaciais SAR (Equação 9), CAR (Equação 10) e GAM (Equação 11), que são os modelos mais comuns presentes na literatura.

Quadro 2 - Modelos de regressão espaciais mais comuns na literatura, nos quais a autocorrelação espacial é permitida e considerada na descoberta de padrões presentes nos dados.

Modelos	Fórmulas
SAR	$Y_i = \beta_0 + \sum_{k=1}^k \beta_k X_{ki} + \rho w_{ij} + e_i \quad (9)$
CAR	$Y_i = \beta_0 + \sum_{k=1}^k \beta_k X_{ki} + e_i, \text{ sendo } e_i = \lambda w_{ij} e_i + \xi \quad (10)$
GAM	$Y_i = \beta_0 + \sum_{k=1}^k \beta_k X_{ki} + S(lat_{ij}, long_{ij}) + e_i \quad (11)$

onde β_0 representa uma constante, β_k é um vetor representando as inclinações associadas aos preditores na matriz preditora X_{ki} , ρ e λ são os coeficientes de auto regressão espacial, ρw_{ij} corresponde a matriz de pesos espaciais, $S(lat_{ij}, long_{ij})$ é uma função de suavização bidimensional.

O modelo SAR pode ser visto como uma generalização de modelos de regressão linear padrão, de modo que a autocorrelação espacial é permitida e levada em consideração. O modelo SAR pode assumir três formas diferentes, dependendo de onde se acredita que o processo autoregressivo espacial ocorre (ANSELIN; SYABRI; KHO, 2006).

A distinção fundamental entre o modelo SAR e CAR está na especificação do modelo (ANSELIN; FLORAX; REY, 2004). Os modelos SAR explicam as relações entre as variáveis de resposta em todos os locais na rede simultaneamente, o efeito espacial é considerado interno. Em contraste, os modelos CAR especificam a distribuição de uma variável de resposta em um

local, dependendo dos valores de seus vizinhos; o efeito espacial dos vizinhos é considerado externo (WOOD, 2006).

O modelo GAM é um modelo linear generalizado, com um preditor linear, envolvendo uma soma de funções de suavização de variáveis independentes contínuas. O modelo permite uma especificação bastante flexível da dependência da resposta nas variáveis independentes, especificando o modelo apenas em termos de “funções de suavização”, em vez de relações paramétricas detalhadas. O modelo GAM inclui um termo adicional ($S(lat_{ij}, long_{ij})$), para indicar a função de suavização bidimensional com base nas coordenadas Leste e Norte. Este termo indica se a variação residual nos dados apresenta qualquer estrutura espacial, depois de controlar as covariáveis explicativas. Este parâmetro de suavização é selecionado minimizando risco de uma estimativa imparcial, o que equivale a minimizar o erro quadrático médio esperado (WOOD, 2006).

Os modelos SAR, CAR e GAM foram utilizados na avaliação da produção apícola (TASSINARI; LORENZON; PEIXOTO, 2013), onde segundo os autores, o modelagem dos contextos espaciais facilita a análise das regiões produtoras de mel, o que não está disponível com outros modelos de regressão clássicos. O modelo SAR também foi utilizado para a estimativa de produtividade de soja por meio de variáveis agrometeorologias (ARAÚJO; URIBE-OPAZO; JOHANN, 2014; GRZEGOZEWSKI et al., 2017) e, em comparação com os modelos de regressão clássicos, apresentou resultados superiores e indicou forte autocorrelação espacial da produtividade.

Embora a mineração de dados espaciais tenha apresentado resultados, existem algumas dificuldades técnicas, como por exemplo, dados excessivos, dados com alta dimensionalidade, dados com ruídos, dados incertos, o fato de haverem vários pontos de vista, e também as diferentes formas de representação do conhecimento (LI et al., 2016).

A abundância de dados espaciais oferece oportunidades para novas direções de pesquisa. No entanto, os dados são muitas vezes de diferentes fontes e recolhidos para finalidades distintas, em diferentes condições, tais como a incerteza de medição, amostragem tendenciosa, variando unidade de área, e a restrição de confidencialidade. É importante compreender a qualidade e as características dos dados escolhidos. Novos tipos de dados e novas áreas de aplicação expandiram significativamente a fronteira da pesquisa da MDE, exigindo o desenvolvimento de novos métodos de mineração de dados e descoberta de novos tipos de padrões (MENNIS; GUO, 2009).

A maioria dos dados espaciais são obtidos atualmente por meio do sensoriamento remoto, a partir de observações da superfície terrestre. A tecnologia de sensoriamento remoto

desempenha um papel importante na captura de uma ampla gama de informações espaço-temporais sobre recursos dinâmicos e os ambientes que cercam os seres humanos. Os dados do sensoriamento remoto são parte fundamental na tomada de decisões sociais, políticas e econômicas (LI et al., 2016).

Com ao advento da AP, o mapeamento da variabilidade espacial de certos fatores, em nível de talhão, para fim de gerenciamento localizado, tornou-se uma necessidade. O sensoriamento remoto tem sido apontado como ferramenta de auxílio no mapeamento do vigor das culturas ou a sua produtividade, o que é fundamental para tomada de decisão em âmbitos agrícolas (ARAÚJO; VETTORAZZI; MOLIN, 2005). Neste contexto, os dados espaciais obtidos das culturas por meio do sensoriamento, associados as técnicas de MDE, podem ser utilizados na geração de modelos confiáveis para a estimativa de produtividade. Para a estimativa da produção, é necessário conhecer a área cultivada com a espécie de interesse e a sua produtividade. O conhecimento da área pode ser obtido por meio do sensoriamento remoto, cujos dados podem ser correlacionados (por meio das técnicas de MDE) com a produtividade, obtida em campo pelas colhedoras de precisão (ASSAD et al., 2007).

3.3 ESTIMATIVA DE PRODUTIVIDADE DE GRÃOS

A previsão da produtividade de grãos pode ser feita por modelos de regressão e de classificação. Nesses modelos, define-se a produtividade como variável dependente e as demais variáveis como independentes. O Quadro 3 apresenta alguns trabalhos visando a estimativa de produtividade de diferentes culturas, os quais fazem uso de conjuntos de dados distintos e com diferentes técnicas de análise.

Quadro 3 - Resumo de trabalhos com objetivo de estimar a produtividade de diferentes culturas e com técnicas distintas para geração de modelos de estimativa.

Cultura	Tipo de dados	Técnica	Autores
Soja e trigo	Dados de satélite	Regressão linear simples	Araújo; Vettoreazzi e Molin (2005)
Cana-de-açúcar	Dados de satélite	Regressão linear simples	Picoli et al. (2009)
Milho e algodão	Dados de satélite, meteorológicos e fotografias digitais	Modelo LUE	Alganci et al. (2014)
Soja e trigo	Dados de satélite	Regressão linear simples	Holzman; Rivas e Piccolo (2014)
Soja	Dados meteorológicos	SAR, CAR e RLM	Araújo; Uribe-Opazo e Johann (2014)
Milho	Imagens de RPA	Regressão linear simples e quadrática	Maresma et al. (2016)
Soja	EVI e dados meteorológicos	Teste de Moran's	Grzegozewski et al. (2017)
Trigo	Dados de satélite	Modelo SAFY	Silvestro et al. (2017)
Soja	Imagens de RPA	SVR	Gerke (2017)

Como mostrado no Quadro 3, diferentes técnicas podem ser aplicadas para estimativa de produtividade de diferentes culturas, com técnicas distintas e utilizando conjuntos de dados distintos na estimativa. No entanto, apenas o trabalho de Araújo; Uribe-Opazo e Johann (2014) faz uso de modelos de regressão espacial. Os autores utilizam os modelos espaciais SAR e CAR e o modelo de regressão linear múltipla clássica (RLM), obtendo valores de coeficiente de determinação (R^2) de 50,58%, 50,61% e 24,15%, respectivamente. Isso indica que os algoritmos de regressão propostos para MDE possibilitaram a geração de modelos com capacidade superior de explicação da variação da produtividade.

O sensoriamento remoto tem sido fortemente utilizado em trabalhos relacionados a estimativa de produtividade agrícola. Os modelos de estimativa gerados pelos dados de sensoriamento remoto têm sido apontados como técnica alternativa à atual metodologia para elaboração de mapas de produtividade, por meio de colhedoras instrumentadas. A produtividade pode ser estimada a partir de sua relação com o vigor da cultura, que por sua vez, pode ser determinado via sensoriamento remoto, pela sua relação com determinados índices de vegetação (IVs) gerados por meio das imagens multiespectrais (ARAÚJO; VETTORAZZI; MOLIN, 2005). Os IVs são indicadores à presença e condição da vegetação e por isso são frequentemente utilizados como variáveis auxiliares para estimar a produtividade de culturas agrícolas.

Comparados com os métodos terrestres tradicionais, como as estimativas visuais e análise de produtividade por meio de pontos amostrais, o sensoriamento remoto oferece observações repetidas da cobertura terrestre, adequadas para o mapeamento e acompanhamento de áreas agrícola de extensões variadas (XIN et al., 2013). Porém, assim como em relação aos trabalhos que fazem uso da MDE, pode-se verificar no Quadro 3 que existem poucos trabalhos utilizando dados de imagens obtidas por RPA para a estimativa de produtividade, sendo a maioria deles com a utilização de imagens orbitais.

3.4 SENSORIAMENTO REMOTO NA AGRICULTURA

O surgimento e a evolução das geotecnologias na área de sensoriamento remoto permitem o desenvolvimento de metodologias inovadoras para o mapeamento da cobertura do solo (NONATO; OLIVEIRA, 2013). Este mapeamento é um dos procedimentos essenciais em estudos ambientais, em avaliações da biodiversidade, no monitoramento agrícola, no apoio às decisões de ações sociais, políticas e econômicas, e na estimativa da produtividade agrícola e previsão de safras (YI; SHIMABUKURO; QUINTANILHA, 2007).

As plataformas de sensoriamento remoto podem ser divididas em três categorias: terrestre (radiômetros, fotômetro, etc.), aerotransportado (aviões e RPAs) e orbital (satélites). Para avaliar essas diferentes plataformas, as resoluções espacial, espectral e temporal devem ser consideradas. A resolução espacial define o tamanho do pixel de imagens de satélite e aéreas que cobrem a superfície terrestre. A resolução espectral indica a largura das faixas espectrais em que o sensor pode captar a radiação refletida. Por último, a resolução temporal indica a frequência com que a plataforma pode obter informações de um mesmo local (WÓJTOWICZ; WÓJTOWICZ; PIEKARCZYK, 2016).

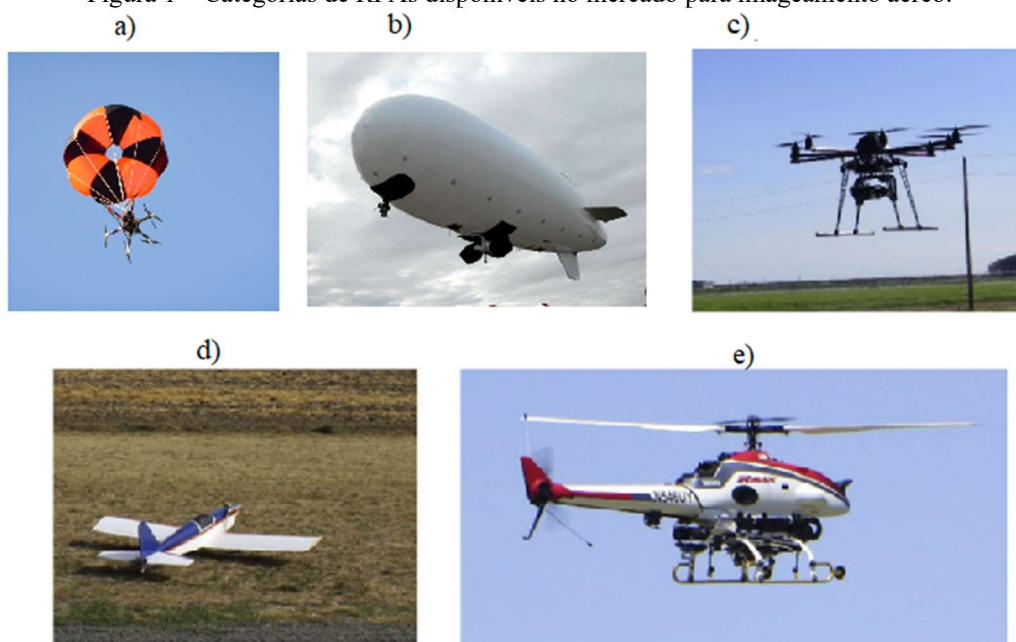
Os satélites e aviões tripulados, consideradas as plataformas mais tradicionais de sensoriamento remoto, não conseguem fornecer informações com as necessárias resoluções espaciais, espectrais e temporais, dependendo do objeto de estudo (EHSANI et al., 2012; MORAN; INOUE; BARNES, 1997). Estas plataformas apresentam vantagens e desvantagens que envolvem fatores tecnológicos, operacionais e econômicos. Os satélites podem traçar grandes áreas ao mesmo tempo, mas, por outro lado podem sofrer de cobertura de nuvens e restrições ao relacionar a temporização de imagens com fases fenológicas específicas das culturas, por causa das aquisições em tempo fixo. Os levantamentos por meio de aviões podem ser planejados de maneira mais flexível, no entanto, a organização desses levantamentos pode ser difícil e de alto custo (MATESE et al., 2015).

Uma alternativa recente para atividades de sensoriamento remoto é a utilização de RPAs, as quais possibilitam observações qualitativas e quantitativas de áreas de interesse com alta flexibilidade de uso, baixo custo operacional, altas resoluções espaciais e temporais (CANDIAGO et al., 2015; SALAMÍ; BARRADO; PASTOR, 2014).

3.4.1 RPAS na agricultura

As RPAs são popularmente conhecidas como drones e veículos aéreos não tripulados (VANTs ou UAV – do inglês *Unmanned aerial vehicle*). O termo RPA foi introduzido em 2015, juntamente com os conceitos de sistemas aéreos pilotados remotamente (RPAS - *Remote Pilot Aircraft System*), o qual é usado para se referir ao conjunto representado pela aeronave e demais equipamentos necessários para a sua operação como, por exemplo, estação de pilotagem remota, enlace de comando e controle, equipamentos para lançamento e recuperação (ANAC, 2015). As RPAs podem ser divididas em cinco diferentes categorias (SANKARAN et al., 2015): paraquedas (Figura 1a), dirigível (Figura 1b), multi-rotores (Figura 1c), asa fixa (Figura 1d) e helicópteros (Figura 1e).

Figura 1 – Categorias de RPAs disponíveis no mercado para imageamento aéreo.



Fonte: O Autor

As imagens coletadas pelas RPAs permitem, além da geração de mosaicos da área de interesse, a geração dos DSMs e DTMs. A disponibilidade dos DSMs e DTMs com alta resolução espacial e precisão vertical é de importância crescente para todas as ciências interessadas na reconstrução tridimensional do meio ambiente (MANCINI et al., 2013). Estes modelos fornecem subsídios para detecção de erosão no solo e novas informações sobre as culturas, como a altura da planta e altitude do terreno, as quais podem ser utilizadas como variáveis auxiliares em estimativas e previsões agrícolas. Alguns softwares disponíveis para processamento das imagens obtidas por RPAs, como o PIX4D¹ permitem a geração de mapas de refletância. Nestes mapas, o interesse está na refletância do terreno. No entanto, o valor do pixel depende de muitos outros fatores, que precisam ser corrigidos para obter uma medida da refletância. Esses fatores incluem: luz capturada pelo sensor, ISO, abertura, velocidade do obturador, resposta do sensor e sistema óptico, vinheta. Ao gerar o mapa de refletância, o PIX4D usa os valores de alguns parâmetros presentes no EXIF (metadados) das imagens para corrigir (pelo menos parcialmente) os fatores mencionados acima.

A aplicação de RPAs na área agrícola e em missões de reconhecimento é favorecida e facilitada pelo atual estágio de desenvolvimento tecnológico, principalmente pela redução do custo, redução no tamanho dos equipamentos e pela necessidade de otimização da produção (JORGE; INAMASU, 2013). A utilização de RPAs, juntamente com sensores de imagem,

¹ <https://support.pix4d.com/hc/en-us/articles/202559509-Index-Calculator-Reflectance-Map-Radiometric-Correction#gsc.tab=0>

alcance e posicionamento, permite a coleta de imagens multiespectrais com resolução espacial que pode chegar a 3,4 cm e oferece várias possibilidades no domínio da AP, manejo florestal e geociências (CANDIAGO et al., 2015).

O desenvolvimento das RPAs surgiu como importante opção na AP, podendo ser utilizada no monitoramento de safras, gestão e logística da produção, estimativa de produção, contagem de plantas, detecção de falhas de plantio, estresse hídrico, nutrição e detecção de pragas e doenças (CHAVES et al., 2015). O Quadro 4 apresenta algumas aplicações de RPAs na agricultura, mostrando os métodos tradicionais e quais alternativas com a utilização deste equipamento.

Quadro 4 – Algumas aplicações de RPAs na agricultura, elencando os métodos tradicionais e como podem ser realizados com esta plataforma (adaptado de Sankaran et al. (2015)).

Aplicação	Método tradicional	Método com RPA	Referência
Altura de plantas	Escala de medição	Sistema LIDAR, importante saber a elevação do terreno	Merz e Chapman, (2012), Zarco-Tejada; Ustin e Whiting (2005)
Biomassa e emergência das plantas	Avaliação visual; amostragem destrutiva; contagem de plantas	Imagens VIS e NIR para determinação do verde da planta	Chapman et al. (2014)
Estresse hídrico; condutância estomática; estresse por calor e por salinidade	Avaliação visual, medições de umidade do solo, porômetro	Imagem VIS, NIR e térmicas para medir a temperatura do dossel e as bandas de absorção de água	Zarco-Tejada; González-Dugo e Bernil (2012)
Nutrição	Análise de nutrientes foliar e / ou pecíolo	Imagem VIS-NIR para estimar o nitrogênio da folha e potencialmente outros nutrientes	Lebourgeois et al. (2012)
Índice de área foliar	Amostragem destrutiva; análise da copa das plantas	Imagem VIS-NIR para estimar biomassa e cobertura da planta	Hunt et al. (2010)
Suscetibilidade de doenças	Avaliação visual da doença	Imagem de VIS-NIR para avaliar a saúde / dano da planta	Di Gennaro et al. (2016); Jansen et al. (2014)
Determinação de nitrogênio	Amostragem destrutiva; análise da química laboratorial	Imagem de VIS-NIR para estimativa de teores de nitrogênio	Maresma et al. (2016)
Estimativa de produtividade em parcelas	Colheita manual ou colhedora de parcelas	Imagens de VIS-NIR para estimativa de produtividade	Gerke (2017), Maresma et al. (2016), Yu et al. (2016)
Estimativa de severidade de doenças	Escalas diagramáticas e análise visual de plantas	Imagens de VIS-NIR para estimativa de severidade	Lacerda (2017)

Pelo fato de poucos trabalhos utilizarem dados de RPAs para a estimativa de produtividade, não existe um protocolo definido da sua utilização para esta finalidade. Gerke (2017), visando a estimativa da produtividade de trigo, faz uso de uma RPA de asa fixa, com câmeras que capturam dados no espectro visível (RGB – do inglês *Red*, *Green* e *Blue*) e infravermelho próximo (NIR – do inglês *Near Infrared*), e também compara três diferentes resoluções espaciais (3,4, 10 e 20 cm/pixel). Os melhores resultados foram obtidos com a

câmera RGB e resolução espacial de 20 cm/pixel, com valores de coeficiente de correlação (r) de até 0,61. Isso indica que nem sempre uma resolução mais alta será capaz de fornecer os melhores resultados, isso dependerá do objeto de estudo e das condições de obtenção das imagens em campo.

No trabalho de Maresma et al. (2016), os autores obtêm altos valores de coeficiente de determinação (R^2) na estimativa de produtividade de milho, variando de 0,60 a 0,92, quando utilizando a regressão linear simples e como atributos preditores a altura das plantas e diferentes índices espectrais. Assim como no trabalho de Gerke (2017), os autores fazem uso de uma RPA de asa fixa e uma câmera multispectral (com bandas R, G e NIR), no entanto, a resolução espacial das imagens obtidas foi de 15 cm.

Yu et al. (2016) visam a estimativa da produtividade da soja utilizando dados obtidos por meio de uma RPA multi-rotor (octocoptero). Os autores também utilizam uma câmera multispectral com as bandas R, G e NIR e a resolução espacial das imagens foi de 3,5 cm/pixel. A estimativa foi realizada por meio da regressão linear simples entre a produtividade da soja e a área de cobertura das plantas nas parcelas experimentais, obtendo valores de r de até 0,74.

Como pôde-se verificar, a utilização de RPAs com a finalidade de estimar a produtividade de culturas agrícolas é um campo de pesquisa recente. Não se sabe de forma concreta quais equipamentos e técnicas fornecerão os dados com maior capacidade de estimativa. Com isso, novas pesquisas devem ser realizadas, visando a geração de modelos de estimativas com capacidade de explicar de maneira significativa a variação da produtividade, podendo se tornar uma alternativa em áreas onde não há a disponibilidade de colhedoras de precisão.

3.4.2 Pré-Processamento das imagens obtidas por meio sensoriamento remoto

Assim como as imagens obtidas por aviões tripulados e satélites, aquelas obtidas por RPAs também necessitam de correções relacionadas a geometria (georreferenciamento e ortorretificação das imagens) e radiometria (eliminação de ruídos e correção atmosférica), as quais são realizadas na etapa de pré-processamento (COSTA et al., 2016). Essas correções têm como objetivo garantir a qualidade dos dados obtidos das imagens e possibilitar a obtenção de resultados mais precisos e coerentes, pois fazem com que as imagens representem o mais próximo da realidade apresentada em campo.

A técnica de correção geométrica mais utilizada baseia-se no uso de pontos de controle no terreno (GCPs – do inglês *Ground Control Points*) localizados na imagem e no mapa correspondente, para determinar empiricamente uma transformação de coordenadas

matemáticas, buscando corrigir a geometria (RICHARDS, 2013). Esta correção é necessária devido aos erros do receptor GNSS da RPA, distorções geradas na ortorretificação das imagens, dentre outros ruídos que afetam a precisão do georreferenciamento. Perin et al. (2016) fazem uma análise comparativa de métodos de correção geométrica em imagens obtidas por RPAs, utilizando GCPs. Os autores constataram que a correção do georreferenciamento disponibilizada no sistema de informação geográfica QGIS², utilizando o tipo de transformação Polinomial 1 e o método de reamostragem vizinho mais próximo, possibilitou as melhores acurácias na correção geométrica, quando comparadas com um receptor GNSS geodésico, o qual possui precisão milimétrica tanto vertical, quanto horizontal.

A ortorretificação é um processo que visa corrigir uma imagem, de modo que cada pixel seja colocado na posição geométrica teórica, como se tivesse sido imageada na vertical ou mais próximo desta situação (RICHARDS, 2013).

A correção radiométrica pode ocorrer de duas formas: absoluta e relativa. A correção absoluta é aquela que utiliza parâmetros de calibração do sensor e da atmosfera para estimar a refletância de superfície (correção atmosférica). Já a correção relativa utiliza somente dados da própria imagem, não exigindo parâmetros atmosféricos (SONG et al., 2001).

As correções nas imagens obtidas por meio de SR são realizadas, em sua maioria, em SIGs. Estes softwares permitem a realização de múltiplas tarefas necessárias para a obtenção de análises visuais e posterior tomada de decisão, sendo úteis para planejamentos na agricultura, pecuária, meio-ambiente e gestão de recursos urbanos variados. Os SIGs fornecem ampla variedade de algoritmos de extração de características, classificação de imagens, extração de dados, etc., e possibilitam a entrada de dados de diferentes fontes e formatos.

3.4.3 Sistemas de Informação Geográfica

Um SIG é constituído por um conjunto de ferramentas especializadas em adquirir, armazenar, recuperar, transformar e emitir informações espaciais. Esses dados geográficos descrevem objetos do mundo real em termos de posicionamento, com relação a um sistema de coordenadas, seus atributos não aparentes (como a cor, pH, custo, incidência de pragas, etc) e as relações topológicas existentes. Shimabukuro; Maeda e Formaggio (2015) estabelecem que, ao longo dos últimos anos, os SIGs passaram a assumir papel importante na convergência de diversas outras tecnologias, como o sensoriamento remoto, GNSS e a internet, promovendo novos níveis de funcionalidade e acessibilidade.

² <https://www.qgis.org/en/site/>

Os SIGs são utilizados em diversas áreas e com diferentes finalidades, como planejamento urbano, conservação do solo, hidrologia, AP, em estudos relativos ao meio ambiente e recursos naturais, em pesquisas para previsão de determinados fenômenos ou no apoio a decisões de planejamento, considerando a concepção de que os dados armazenados representam um modelo do mundo real (BURROUGH, 1986). Por possibilitarem a integração de dados de diferentes fontes e formatos, esses sistemas facilitam a interpretação dos dados na AP, os quais são obtidos de diferentes tipos de sensores e implementos. Assim, os SIGs se estabelecem como ferramenta fundamental na tomada de decisões em âmbitos agrícolas, auxiliando no desenvolvimento de pesquisas relacionadas a diferentes estimativas, como a estimativa de produtividade que é o foco deste trabalho. O Quadro 5 apresenta alguns dos principais SIGs utilizados na literatura.

Quadro 5 - Principais SIGs utilizados em pesquisas relacionadas a análise da cobertura terrestre

Software	Licença	Plataformas
³ ARCGIS ®	Software proprietário	Windows
⁴ Mapinfo TM	Software proprietário	Windows
⁵ SPRING ©	Livre	Multiplataforma
⁶ ENVI ®	Software proprietário	Multiplataforma
⁷ TerraView ©	GPL	Multiplataforma
⁸ QGIS ®	GPL	Multiplataforma
⁹ GvSIG ©	GPL	Multiplataforma

O software QGIS dispõe de uma variedade de algoritmos para diferentes finalidades e possui métodos intuitivos para visualização de dados, extração de características, exportação de dados, correções no georreferenciamento, etc. Por ser um software com licença pública geral (GPL – do inglês *General Public License*), permite a utilização de todas suas funcionalidades sem custos adicionais, sendo assim, uma ferramenta indicada para utilização em pesquisas acadêmicas.

³ <https://www.arcgis.com/features/index.html>

⁴ <http://meixnerimaging.com/mapinfo-pro/>

⁵ <http://www.dpi.inpe.br/spring/>

⁶ <http://www.envi.com.br/index.php/>

⁷ <http://www.dpi.inpe.br/terralib5/wiki/doku.php>

⁸ <https://www.qgis.org/en/site/>

⁹ <http://www.gvsig.com/>

4 MATERIAL E MÉTODOS

4.1 ÁREA EXPERIMENTAL

Os experimentos foram realizados em uma área comercial localizada na Fazenda Paiquerê, município de Pirai do Sul, Paraná, Brasil, cujas coordenadas de referência são 50° 7' 51.70" S e 24°21'23.485" W, com altitude média de 970m (Figura 2). O clima da região é subtropical úmido, com temperatura média anual nos meses mais frios de 18°C e nos meses mais quentes de 22°C, e precipitação pluvial média anual acima de 1300 mm¹⁰. O solo da área é classificado como Latossolo Vermelho-Amarelo com textura argilosa, cultivado sobre plantio direto. Foram analisadas as culturas de trigo (safra 2016) e soja (safra 2016-2017).

Figura 2 - Área sobrevoada nas safras de trigo 2016 e soja 2016-2017.



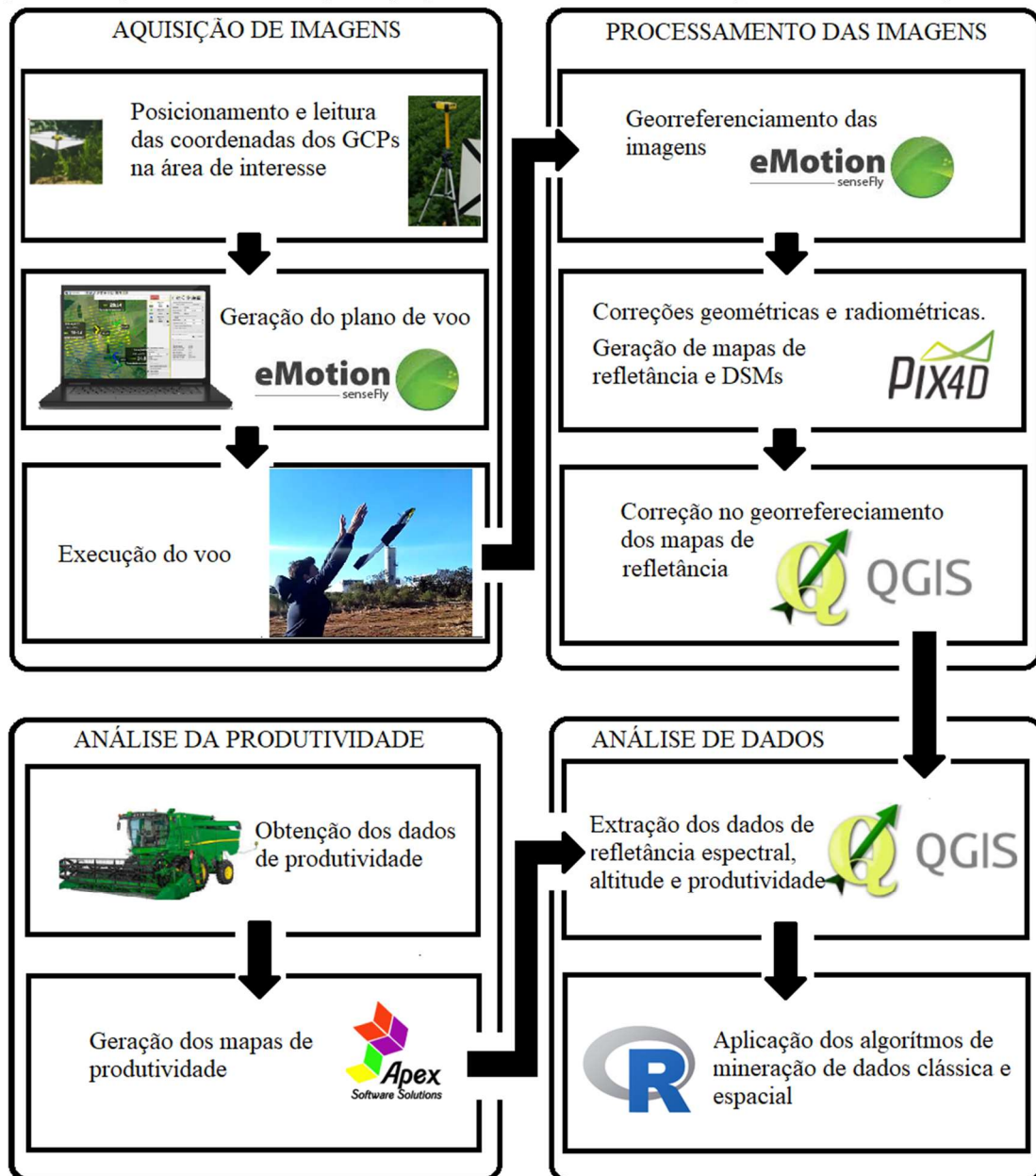
Fonte: O Autor

A área imageada nas safras de trigo e soja está localizada no talhão T05 da Fazenda Paiquerê, que corresponde a uma área de aproximadamente 60,4 ha. Na safra de trigo 2016, a cultivar utilizada foi a TBIO Sinuelo, semeada no dia 28 de junho de 2016, com espaçamento de 0,17 m entre as linhas e densidade de 70 sementes por metro linear. Na safra da soja utilizou-se a cultivar BMX Alvo, semeada no dia 27 de novembro de 2016, com espaçamento entre linhas de 0,5 m e densidade de 280.000 plantas/ha. A adubação na safra de soja consistiu em fertilizante Super Simples (150 kg/ha) e Cloreto de Potássio (KCl – 200kg/ha).

¹⁰ Embrapa Boletim de Pesquisa e Desenvolvimento. Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/338503/1/ParanaPirai.pdf>

Definida a área de interesse, o próximo passo foi a criação da sequência metodológica empregada para a obtenção de modelos confiáveis para a estimativa de produtividade. A Figura 3 apresenta a sequência metodológica utilizada neste trabalho, contendo as etapas de aquisição de imagens, processamento das imagens, aquisição de dados de produtividade e por fim a análise de dados. Estas etapas serão explicadas detalhadamente nos tópicos seguintes.

Figura 3 - Sequência da metodologia empregada na obtenção dos modelos para a estimativa da produtividade



Fonte: O Autor

4.2 AQUISIÇÃO DAS IMAGENS EM CAMPO

O imageamento da área experimental foi realizado com a RPA de asa fixa modelo *eBee* (Figura 4), fabricada pela empresa SenseFly (SENSEFLY, 2017). Essa RPA permite a tomada de imagens de alta resolução espacial, de forma rápida, fácil e sob demanda. Possui

embarcado sistema inercial, receptor GPS, piloto automático, sensor de altura e reverso no motor, permitindo pousos suaves em linha reta ou em espiral. Essa RPA proporciona a realização de voos de forma totalmente automática desde o lançamento, realizado manualmente, até o pouso.

Figura 4 – RPA eBee da empresa SenseFly utilizada no imageamento da área de estudo



Fonte: SenseFly (2017)

Para captura dos dados de imagem foram utilizadas duas câmeras com diferentes resoluções espectrais, a Sony Cyber-shot WX220 (RGB) e a Canon S110 (NIR), mostradas na Figura 5. Essas câmeras são adaptadas de modo que possam ser controladas pelo piloto automático da aeronave.

Figura 5 - Câmeras utilizadas no imageamento da área de estudo



Fonte: SenseFly (2017)

A câmera RGB (Figura 5a) adquire imagens no espectro visível e seus parâmetros de exposição são definidos automaticamente. Já a câmera NIR (Figura 5b) adquire dados de imagem na banda do infravermelho próximo (NIR), a região onde ocorre alta refletância da planta, seus parâmetros de exposição podem ser ajustados manualmente. O Quadro 6 apresenta as principais especificações das câmeras utilizadas.

Quadro 6 - Especificações das câmeras RGB e NIR utilizadas para obtenção das imagens em campo

Câmera	Especificações				
	Sensor	Tamanho imagem	Formato Imagem	Bandas	Intensidade de bits
RGB	7,76 mm (1/2.3 type) Exmor RTM CMOS sensor	18,2 Mpix	JPEG	R (660 nm) G (520 nm) B (450 nm)	24
NIR	1/1,7" back-illuminated CMOS	12,1 Mpix	JPEG e RAW	R (625 nm) G (550 nm) NIR (850 nm)	48

A aquisição das imagens em campo foi realizada por meio de uma sequência de etapas, as quais irão garantir a qualidade dos dados obtidos das mesmas. Dentre estas etapas está a

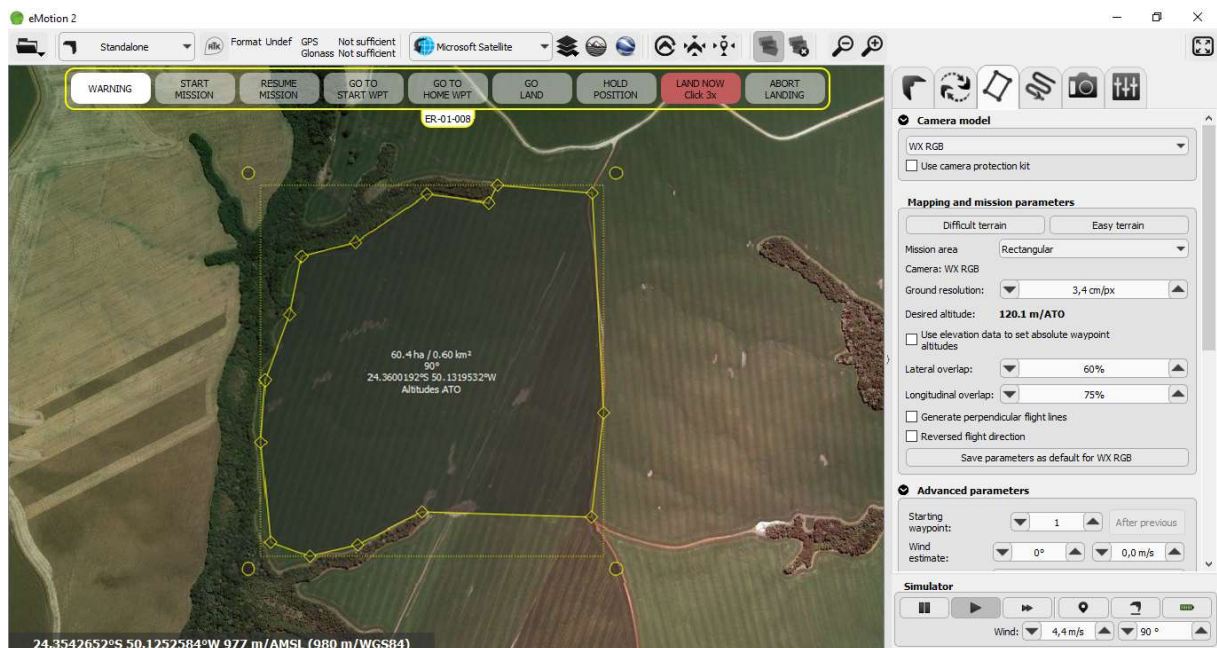
geração do plano de voo, posicionamento dos GCPs, testes e verificações na RPA, lançamento e pouso, etc.

4.2.1 Geração do plano de voo

A geração do plano de voo da RPA foi realizada no software *Emotion 2*¹¹ (Figura 6), utilizado desde a configuração prévia e simulações de plano de voo em laboratório, até o levantamento em campo. O usuário define as características dos dados que serão obtidos pela câmera acoplada na RPA, como a resolução espacial das imagens, os valores de sobreposição lateral e longitudinal. Com base nas informações especificadas pelo usuário, o software gera o melhor percurso a ser percorrido pela aeronave, assim como define a altitude da mesma durante o voo.

O usuário também pode acompanhar toda a execução do voo, bem como a situação da bateria, velocidade do vento (m/s), altura e altitude da aeronave, dentre outros aspectos que devem ser analisados durante o imageamento. No software *Emotion 2* o usuário também deve informar qual é a área que a aeronave poderá percorrer durante o pouso.

Figura 6 - Software *Emotion 2*, no qual é realizada a configuração e o acompanhamento do voo da RPA.



Fonte: O Autor

4.2.2 Posicionamento dos GCPs

Para obtenção das imagens em campo, inicialmente foram posicionados GCPs (Figura 7), para realizar as correções no georreferenciamento dos subprodutos obtidos por meio das imagens.

¹¹ Software comercial, disponibilizado juntamente com a RPA eBee para controle da aeronave, mais informações em: <https://www.sensefly.com/software/emotion/>

Figura 7 - GCP posicionado antes do imageamento da área de estudo.



Fonte: O Autor

Os GCPs utilizados possuem 40x40cm e foram posicionados a 50 cm do solo, possibilitando a visualização dos mesmos nas imagens obtidas em campo. Foram utilizados 11 GCPs na área de trigo 2016 (Figura 8a) e 10 GCPs na área de soja 2016-2017 (Figura 8b). Foram coletadas as coordenadas geográficas do ponto central de cada GCP, as quais foram utilizadas posteriormente para correção do georreferenciamento. As áreas representadas na Figura 8 são diferentes para a cultura de trigo e soja pois foram coletadas inicialmente para diferentes finalidades, assim, as áreas representam apenas os locais de posicionamento dos GCPs. Na geração dos modelos de regressão para a estimativa de produtividade das diferentes culturas, a área de extração dos dados corresponde a área de estudo presente na Figura 2.

Figura 8 – Locais de posicionamentos dos GCPs nas áreas de trigo 2016 (a) e soja 2016-2017 (b).



Fonte: O Autor

A leitura das coordenadas geográficas dos locais de posicionamento dos GCPs, na safra de trigo 2016, foi realizada com o receptor GNSS Trimble R4 RTK e com o coletor de dados Juno 5D (Figura 9a). Na safra de soja 2016 – 2017, a leitura foi realizada com o receptor GNSS Trimble ProXT e com o coletor de dados Juno SB (Figura 9b).

Figura 9 - Receptores GNSS utilizados na leitura das coordenadas geográficas a) GNSS Trimble R4 RTK e coletor de dados Juno 5D b) GNSS Trimble ProXT e o coletor de dados Juno SB.



Fonte: O Autor

4.2.3 Imageamento

Estando o plano de voo configurado e os GCPs posicionados, o próximo passo foi a execução do plano de voo, iniciando com os testes nos sensores da aeronave (câmera, altitude, velocidade do vento, giroscópio, etc.), verificação do estado da bateria e lançamento.

As sobreposições lateral e longitudinal utilizadas em todos os voos foram de 60% e 75%, respectivamente. Visando diminuir o efeito da angulação do sol e as sombras das culturas, os voos foram realizados nos horários entre 10h00 e 16h00. Procurou-se também realizar o imageamento em dias claros e sem a incidência de nuvens. A Tabela 1 apresenta as datas nas quais foram realizados os voos na área de estudo, as câmeras utilizadas, resoluções espaciais e número de imagens por voo nas safras de trigo 2016 e soja 2016-2017.

Tabela 1 - Informações sobre os voos realizados nas safras de trigo (2016) e soja (2016-2017).

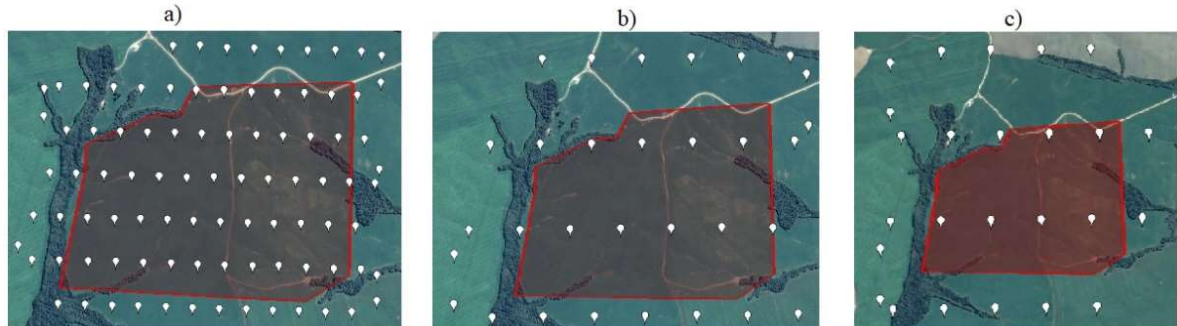
Data	Estádio Fenológico	Cultura	Câmera	Resolução Espacial cm/pixel	Nº Imagens	Altura Média (m)
10/10/2016	Florescimento	Trigo	RGB	10	39	353
				20	18	716
			NIR	10	46	296
				20	21	582
20/02/2017	Reprodutivo R 5.1	Soja	RGB	10	91	353
				20	32	716
				26	25	928

As datas presentes na Tabela 1 foram escolhidas buscando analisar cada cultura o mais próximo possível do seu ápice vegetativo, levando em consideração as condições do tempo e clima, para garantir a qualidade dos dados. Na safra de soja 2016-2017, não foi possível a utilização da câmera NIR, devido alguns problemas técnicos com as baterias da RPA e também pelo fato da área ser sobrevoada com três resoluções espaciais diferentes para câmera RGB.

Na Figura 10 é possível verificar os pontos de tomada das imagens na área de soja, nas resoluções de 10cm/pixel (Figura 10a), 20 cm/pixel (Figura 10b) e 26 cm/pixel (Figura 10c), com 91, 32 e 25 imagens, respectivamente. A câmera RGB, por gerar imagens com 18,2 MPixels, permite que a RPA sobrevoe mais alto em relação a câmera NIR (12 MPixels). Assim,

quanto maior a resolução espacial utilizada, maior o número de imagens coletadas, e menor a altura da RPA em relação ao solo.

Figura 10 – Pontos de coleta das imagens pela PRA, nos voos realizados na safra de soja 2016-2017, no dia 20/02/2017, nas resoluções espaciais de 10 (a), 20 (b) e 26 (c) cm/pixel.



Fonte: O Autor

4.3 GERAÇÃO DOS MAPAS DE REFLETÂNCIA

As imagens obtidas pela RPA *eBee* são salvas no cartão de memória da câmera durante o imageamento. As informações sobre o voo, coordenadas geográficas, angulação da aeronave em relação ao solo, altitude e altura de onde a imagem foi obtida, resolução espacial e demais informações do imageamento são salvas no *log* do voo, tanto na aeronave, quando no computador conectado a RPA pelo link de comunicação. Este *log* é utilizado na etapa de processamento dos dados de voo, realizado também no software *Emotion 2*, que irá gravar essas informações nos metadados EXIF de cada imagem. Assim, após cada voo, estes dados foram transferidos para o computador que estava controlando a aeronave remotamente.

Para as imagens obtidas com a câmera RGB, o formato de entrada e saída na etapa de processamento dos dados de voo é o JPEG. Quando utilizada a câmera NIR, o formato de entrada das imagens é RAW e o formato de saída é TIF. O processamento dos dados de voo é necessário para a geração dos DSMs, ortomosaicos e mapas de refletância, dentre outros índices que podem ser gerados por meio das imagens georreferenciadas.

A geração dos mapas de refletância e DSMs, responsáveis por fornecer as bandas espectrais e altitude, respectivamente, foi realizada no software Pix4D¹². Este software calcula as posições e orientações das imagens georreferenciadas pelo software *Emotion 2* de forma automática através da triangulação aérea (AAT – do inglês *Automatic Aerial Triangulation*) e do ajuste de feixes em bloco (BBA – do inglês *Bundle Block Adjustment*). Com base na nuvem de pontos 3D (Figura 11), obtidos durante a AAT e BBA, o DSM é gerado pela ligação destes pontos. Por fim o mapa de refletância é então criado a partir da projeção e combinação das

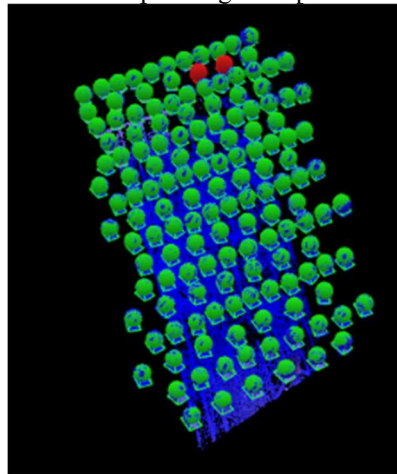
¹² Software comercial adquirido juntamente com a RPA *eBee*. Versão de teste disponível para download em <https://pix4d.com/product/pix4dmapper-photogrammetry-software/>

imagens originais com o DSM. Para os processamentos realizados no Pix4D, existem quatro opções de correções radiométricas, como apresentado a seguir:

- *No correction*: não será aplicada nenhuma correção radiométrica
- *Camera only*: as correções serão aplicadas para os parâmetros que estão escritos nos metadados EXIF e se relacionam com a câmera (ISO, pixel escuro, etc)
- *Camera and Sun Irradiance*: as correções serão aplicadas para os parâmetros da câmera a partir do ponto acima, bem como para as informações de irradiação do sol escritas na etiqueta EXIFF XMP.Camera.Irradiance.
- *Camera, Sun Irradiance and Sun angle*: as correções serão aplicadas para levar em consideração a posição do sol, bem como as informações da câmera e os dados da irradiância.

A opção selecionada no processamento realizado no Pix4D foi a “*câmera only*” pois as demais opções exigiam parâmetros que necessitavam de equipamentos específicos de coleta, os quais não estavam disponíveis.

Figura 11 – Nuvem de pontos gerada pelo software Pix4D.



Fonte: O Autor

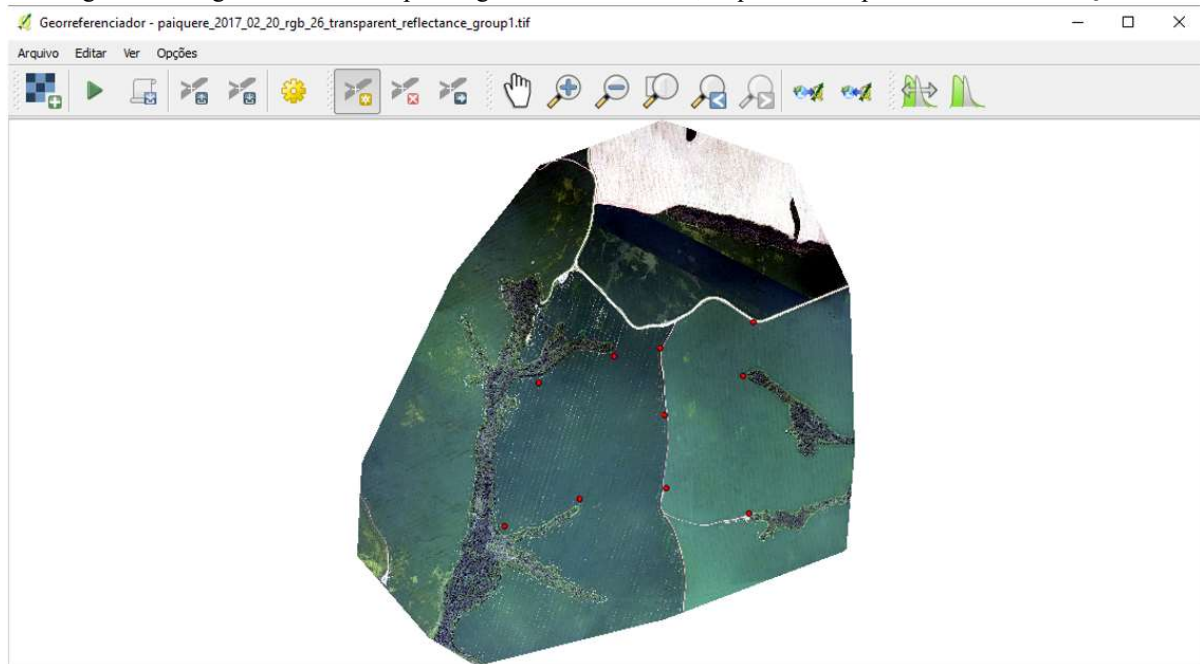
4.3.1 Correção do georreferenciamento dos mapas de refletância

Com base nas coordenadas geográficas dos GCPs, coletadas nas datas de imageamento, o georreferenciamento dos mapas de refletância foi corrigido no software QGIS 2.18.12¹³, um ambiente SIG que permite trabalhar com dados espaciais de diferentes fontes e formatos. A Figura 12 apresenta a estrutura de georreferenciamento desse software, que realiza a transformação das coordenadas geográficas da imagem conforme as coordenadas dos GCPs (pontos em vermelho no mapa de refletância da Figura 12).

¹³ Software de domínio público, disponível para download em: <https://www.qgis.org>

O georreferenciamento consiste em encontrar nas imagens os pontos nos quais foram posicionados os GCPs e atribuir a eles as coordenadas geográficas observadas em campo. Assim, quanto mais preciso o receptor GNSS utilizado na leitura das coordenadas dos GCPs, mais precisão será possível na etapa de correção do georreferenciamento. Na correção de todos os mapas de refletância utilizou-se o tipo de transformação Polinomial 1 e o método de reamostragem vizinho mais próximo, disponíveis no software QGIS.

Figura 12 - Algoritmo utilizado para o georreferenciamento, o qual está disponível no software QGIS.



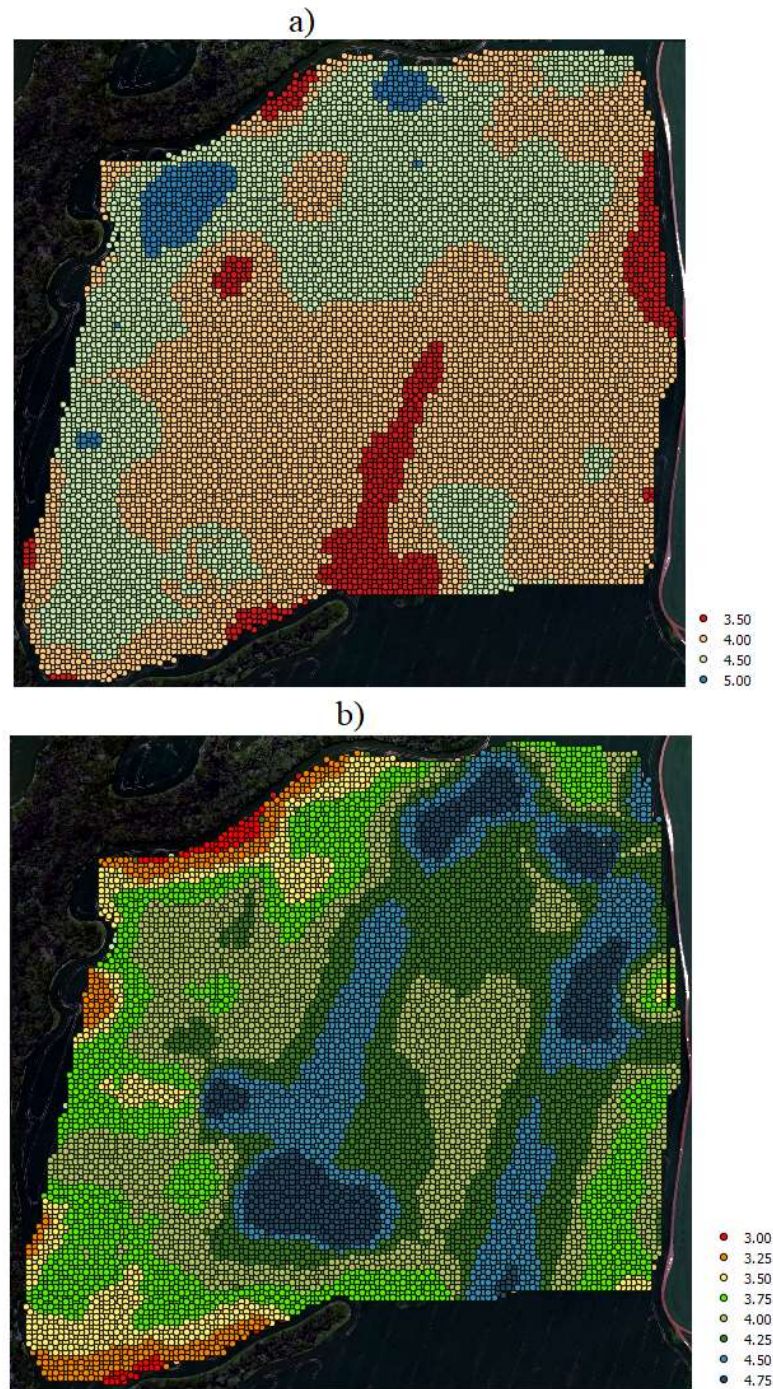
Fonte: O Autor

4.4 OBTENÇÃO DOS DADOS DE PRODUTIVIDADE EM CAMPO

Foram utilizadas as colhedoras John Deere 9670 STS com monitor de colheita GS2 2630 e a colhedora John Deere 9650 STS com o monitor GreenStar Original para a colheita das áreas sobrevoadas com a RPA. O software utilizado para geração dos mapas de produtividade, por meio dos dados de colheita, foi o GreenStar APEX 3.7.1.0 (software que dispõe de um gerenciamento completo de fazendas).

A colheita da área de trigo foi realizada no dia 15 de novembro de 2016. Na safra de soja, o talhão T05 foi colhido no dia 03 de abril de 2017. A Figura 13 mostra os mapas da produtividade gerados pelo software GreenStar APEX, por meio dos dados de colheita da safra de trigo 2016 (Figura 13a) e soja 2016-2017 (Figura 13b).

Figura 13 - Mapas de produtividade da safra de trigo 2016 (a) e de soja 2016-2017 (b), com valores em t/ha.



Fonte: O Autor

4.5 EXTRAÇÃO DOS DADOS DOS MAPAS DE REFLETÂNCIA E MAPAS DE PRODUTIVIDADE

A extração dos dados foi realizada no software QGIS versão 2.18.12, sendo a sequência de etapas mostradas na Figura 14. Foram gerados pontos regulares das áreas de interesse, sendo que, cada ponto estava a uma distância de aproximadamente 7 metros dos vizinhos. Esses pontos regulares foram transformados em polígonos utilizando o algoritmo

Polígonos de Voronoi do software QGIS. Dentro de cada polígono foram calculadas as médias de cada banda espectral, para cada mapa de refletância, utilizando o algoritmo Estatística Zonal. Posteriormente, foram gerados os centroides dos polígonos contendo as médias de cada banda espectral, R, G e B para a câmera RGB e R, G e NIR para a câmera NIR. O centroide corresponde ao ponto no interior do polígono que define o seu centro geométrico e que herda os mesmos atributos do polígono. A Figura 14 mostra a sequência de etapas citadas anteriormente, e como pode-se verificar pelas cores diferenciadas, a saída de uma etapa, é a entrada da etapa posterior (podendo haver outros arquivos de entrada).

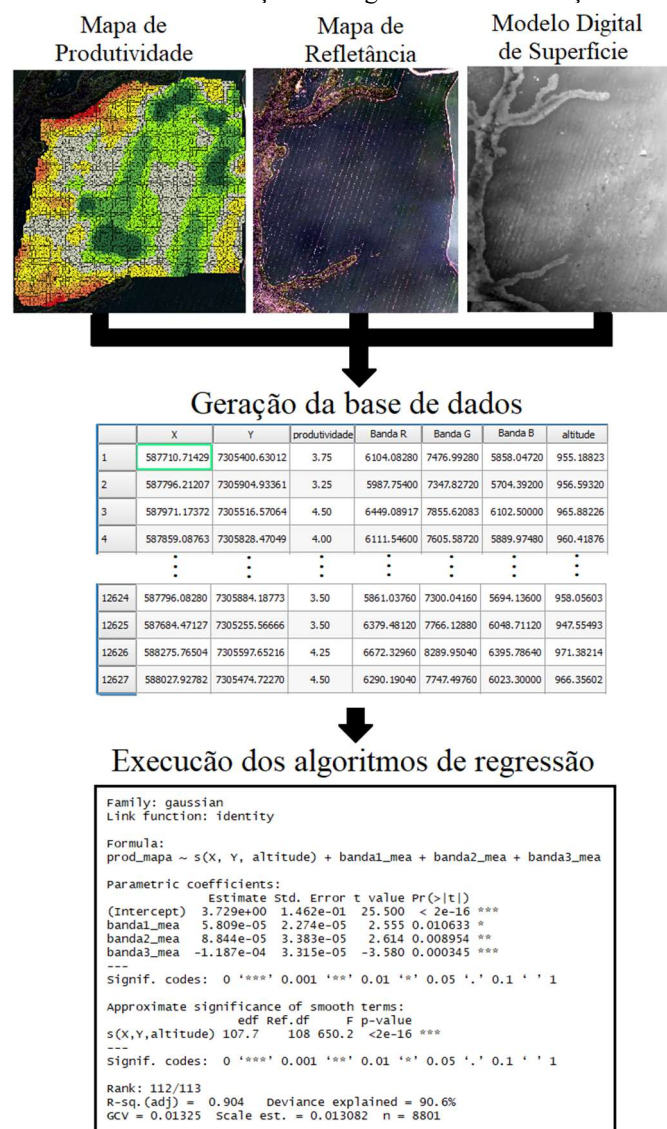
Figura 14 – Sequência de etapas para extração dos dados de refletância espectral, altitude e produtividade. Todas as etapas são realizadas no software QGIS.

	Descrição	Algoritmo	Entrada		Saída
1ª	Geração dos pontos regulares na área de interesse	Pontos Regulares	Densidade de pontos na área de interesse		Pontos regulares da área de interesse
2ª	Conversão de pontos em polígonos	Polígonos de Voronoi	Pontos regulares da área de interesse		Polígonos da área de estudo
3ª	Média dos valores da Banda R	Estatística Zonal	Mapa de refletância	Polígonos da área de estudo	Polígonos com as médias da banda R
4ª	Média dos valores da Banda G	Estatística Zonal	Mapa de refletância	Polígonos com as médias da banda R	Polígonos com as médias das bandas R e G
5ª	Média dos valores da Banda B	Estatística Zonal	Mapa de refletância	Polígonos com as médias das bandas R e G	Polígonos com as médias da banda R, G e B
6ª	Converter polígonos em pontos	Centroide de Polígonos	Polígonos com as médias das bandas R, G e B		Pontos com as médias das bandas R, G, B
7ª	Extração dos valores de altitude e produtividade	Point Sampling Tools	DSM e mapa de produtividade	Polígonos com médias das bandas R, G, B	Pontos com as médias das bandas R, G, B e valores de altitude e produtividade
8ª	Salvar pontos no formato CSV	Salvar Arquivo	Pontos com as médias das bandas R, G, B e valores de altitude e produtividade		Base de dados final

Fonte: O Autor

Estando as informações de refletância espectral salvas nos pontos amostrais, estes pontos foram utilizados para a extração dos valores de produtividade e altitude, presentes dos mapas de produtividade e DSMs, respectivamente. A extração desses dados foi realizada com o plugin *Point Sampling Tools* do software QGIS. Esse plugin utiliza os pontos amostrais para extrair os dados das diferentes camadas importadas no QGIS. A Figura 15 apresenta a sequência de etapas desde a entrada de dados no QGIS por meio dos mapas de produtividade, mapas de refletância e DSMs, utilizados para geração da base de dados contendo os valores de produtividade, bandas espectrais e altitude, e, por fim, a execução das técnicas de MD.

Figura 15 - Sequência de etapas para obtenção de dados de produtividade, bandas espectrais e altitude, geração da base de dados e execução dos algoritmos de mineração de dados.



Fonte: O Autor

O plugin *Point Sampling Tools* salva os dados extraídos no formato *shapefile* (do tipo ponto). Sendo assim, esses pontos precisaram ser salvos posteriormente nos formatos de entrada dos algoritmos de MD. Assim, cada ponto amostral se torna uma instância na base de dados

contendo valores das médias de refletância das bandas R, G e B (para dados da câmera RGB) ou R, G e NIR (para dados da câmera NIR), assim como valores de altitude e produtividade.

4.6 MINERAÇÃO DE DADOS

As técnicas da MD clássica e espacial foram executadas no software *RStudio*¹⁴, um ambiente desktop que dispõe de uma interface para a ferramenta estatística R¹⁵ (utilizada na versão 3.3.3). O *RStudio* facilita a importação e visualização de dados, importação de pacotes, visualização da descrição de pacotes e métodos, dentre outros componentes que tornam a utilização do R dinâmica e autoexplicativa.

4.6.1 Execução da Mineração de dados Clássica

Para execução das técnicas da MD clássica, ou seja, aquelas que não consideram a dimensão espacial dos dados na descoberta de padrões, as bases de dados foram importadas no software *RStudio* no formato CSV (separado por vírgulas). Os dados foram divididos em dois subconjuntos, um para treinamento e outro para teste, sendo utilizados 70% e 30% para cada um, respectivamente (Quadro 7). Assim, os modelos foram criados utilizando a base de treinamento, e validados utilizando a base de teste.

Quadro 7 - Script para definição da pasta de trabalho contendo as bases de dados e para a divisão da base em conjuntos treinamento e teste.

```
> #Definindo ambiente de trabalho
> setwd('C:/workspace/')
> dados = read.csv("dados.csv")
> set.seed(100)
> #Dividindo os dados em dois subconjuntos, um com 70% e outro com 30%
> base <- sample(2, nrow(dados), replace = TRUE, prob = c(0.7, 0.3))
> #Atribuindo a base de treinamento a variável treinamento
> treinamento <- dados[base == 1, ]
> #Atribuindo a base de teste a variável teste
> teste <- dados[base == 2, ]
```

Foram definidas quatro fórmulas utilizadas como parâmetro para a execução das técnicas de MD no software R, como apresentado no Quadro 8. Essas fórmulas representam qual é o atributo meta da base de dados (PRODUTIVIDADE) e quais são os atributos preditores. Por exemplo, sendo “~” o símbolo que separa o atributo meta (antes do símbolo) dos atributos preditores (após o símbolo), em (PRODUTIVIDADE ~ R + G + B) tem-se que o atributo meta é PRODUTIVIDADE e os atributos preditores são R, G e B. Das fórmulas

¹⁴ <https://www.rstudio.com/>

¹⁵ <https://www.r-project.org/>

apresentadas no Quadro 8, duas representam os dados obtidos pela câmera RGB, e duas os dados obtidos pela câmera NIR, sendo uma para o modelo sem incorporação da altitude do terreno e outra com a utilização da altitude na estimativa.

Quadro 8 - Fórmulas utilizadas na MD clássica para a estimativa da produtividade com os dados das câmeras RGB e NIR, com e sem a utilização da altitude

```
# Para estimar a produtividade com as bandas espectrais da câmera RGB
> formulaRGB <- PRODUTIVIDADE ~ R + G + B

# Para estimar a produtividade com as bandas espectrais da câmera RGB e altitude
> formulaAltitudeRGB <- PRODUTIVIDADE ~ R + G + B + altitude

# Para estimar a produtividade com as bandas espectrais da câmera NIR
> formulaNIR <- PRODUTIVIDADE ~ R + G + NIR

# Para estimar a produtividade com as bandas espectrais da câmera NIR e altitude
> formulaAltitudeNIR <- PRODUTIVIDADE ~ R + G + NIR + altitude
```

As técnicas de MD clássica utilizadas foram a Regressão Linear Múltipla (RLM), a SVR e RNA. O Quadro 9 apresenta a sequência de etapas para a execução das técnicas da MD clássica no software R, desde a instalação e importação dos pacotes, execução e validação com o método *predict*.

Quadro 9 - Comandos e parâmetros para execução das técnicas da MD clássica no software R

```
> #Execução da técnica de regressão linear múltipla
> ModeloRLM <- lm(formula, treinamento)
> validaRLM <- predict(ModeloRLM, teste)

> #Instalação e importação do pacote para execução da SVR
> install.packages("kernlab")
> library(kernlab)

> #Execução da técnica de máquina vetores de suporte para regressão
> ModeloSVR <- ksvm(formula, data=treinamento, type="eps-svr", kernel="rbfdot", c
=5, prob.model=TRUE, kpar="automatic")
> validaSVR <- predict(ModeloSVR, teste)

> #Instalação e importação do pacote para execução da RNA
> install.packages("nnet")
> library(nnet)

#Execução da técnica de redes neurais artificiais
> ModeloRNA <- nnet(formula, data=treinamento, size=10, linout=T, rang=0.07,decay
=5e-2,maxit=3000)
> validaRNA <- predict(model, teste)
```

Como pode ser observado no Quadro 9, para a execução da RLM não foi necessária a importação de pacotes adicionais, sendo utilizado o método *lm* (do inglês *Linear Model*) no software R, esse método já é disponibilizado na versão de download do R. Já para a execução da técnica SVR utilizou-se o método *ksvm*, do tipo *eps-svr*, disponível no pacote “*kernlab*”. Foram utilizadas as opções de *kernel* *rbfdot*, *polydot*, *vanilladot*, *laplacedot* e *besseldot*, que correspondem às funções de kernel *Gaussian Radial Basis*, *Polynomial*, *Linear*, *Laplacian* e *Bessel*, respectivamente.

Em relação a RNA, foi utilizada uma arquitetura multicamadas, com uma camada escondida e o algoritmo de propagação *FeedForward*, sem retropropagação. Para execução da RNA foi utilizado o método *nnet*. O parâmetro *size*, que representa o número de neurônios na camada escondida, foi fixado em 10, e o parâmetro *maxit*, que define o máximo de interações do algoritmo até convergir em um resultado, foi definido com o valor 3000. Ambos os parâmetros foram definidos depois de uma análise de valores que possibilitassem a obtenção de melhores resultados.

Para avaliação dos modelos gerados por meio das técnicas da MD clássica foi utilizado o método *predict*. Esse método utiliza o modelo gerado na etapa de treinamento para realização da estimativa de produtividade com conjunto de dados de teste.

4.6.2 Execução da Mineração de dados Espaciais

Inicialmente foram testados três modelos de regressão espacial frequentemente utilizados como técnicas de MDE, sendo eles o CAR, SAR e GAM. No entanto, os métodos SAR e CAR apresentaram limitações, relacionadas a confiabilidade dos modelos gerados. Não foi encontrado nenhum software que fornecesse a possibilidade de executar o método CAR com a utilização de bases de treinamento e teste, visto que este método exige que o conjunto de dados de entrada seja simétrico. Já o método SAR possibilita a divisão da base em conjuntos de treinamento e teste, porém, nas execuções realizadas o desempenho do método na etapa de teste apresentou coeficientes de correlação (*r*) inferiores a 0,45, não apresentando resultados significativos. Com isso, como técnica de MDE foi utilizado apenas o método GAM.

Assim como para os algoritmos da MD clássica, para o algoritmo GAM os dados foram importados para o *RStudio* no formato CSV e os subconjuntos de treinamento e teste foram utilizados conforme as subdivisões do Quadro 7. No entanto, o formato de entrada do atributo meta e atributos preditores é diferente em relação a execução das técnicas da MD clássica, pois é necessário acrescentar dados da posição geográfica (latitude[Y], longitude[X]), como mostrado no Quadro 10. A função “*s(x, y)*” é responsável pela suavização

bidimensional, com base nas coordenadas geográficas. Nos modelos com a incorporação da altitude do terreno, os valores correspondentes a altitude foram inseridos juntamente com as coordenadas geográficas, permitindo uma suavização tridimensional($s(x, y, altitude)$).

Quadro 10 - Fórmulas utilizadas no método GAM

```
> # Para estimar a produtividade com as bandas espectrais da câmera RGB
> FormulaRGB <- PRODUÇÃO ~ s(X, Y) + R + G + B

> #Para estimar a produtividade com as bandas espectrais da câmera RGB e altitude
> FormulaAltitudeRGB <- PRODUÇÃO ~ s(X, Y, altitude) + R + G + B

> # Para estimar a produtividade com as bandas espectrais da câmera NIR
> FormulaNIR <- PRODUÇÃO ~ s(X, Y) + R + G + NIR

> #Para estimar a produtividade com as bandas espectrais da câmera RGB e altitude
> FormulaAltitudeNIR <- PRODUÇÃO ~ s(X, Y, altitude) + R + G + NIR
```

O modelo GAM disponibilizado para o software R é ajustado usando o algoritmo de pontuação local, que se encaixa iterativamente em modelos aditivos ponderados pelo back-fitting. O algoritmo de back-fitting é um método de Gauss-Seidel para a montagem de modelos aditivos, suavizando iterativamente os resíduos parciais, por meio da função de suavização $s(x, y)$, por exemplo. O algoritmo separa a parte paramétrica da parte não paramétrica do ajuste e ajusta a parte paramétrica usando mínimos quadrados lineares ponderados dentro do algoritmo de back-fitting. O modelo GAM possui diferentes funções de ajuste que podem ser utilizadas na geração dos modelos de regressão espaciais, dentre elas a *binomial*, *gaussian*, *gamma*, *Poisson*, *quasi*, etc, disponíveis no software R. Neste trabalho, foram utilizadas as funções *gaussian*, *Poisson* e *quasi*, por apresentarem melhores resultados. Os demais parâmetros da função de suavização são ajustados automaticamente pelo método GAM ou o usuário pode informar um valor fixo, sendo utilizada a primeira opção em todas as execuções. Para execução do algoritmo GAM, foi necessária a instalação e importação do pacote “gam” no software *RStudio*. Os passos para geração e validação do modelo são mostrados no Quadro 11.

Quadro 11 - Script para execução do algoritmo GAM

```
> #Instalação e importação do pacote para execução do método GAM
> install.packages("gam")
> library(gam)
> #Execução do método GAM
> ModeloGAM <- gam(FormulaRGB, family= "gaussian", data = treinamento)
> validaGAM <- predict.gam(ModeloGAM, newdata=teste)
```

Para fins de comparação da utilização da MDE, assim como das técnicas da MD clássica, o método de krigagem também foi utilizado como técnica de regressão espacial. Este método é tradicionalmente utilizado em SIGs para realização de interpolação de dados e em

análises de dados espaciais diversos. Foi escolhido o pacote `gstat` para a execução da krigagem no software RStudio, sendo utilizadas as mesmas fórmulas disponíveis no Quadro 8 para indicar o formato de entrada de dados.

O Quadro 12 mostra os passos para a execução da krigagem, lembrando que a importação e a divisão da base de dados em conjuntos de treinamento e teste são realizadas anteriormente. O parâmetro *nmax* do método `Gstat`, que indica o raio dos pontos amostrais vizinhos que serão considerados na estimativa em um determinado local, foi fixado em 15, depois de uma análise considerando a variação dos valores de produtividade na base de dados e dos resultados apresentados com a variação deste parâmetro.

Quadro 12 - Script para execução do algoritmo de Krigagem com método `Gstat`

```
> #Instalação e importação do pacote para execução do método de Krigagem
> install.packages("gstat")
> library(gstat)

> #indicando quais atributos correspondem as coordenadas geográficas na base de
> dados de treinamento e teste
> coordinates(treinamento) <- ~X+Y
> coordinates(teste) <- ~X+Y

> #Execução do método de Krigagem
> Modelo <- gstat(FormulaRGB, data = treinamento, nmax = 15)
> validacao <- predict(Modelo, newdata=teste)
```

4.7 ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

Como medida de análise dos modelos de regressão gerados pelas diferentes técnicas de MD, utilizou-se o coeficiente de correlação (r), o coeficiente determinação (R^2) e a raiz do erro quadrático médio (RMSE), expressos pelas Equações 11, 12 e 13, respectivamente. Estas medidas foram calculadas sobre os valores de estimativa gerados na etapa de validação dos modelos de regressão (retornados pelo método *predict* do software R).

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (11)$$

onde $x_1 \dots x_N$ e $y_1 \dots y_N$ representam os valores das duas variáveis para correlacionar e $\bar{x}$ e $\bar{y}$ são as médias aritméticas dessas variáveis.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (12)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (13)$$

onde y_i representa o valor de produtividade observado em campo, $\hat{y}_i$ o valor predito, $\bar{y}_i$ a média da produtividade das culturas e N é o número de pontos amostrais (instâncias) na base de dados. Os valores de RMSE foram expressos em toneladas por hectare (t/ha), visto que é a unidade de medida utilizada na base de dados para representação da produtividade das culturas de trigo e soja.

5 RESULTADOS

A seguir serão apresentados os resultados para a estimativa de produtividade de trigo (sessão 5.1) e soja (sessão 5.2), e também a discussão com trabalhos correlatos (sessão 5.3).

5.1 RESULTADOS OBTIDOS NA ESTIMATIVA DA PRODUTIVIDADE DA CULTURA DE TRIGO

A Tabela 6 apresenta a estatística descritiva dos dados obtidos na safra de trigo 2016, contendo valores de mínimo (min), máximo (max), média, desvio padrão (DP) e coeficiente de variação (CV) das bandas espectrais e da altitude do terreno. Estas informações são mostradas para as diferentes resoluções e câmeras utilizadas. Como já era esperado, os valores de refletância das bandas espectrais obtidos pela câmera NIR são consideravelmente maiores que os obtidos pela câmera RGB, visto que possuem diferentes resoluções radiométricas.

Tabela 2 – Estatística descritiva dos dados da cultura de trigo safra 2016 e dos valores espectrais para as câmeras e resoluções espaciais utilizadas

Câmera/Resolução espacial (cm/pixel)	Atributos	Dados estatísticos				
		Min	Max	Média	DP	CV
RGB – 10	Banda R	3659,21	11001,5	6139,78	606,93	9,88
	Banda G	4750,10	9900,192	7117,97	553,27	7,77
	Banda B	3351,72	8246,24	4838,47	456,62	9,43
	Altitude	941,41	980,56	963,56	7,99	0,83
RGB – 20	Banda R	5443,63	11893,45	6820,47	648,13	9,50
	Banda G	6707,96	10804,47	7880,158	606,05	7,69
	Banda B	4779,39	9240,30	5905,30	492,91	8,35
	Altitude	941,44	985,41	965,66	8,28	0,86
NIR – 10	Banda R	12499,04	35062,29	22995,84	3552,071	15,44
	Banda G	14529,34	36234,13	26679,96	4140,23	15,51
	Banda NIR	38352,81	63501,04	55259,91	2327,31	4,21
	Altitude	939,99	979,23	962,42	8,17	0,85
NIR – 20	Banda R	22808,79	38719,04	26125,04	1065,50	4,08
	Banda G	27085,48	40358,36	30578,21	1114,09	3,64
	Banda NIR	42592,54	61527,26	56616,72	1845,03	3,26
	Altitude	939,99	979,23	962,89	8,29	0,86
Produtividade (t/ha)		3,5	5,0	4,18	0,32	7,75

A Tabela 3 apresenta os resultados da estimativa da produtividade de trigo, por meio das técnicas da MD clássica MLR, RNA e SVR, utilizando os dados obtidos pelas câmeras RGB e NIR. Os melhores resultados para os dados obtidos com a câmera RGB foram apresentados pela técnica SVR, utilizando o kernel de Laplace, com r de até 0,72 e R^2 de até 0,52, para as duas resoluções espaciais empregadas. Isso indica que os dados das bandas espectrais R, G e B, associados aos valores de altitude do terreno foram capazes de explicar 52% da variação da produtividade de trigo no talhão.

Utilizando as bandas espectrais R, G e NIR e a altitude obtidas com a câmera NIR, no voo com resolução espacial de 10 cm/pixel, foram obtidos valores de $r = 0,73$ e $R^2 = 0,54$ com a técnica SVR e o kernel de Laplace. Assim, a SVR gerou um modelo com capacidade de explicar 54% da variação da produtividade de trigo. Resultado similar foi alcançado com a estimativa utilizando os dados obtidos por meio da câmera RGB, no voo com resolução espacial de 20 cm/pixel.

Tabela 3 - Resultados da execução das técnicas da MD clássica para a estimativa da produtividade do trigo por meio dos dados obtidos pela câmera RGB e NIR, nas resoluções espaciais de 10 e 20 cm/pixel

Técnica	Câmera	RE (cm/pixel)	Sem altitude			Com altitude		
			r	R ²	RMSE	R	R ²	RMSE
MLR	RGB	10	0,36	0,13	0,30	0,37	0,14	0,31
		20	0,47	0,21	0,29	0,46	0,21	0,29
	NIR	10	0,22	0,05	0,32	0,23	0,05	0,32
		20	0,19	0,04	0,32	0,26	0,07	0,32
RNA	RGB	10	0,44	0,19	0,29	0,50	0,25	0,28
		20	0,51	0,26	0,28	0,42	0,18	0,29
	NIR	10	0,42	0,17	0,29	0,49	0,24	0,28
		20	0,34	0,12	0,30	0,38	0,14	0,30
SVR - Gaussian RBF	RGB	10	0,42	0,18	0,30	0,64	0,41	0,25
		20	0,48	0,23	0,28	0,62	0,38	0,25
	NIR	10	0,46	0,21	0,29	0,64	0,41	0,25
		20	0,38	0,14	0,31	0,64	0,41	0,25
SVR - Polynomial	RGB	10	0,36	0,13	0,32	0,36	0,13	0,31
		20	0,46	0,21	0,28	0,45	0,20	0,29
	NIR	10	0,17	0,03	0,35	0,17	0,03	0,35
		20	0,08	0,007	0,35	0,26	0,07	0,33
SVR – Linear	RGB	10	0,36	0,13	0,32	0,36	0,13	0,31
		20	0,46	0,21	0,29	0,45	0,20	0,29
	NIR	10	0,22	0,05	0,35	0,18	0,03	0,36
		20	0,19	0,04	0,35	0,25	0,06	0,34
SVR – Laplace	RGB	10	0,46	0,21	0,30	0,72	0,52	0,23
		20	0,50	0,25	0,28	0,72	0,52	0,22
	NIR	10	0,49	0,24	0,29	0,73	0,54	0,22
		20	0,44	0,19	0,27	0,72	0,52	0,22
SVR – Bessel	RGB	10	0,36	0,13	0,31	0,49	0,24	0,30
		20	0,45	0,20	0,29	0,49	0,24	0,28
	NIR	10	0,31	0,10	0,32	0,44	0,20	0,29
		20	0,29	0,08	0,32	0,43	0,17	0,30

Os dados obtidos do voo com resolução espacial de 20 cm/pixel apresentaram melhores resultados em comparação com os obtidos na resolução de 10 cm/pixel, quando utilizando apenas as bandas espectrais R, G e B para realizar a estimativa da produtividade do trigo, como pode ser visto no Tabela 3. No entanto, quando estas bandas foram associadas a

altitude do terreno, utilizando os kernels Gaussian RBF e Laplace, os dados do voo com resolução de 10cm/pixel apresentaram capacidades de estimativa similares.

Os resultados obtidos com as técnicas MLR e SVR-Linear, quando utilizando os dados obtidos pela câmera RGB, não diferem, como pode ser verificado na Tabela 3, o que já era esperado, visto que ambas buscam explicar a variação da produtividade por meio da regressão linear multivariada. Estas técnicas não apresentaram resultados significativos, o que indica que a variação dos atributos preditores em relação a variação da produtividade não ocorre de forma linear. Já para os resultados obtidos com os dados da câmera NIR, as técnicas MLR e SVR-Linear apresentaram comportamentos não esperados, obtendo valores dos coeficientes r e R^2 diferentes.

A técnica RNA, quando utilizada na estimativa da produtividade do trigo por meio das bandas espectrais R, G e B, obteve desempenho similar ao obtido pela técnica SVR- Laplace (Tabela 3). Em contrapartida, a técnica SVR-Laplace possibilitou melhores ajustes nos modelos de estimativa quando incorporada a altitude, com consequente aumento nos coeficientes r e R^2 , enquanto a RNA não apresentou ganho significativo com a utilização da altitude.

Em comparação com os resultados obtidos com os dados da câmera RGB, que apresentou melhor capacidade de estimativa para os dados extraídos do voo com resolução espacial de 20 cm/pixel, para câmera NIR, as melhores estimativas foram obtidas com os dados do voo com resolução espacial de 10 cm/pixel, na maioria das técnicas testadas.

Os valores de RMSE das técnicas da MD clássica, tanto para os resultados obtidos para câmera RGB, quanto para a da câmera NIR foram baixos, variando de 0,22 (220 kg/ha) a 0,36 (360 kg/ha). Esses valores indicam um erro de 4,4% e 7,2%, respectivamente, em relação ao valor máximo de produtividade de trigo obtido na safra de 2016.

Os resultados obtidos com os algoritmos de MDE e com a técnica de geoestatística krigagem são mostrados na Tabela 4. Pode-se verificar que, assim como nos resultados obtidos com as técnicas da MD clássica, a altitude do terreno proporcionou um ganho significativo no desempenho do método GAM, quando utilizando as funções de ajuste Gaussian e Quasi, aumentando de 52% para 81% a explicação da variação da produtividade no talhão. Para o método Gstat, que utiliza a krigagem na geração dos modelos de regressão, assim como para o método GAM com a função Poisson, o uso da altitude não proporcionou a obtenção de melhores resultados, para os dados obtidos de ambas as câmeras (RGB e NIR). Com isso, sabendo que para a maioria das técnicas da MD clássica, assim como para a regressão espacial com o método GAM Gaussian e Quasi, os melhores resultados foram obtidos quando incorporada a altitude

nos modelos de estimativa, os métodos GAM – Poisson Gstat não apresentaram resultados coerentes.

Tabela 4 - Resultados obtidos com a execução das técnicas de MDE e krigagem para a estimativa da produtividade do trigo por meio dos dados obtidos com as câmeras RGB e NIR

Algoritmos	Câmera	RE (cm/pixel)	Sem altitude			Com altitude		
			r	R ²	RMSE	R	R ²	RMSE
GAM – Gaussian	RGB	10	0,72	0,52	0,23	0,90	0,81	0,14
		20	0,72	0,52	0,22	0,90	0,81	0,14
	NIR	10	0,71	0,51	0,23	0,90	0,81	0,14
		20	0,71	0,50	0,23	0,89	0,79	0,15
GAM – Poisson	RGB	10	0,67	0,44	2,76	0,70	0,49	2,76
		20	0,65	0,43	2,76	0,68	0,46	2,76
	NIR	10	0,65	0,42	2,76	0,69	0,48	2,76
		20	0,65	0,42	2,77	0,69	0,48	2,77
GAM - Quasi	RGB	10	0,72	0,52	0,23	0,90	0,81	0,14
		20	0,72	0,52	0,22	0,89	0,80	0,14
	NIR	10	0,71	0,51	0,23	0,90	0,81	0,14
		20	0,71	0,50	0,23	0,89	0,79	0,15
Gstat	RGB	10	0,94	0,88	0,11	0,94	0,87	0,12
		20	0,93	0,87	0,11	0,93	0,86	0,12
	NIR	10	0,94	0,88	0,12	0,93	0,86	0,12
		20	0,94	0,88	0,12	0,94	0,88	0,12

Os valores de RMSE obtidos com as técnicas de MDE, em sua maioria, foram menores dos obtidos pelas técnicas da MD clássica. Se considerando o método GAM com a função de ajuste gaussiana, o valor de RMSE foi de 0,14 t/ha, que equivale a um erro de 2,8% em relação ao valor máximo de produtividade de trigo obtida. No entanto, quando utilizando a função de ajuste Poisson, o valor de RMSE foi de 2,76 t/ha, que equivale a um erro de 55% em relação a produtividade máxima de trigo no talhão, sendo assim, o método GAM – Poisson não foi adequado para a estimativa da produtividade de trigo.

5.2 RESULTADOS OBTIDOS NA ESTIMATIVA DA PRODUTIVIDADE DA CULTURA DA SOJA

Na Tabela 5 pode-se observar a estatística descritiva dos dados obtidos na safra de soja 2016-2017. Os dados apresentam pouca variação em relação as médias de seus valores, sendo a altitude das amostras o atributo com menores valores de CV, não chegando a 1% de variação, enquanto as bandas espectrais apresentam CV entre 3% e 6%. Assim como nos resultados obtidos na estimativa da produtividade de trigo, na estimativa da produtividade da soja a inclusão dos valores de altitude do terreno nos modelos de regressão melhorou o desempenho dos algoritmos da MD clássica (Tabela 6).

Tabela 5 - Estatística descritiva dos dados da cultura de soja safra 2016-2017 e dos valores espectrais para as câmeras e resoluções espaciais utilizadas

Resolução espacial (cm/pixel)	Atributos	Dados estatísticos				
		Min	Max	média	DP	CV (%)
10	Banda R	4794,47	8088,36	5705,85	331,22	5,80
	Banda G	6047,10	9351,07	7056,10	373,14	5,29
	Banda B	3869,11	6015,76	4580,54	238,64	5,21
	Altitude	94554	982,57	966,94	8,01	0,83
20	Banda R	5464,96	9195,27	6171,87	290,82	4,71
	Banda G	6954,32	9529,66	7596,92	249,81	3,29
	Banda B	4880,09	7560,19	5582,00	233,11	4,18
	Altitude	943,34	994,69	965,82	8,21	0,85
26	Banda R	5674,82	9708,65	6406,03	320,47	5,00
	Banda G	7051,88	9861,38	7901,32	374,43	4,74
	Banda B	5430,27	8323,75	6118,33	251,54	4,11
	Altitude	941,52	987,80	964,90	7,87	0,81
Produtividade (t/ha)		3,0	4,75	4,07	0,37	9,17

Tabela 6 - Resultados da execução das técnicas de mineração de dados clássicas para a estimativa de produtividade da soja utilizando a câmera RGB

Técnica	RE (cm/pixel)	Sem altitude			Com altitude		
		R	R ²	RMSE	R	R ²	RMSE
RLM	10	0,19	0,04	0,37	0,56	0,31	0,31
	20	0,32	0,10	0,35	0,53	0,28	0,32
	26	0,36	0,13	0,35	0,56	0,31	0,31
RNA	10	0,32	0,10	0,36	0,63	0,39	0,29
	20	0,41	0,16	0,34	0,67	0,44	0,28
	26	0,46	0,22	0,33	0,69	0,48	0,27
SVR - Gaussian RBF	10	0,38	0,14	0,35	0,77	0,60	0,24
	20	0,51	0,26	0,32	0,79	0,63	0,23
	26	0,53	0,28	0,32	0,81	0,65	0,22
SVR - Polynomial	10	0,19	0,04	0,37	0,55	0,30	0,31
	20	0,32	0,10	0,36	0,53	0,29	0,32
	26	0,34	0,12	0,35	0,55	0,31	0,31
SVR - Linear	10	0,19	0,04	0,37	0,55	0,30	0,31
	20	0,32	0,10	0,36	0,53	0,28	0,32
	26	0,34	0,12	0,35	0,56	0,31	0,31
SVR - Laplacian	10	0,39	0,15	0,35	0,82	0,67	0,21
	20	0,51	0,26	0,33	0,81	0,66	0,22
	26	0,55	0,30	0,32	0,83	0,69	0,21
SVR - Bessel	10	0,25	0,06	0,36	0,70	0,50	0,27
	20	0,43	0,19	0,34	0,74	0,55	0,25
	26	0,44	0,20	0,33	0,74	0,54	0,25

Para a estimativa da produtividade da soja, assim como para o trigo, os melhores resultados dos algoritmos da MD clássica foram obtidos com a técnica SVR, utilizando o kernel

de Laplace, quando utilizando as médias das bandas espectrais e a altitude na estimativa (Tabela 6). Considerando o R^2 de 0,69, os atributos utilizados foram capazes de explicar até 69% da variação da produtividade da soja no talhão T05. A estimativa da soja apresentou resultados superiores aqueles obtidos para o trigo.

Dentre as três diferentes resoluções espaciais sobrevoadas na cultura da soja, os dados obtidos das imagens com resolução de 26 cm/pixel possibilitaram a obtenção de resultados iguais ou superiores as demais resoluções, na maioria das técnicas da MD clássica utilizadas. Isso indica que sobrevoar com resoluções mais altas, além de aumentar a duração de voo e o custo computacional, tanto do processamento, quanto da extração dos dados das imagens, não possibilitou a obtenção de resultados com maior capacidade de estimativa.

A Tabela 7 apresenta os resultados obtidos pelos algoritmos de regressão espacial para a estimativa da produtividade da soja. Como pode ser observado, os resultados apresentados pela MDE foram melhores que os obtidos pelos algoritmos da MD clássica (Tabela 6), com valores de r até 0,96 e R^2 até 0,92 na utilização do método GAM. A MDE proporcionou melhores estimativas da produtividade da soja, em relação a estimativa da produtividade de trigo, variado de 81% para 92% de explicação da variação da produtividade de trigo e soja, respectivamente. Os valores de RMSE também foram inferiores na estimativa da produtividade da soja, com erros de 110 kg/ha, que representa 2,32% do valor máximo da produtividade no talhão.

Tabela 7 - Resultados obtidos pelos métodos de MDE e de krigagem para a estimativa da produtividade da soja utilizando a câmera RGB

Algoritmos	RE (cm/pixel)	Sem altitude			Com altitude		
		R	R ²	RMSE	R	R ²	RMSE
GAM - Gaussian	10	0,87	0,75	0,19	0,96	0,91	0,11
	20	0,87	0,75	0,19	0,96	0,92	0,11
	26	0,87	0,75	0,19	0,95	0,91	0,12
GAM - Poisson	10	0,85	0,72	2,69	0,86	0,73	2,69
	20	0,85	0,72	2,69	0,86	0,73	2,69
	26	0,85	0,72	2,69	0,85	0,72	2,69
GAM - Quasi	10	0,87	0,75	0,19	0,96	0,92	0,11
	20	0,87	0,75	0,19	0,96	0,92	0,11
	26	0,87	0,75	0,19	0,95	0,91	0,11
Gstat	10	0,97	0,94	0,09	0,97	0,94	0,09
	20	0,97	0,94	0,09	0,97	0,94	0,09
	26	0,97	0,94	0,09	0,96	0,92	0,10

De maneira semelhante aos coeficientes obtidos na estimativa da produtividade de trigo, a inclusão da altitude do terreno nas técnicas de regressão Gstat e GAM-Poisson não melhorou a capacidade estimativa dos modelos.

5.3 DISCUSSÃO DOS RESULTADOS

As três abordagens utilizadas neste trabalho para geração dos modelos de regressão, por meio das técnicas da MD clássica, MDE e geoestatística possibilitaram a obtenção de bons resultados. A MDE e krigagem apresentaram resultados semelhantes, no entanto, apenas para as técnicas da MD e MDE os resultados obtidos apresentaram um padrão de comportamento, com maior capacidade de explicação da variação da produtividade, mostrando o impacto das bandas espectrais e a influência da altitude na geração de modelos de produtividade de soja e trigo. Em testes realizados com a técnica de krigagem, foi observado que, independente de quantas ou quais bandas espectrais fossem utilizadas na estimativa, os valores altos de coeficiente de correlação e determinação se mantinham os mesmos. Isso pode ter ocorrido pelo fato de não haver grande variação da produtividade das culturas nas safras analisadas, visto que o coeficiente de variação de produtividade para ambas as culturas não atingiu 10%.

Diferentes metodologias são aplicadas para a estimativa de produtividade de culturas agrícolas, no entanto, a maioria delas fazem uso de imagens de satélite, sendo realizadas estimativas em grandes extensões territoriais, tanto regionais, quanto nacionais. As estimativas utilizando RPAs vêm se desenvolvendo nos últimos anos, mas ainda existem poucos trabalhos publicados com esta finalidade. A Tabela 8 mostra os principais resultados obtidos na estimativa das culturas de trigo e soja, considerando os algoritmos SVR – Laplace (regressão clássica) e GAM – Gaussian (regressão espacial), assim como os modelos com incorporação da altitude do terreno.

Tabela 8 - Resumo dos melhores resultados obtidos para a estimativa de produtividade das culturas de trigo e soja.

Cultura	Algoritmo	RE (cm/pixel)	Com altitude		
			R	R ²	RMSE
Trigo	SVR – Laplace	10	0,73	0,54	0,22
	GAM-Gaussian	20	0,90	0,81	0,14
Soja	SVR – Laplace	26	0,83	0,69	0,21
	GAM-Gaussian	20	0,96	0,92	0,11

Os modelos espaciais SAR e CAR, bem como a regressão linear múltipla clássica (RLM) foram utilizados por Araújo; Uribe-Opazo e Johann (2014), na estimativa de produtividade da soja, por meio de variáveis meteorológicas, com os quais os autores obtiveram valores de coeficiente de determinação (R²) de 50,58%, 50,61% e 24,15%, respectivamente. Assim como neste trabalho, os resultados obtidos pelos autores com as técnicas de regressão espacial foram superiores aos obtidos pela técnica clássica de RLM.

No trabalho de Gerke (2017) o autor utilizou dados obtidos por meio de RPA para realização da estimativa de produtividade da cultura de trigo, obtendo valores de r de até 0,61

utilizando a SVR. Este trabalho apresentou resultados de estimativa superiores, com valores de r até 0,73 na estimativa de produtividade de trigo usando SVR. Isso pode ter ocorrido pelo fato deste trabalho ter sido desenvolvido em uma área comercial, enquanto no trabalho de Gerke (2017) o objeto de estudo eram parcelas experimentais com diferentes tratamentos, o que acarreta em diferentes metodologias de extração e análise dos dados espectrais, diferentes formas de colheita (colhedora de precisão e colheita manual), diferentes kernels para SVR e softwares testados.

Na estimativa da produtividade de soja, também cultivada em parcelas experimentais e com a utilização de dados obtidos por meio de RPA, Yu et al. (2016) alcançaram valores de r de até 0,74. Os autores calcularam a área de cobertura das plantas por meio das imagens de RPA e correlacionaram com os dados de produtividade das parcelas. No presente trabalho, utilizando os valores das médias das bandas espectrais (R, G e B - em áreas de aproximadamente 7 m²) e a altitude central desta área na estimativa, foi possível a obtenção de correlações superiores, tanto nas técnicas da MD clássica, quanto nas técnicas de MDE, com r de 0,83 e 0,96, para as técnicas SVR - Laplace e GAM - Gaussian, respectivamente.

Maresma et al. (2016) obtiveram altos valores de coeficiente de determinação (R^2) na estimativa de produtividade de milho, variando de 0,60 a 0,92, quando utilizando a regressão linear simples, tendo como variáveis preditoras a altura das plantas e diferentes índices espectrais. Isso indica que a associação de variáveis espectrais com outras informações obtidas das plantas pode fornecer modelos com capacidade superior de estimativa de produtividade.

Araújo, Vettorazzi e Molin (2005) realizam a estimativa de produtividade das culturas de trigo e soja, utilizando dados extraídos de imagens obtidas por meio de videografia aérea multiespectral, com resolução espacial de 1m/pixel. Os melhores resultados obtidos pelos autores foram com a estimativa realizada por meio da regressão linear simples entre o NDVI e a produtividade das culturas, com valores de $R^2 = 0,54$ para a soja e $R^2 = 0,82$ para o trigo. Diferente dos resultados obtidos neste trabalho, os autores obtiveram melhores resultados para a estimativa de trigo e atribuem esse comportamento ao fato dos mapas de produtividade do trigo apresentarem uma definição mais clara das classes de produtividade.

No trabalho de Xin et al. (2013), utilizando dados de satélite, os autores realizam a estimativa da produtividade de milho e soja irrigados de um conjunto de estados dos Estados Unidos. A estimativa foi realizada utilizando duas abordagens, a nível de município, alcançando valores de $R^2 = 0,77$ para milho e $R^2 = 0,66$ para soja, e a nível de estado, com valores de $R^2 = 0,51$ para milho e $R^2 = 0,58$ para soja. Se comparado com a estimativa da produtividade de milho, que também é uma gramínea, os resultados obtidos neste trabalho para estimativa da

produtividade de trigo com o método GAM foram superiores, com coeficientes $R^2 = 0,81$. Já para a cultura da soja, tanto os melhores resultados alcançados pelas técnicas clássicas de MD, quanto para MDE, foram superiores aos obtidos pelos.

As técnicas RNA e SVR foram utilizadas por Kuwata e Shibasaki (2016) para a estimativa de produtividade de milho nos EUA. Assim como neste trabalho, o melhor resultado foi obtido com a técnica SVR, com $R^2 = 0,79$. No entanto, com a RNA os autores também obtiveram bom desempenho, com $R^2 = 0,78$. Em comparação com a estimativa da produtividade de trigo, com $R^2 = 0,25$ e $R^2 = 0,54$, para a RNA (câmera RGB) e SVR (câmera NIR), respectivamente, os resultados apresentados pelos autores foram melhores. Essa diferença pode ser atribuída a metodologia utilizada, visto que Kuwata e Shibasaki (2016) realizam a estimativa em escala nacional, com dados do sensor MODIS e resolução espacial de 1 km/pixel. Já no presente trabalho, a estimativa da produtividade do trigo foi realizada em talhão, com resoluções espaciais maiores (10 e 20 cm/pixel). No entanto, a técnica de regressão espacial GAM apresentou resultados similares aos obtidos pelos autores.

No trabalho de Wei et al. (2015), os autores realizaram a estimativa da produtividade da soja utilizando dados de evapotranspiração da cultura, produtividade de água, dentre outras variáveis, alcançando valores de RMSE de no mínimo 381 kg/há. No presente estudo, a MD clássica e a MDE possibilitaram a obtenção de valores menores de RMSE, sendo 220 kg/ha e 110 kg/ha, respectivamente.

A altitude do terreno obtida por meio dos DSMs possibilitou aumento considerável nos coeficientes r e R^2 para a estimativa de trigo e soja, com consequente diminuição dos valores de RMSE. Resultados similares foram obtidos por Souza et al. (2010), onde a altitude do terreno revelou uma correlação positiva com a produtividade de cana-de-açúcar, indicando que, nas áreas menos elevadas do talhão, a produtividade é menor. Montezano; Corazza e Muraoka (2008) avaliaram a variabilidade de nutrientes em plantas de milho cultivado em talhão manejado homogeneamente e também verificaram que a altitude proporcionou correlação positiva com a produtividade de milho, demonstrando que essa variável influencia o fluxo da água no solo, diretamente associado à produtividade das culturas.

6 CONCLUSÃO

A MDE se mostrou mais eficiente que a MD e com níveis de correlação similares a krigagem, na geração de modelos de estimativa de produtividade de trigo e soja. Porém, a técnica geoestatística de krigagem, mesmo alcançando bons resultados, não permitiu avaliar o impacto dos valores de bandas espectrais e de altitude do terreno na estimativa de produtividade, visto que independente dos atributos utilizados na estimativa, os altos valores dos coeficientes de correlação e determinação se mantiveram os mesmos quando usando tal técnica.

O método GAM utilizando a função de ajuste gaussiana apresentou os resultados mais coerentes, quando comparados com as demais técnicas de regressão clássica e espacial utilizadas. Os modelos gerados pelo método GAM – Gaussian foram capazes de explicar até 81% da variação da produtividade de trigo e 92% da produtividade da soja, quando incorporando a altitude do terreno aos atributos de predição.

Dentre as câmeras RGB e NIR utilizadas no imageamento da cultura de trigo, a câmera RGB apresentou os melhores resultados para os modelos sem a incorporação da altitude. No entanto, quando os valores de altitude foram adicionados, para a maioria das técnicas, os dados obtidos pela câmera NIR apresentaram capacidade superior de estimativa de produtividade. Porém, se analisando os melhores resultados, os coeficientes de correlação são similares, sendo 0,73 para a NIR e 0,72 para a RGB. Isso indica que mais estudos devem ser conduzidos, visando analisar o comportamento das diferentes câmeras em outras situações.

Na estimativa da produtividade do trigo, as resoluções espaciais de 10 e 20 cm/pixel não apresentaram diferenças significativas nos resultados, tanto para a estimativa com os dados da câmera RGB, quanto para NIR. Já para a cultura da soja, considerando as técnicas da MD clássica, as quais foram capazes de distinguir melhor os resultados para as diferentes resoluções, a menor resolução espacial utilizada (26 cm/pixel) foi a que possibilitou a obtenção dos melhores resultados. Isso indica que, sobrevoar com resoluções mais altas, além de aumentar o tempo de execução dos voos e o custo computacional para o processamento das imagens, não possibilitou a geração de melhores modelos de estimativa.

Com isso, pode-se concluir que a mineração de dados espaciais é promissora na geração de modelos de estimativa de produtividade de grãos a partir de imagens obtidas por RPA. Assim, os dados obtidos por meio do sensoriamento remoto com RPA podem ser utilizados como alternativa na geração de mapas de produtividade, visto que o custo para a aquisição de colhedoras de precisão é muito alto.

Como trabalhos futuros, a análise de novos modelos de regressão espacial, assim como a aplicação dos modelos testados neste trabalho em diferentes áreas de cultivo que apresentem maior variação de produtividade é importante, para nova comparação dos resultados obtidos com MD, MDE e krigagem. A análise da resolução espacial em novos trabalhos também é importante, para consolidar o conhecimento aqui adquirido de que resoluções mais altas não garantem melhores níveis de correlação na estimativa de produtividade de soja e trigo, mostrando ser vantajoso voar com as menores resoluções, pelo fato de acarretar em menores custos de tempo de voo e processamento.

REFERÊNCIAS

- ALGANCI, U. et al. Estimating maize and cotton yield in southeastern Turkey with integrated use of satellite images, meteorological data and digital photographs. **Field Crops Research**, v. 157, p. 8–19, 2014.
- ALKATHIRI, M.; ABDUL, J.; POTDAR, M. B. Geo-spatial Big Data Mining Techniques. **International Journal of Computer Applications**, v. 135, n. 11, p. 975–8887, 2016.
- ANAC. **Rpas - Sistemas De Aeronaves Remotamente Pilotadas**. Disponível em: <<http://www2.anac.gov.br/rpas/>>. Acesso em: 11 jan. 2018.
- ANSELIN, L.; FLORAX, R. J. G. M.; REY, S. J. **Advances in Spatial Econometrics**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.
- ANSELIN, L.; SYABRI, I.; KHO, Y. **GeoDa: An introduction to spatial data analysis** *Geographical Analysis*, 2006.
- APPICE, A.; LANZA, A.; MALERBA, D. **An Integrated Platform for Spatial Data Mining within a GIS Environment**. 2007 IEEE 23rd International Conference on Data Engineering Workshop. **Anais...IEEE**, abr. 2007 Disponível em: <http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4401036>
- ARAUJO, M. A.; SOUZA, J. L. M. DE; TSUKAHARA, R. Y. Modelos agro-meteorológicos na estimativa da produtividade da cultura da soja na região de Ponta Grossa, Estado do Paraná. **Acta Scientiarum. Agronomy**, v. 33, n. 1, p. 23–31, 2 mar. 2011.
- ARAÚJO, E. C. DE; URIBE-OPAZO, M. A.; JOHANN, J. A. Modelo de regressão espacial para estimativa da produtividade da soja associada a variáveis agrometeorológicas na região oeste do estado do Paraná. **Engenharia Agrícola**, v. 34, n. 2, p. 286–299, abr. 2014.
- ARAÚJO, J. C. DE; VETTORAZZI, C. A.; MOLIN, J. P. Estimativa da produtividade e determinação de zonas de manejo, em culturas de grãos, por meio de videografia aérea multiespectral. **Acta Scientiarum. Biological Sciences Acta Sci. Agron. Maringá**, v. 27, n. 3, p. 437–447, 2005.
- ASSAD, E. D. et al. Sistema de previsão da safra de soja para o Brasil. v. 42, n. 5, p. 615–625, 2007.
- BAE, D.-H. et al. **SD-Miner: A spatial data mining system**. 2009 IEEE International Conference on Network Infrastructure and Digital Content. **Anais...IEEE**, nov. 2009 Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5360924>>
- BAILEY, T. C; GATRELL, A. C. **Interactive Spatial Data Analysis** *Interactive Spatial Data Analysis*, 1995.
- BASAK, D.; PAL, S.; PATRANABIS, D. C. Support Vector Regression. **Neural Information Processing – Letters and Reviews**, v. 11, n. 10, 2007.
- BELAYNEH, A. et al. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. **Journal of Hydrology**, v. 508, p. 418–429, jan. 2014.

BOGORNY, V. et al. **Weka-GDPM – Integrating Classical Data Mining Toolkit to Geographic Information Systems**. Workshop on Data Mining Algorithms and Applications, 2006. Disponível em: <<http://www.inf.ufsc.br/~vania.bogorny/artigos/WekaGDPM2006.pdf>>

BURROUGH, P. A. **Principles of geographical information systems for land resources assessment**. [s.l: s.n.].

CAMPBELL, J. B.; WYNNE, R. H. **Introduction to Remote Sensing**. 5^a ed. [s.l.] Guilford Press, 2011.

CANDIAGO, S. et al. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. **Remote Sensing**, v. 7, n. 4, p. 4026–4047, 2015.

CHAPMAN, S. et al. Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. **Agronomy**, v. 4, n. 2, p. 279–301, 2014.

CHAVES, A. A. et al. Using UAVs and digital image processing to quantify areas of soil and vegetation. **Journal of Physics: Conference Series**, v. 633, p. 12112, 21 set. 2015.

COSTA, E. M. et al. Processamento de imagens RapidEye no mapeamento de uso do solo em ambiente de Mar de Morros. **Pesquisa Agropecuária Brasileira**, v. 51, n. 9, p. 1417–1427, set. 2016.

CRESSIE, N. A. C. **Statistics for Spatial Data**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1993.

DI GENNARO, S. F. et al. Unmanned Aerial Vehicle (UAV)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex. **Phytopathologia Mediterranea**, v. 55, n. 2, p. 262–275, 2016.

EHSANI, R. et al. Affordable multi-rotor remote sensing platform for applications in precision horticulture. **11th International Conference on Precision Agriculture**, 2012.

FAYYAD, U.; UTHURUSAMY, R. Data mining and knowledge discovery in databases. **Communications of the ACM**, v. 39, n. 11, p. 24–26, 1 nov. 1996.

GERKE, T. **MINERAÇÃO DE DADOS DE IMAGENS OBTIDAS COM AERONAVE REMOTAMENTE PILOTADA PARA ESTIMATIVA DE PRODUTIVIDADE DO TRIGO**. [s.l.] Universidade Estadual de Ponta Grossa, 2017.

GRZEGOZEWSKI, D. M. et al. SPATIAL CORRELATION OF SOYBEAN PRODUCTIVITY, ENHANCED VEGETATION INDEX (EVI) AND AGROMETEOROLOGICAL VARIABLES. **Journal of the Brazilian Association of Agricultural Engineering**, v. 37, n. 3, 2017.

HAMDAD, L. et al. **EasySDM: An integrated and easy to use spatial data mining platform**. IC3K 2015 - Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. **Anais...2015**

HAN, J.; KOPERSKI, K.; STEFANOVIC, N. GeoMiner: a system prototype for spatial data mining. **Proceedings ACM SIGMOD International Conference on Management of Data**,

SIDMOD '97, p. 553–556, 1997.

HOLZMAN, M. E.; RIVAS, R.; PICCOLO, M. C. Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. **International Journal of Applied Earth Observation and Geoinformation**, v. 28, p. 181–192, 2014.

HUNT, JR., E. R. et al. Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring. **Remote Sensing**, v. 2, n. 1, p. 290–305, 11 jan. 2010.

JANSEN, M. et al. Non-Invasive Spectral Phenotyping Methods can Improve and Accelerate Cercospora Disease Scoring in Sugar Beet Breeding. **Agriculture**, v. 4, n. 2, p. 147–158, 2014.

JORGE, L. A. DE C.; INAMASU, R. Y. Uso de veículos aéreos não tripulados (VANT) em agricultura de precisão. **Ferramentas para agricultura de precisão**, p. 109–134, 2013.

KARATZOGLOU, A. et al. kernlab - An S4 Package for Kernel Methods in R. **Journal of Statistical Software**, v. 11, n. 9, 2004.

KÖRTING, T. S.; GARCIA FONSECA, L. M.; CÂMARA, G. GeoDMA-Geographic Data Mining Analyst. **Computers and Geosciences**, v. 57, p. 133–145, 2013.

KUWATA, K.; SHIBASAKI, R. ESTIMATING CORN YIELD IN THE UNITED STATES WITH MODIS EVI AND MACHINE LEARNING METHODS. **ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. III-8, p. 131–136, 2016.

LACERDA, V. S. **ESTIMATIVA DO ÍNDICE DE SEVERIDADE DE FERRUGEM ASIÁTICA NA CULTURA DA SOJA POR MEIO DE IMAGENS OBTIDAS COM AERONAVE REMOTAMENTE PILOTADA**. [s.l.] Universidade Estadual de Ponta Grossa, 2017.

LEBOURGEOIS, V. et al. A light-weight multi-spectral aerial imaging system for nitrogen crop monitoring. **Precision Agriculture**, v. 13, n. 5, p. 525–541, 13 out. 2012.

LI, D. et al. Software and applications of spatial data mining. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 6, n. 3, p. 84–114, maio 2016.

MANCINI, F. et al. Using Unmanned Aerial Vehicles (UAV) for High-Resolution Reconstruction of Topography: The Structure from Motion Approach on Coastal Environments. **Remote Sensing**, v. 5, n. 12, p. 6880–6898, 9 dez. 2013.

MARESMA, Á. et al. Analysis of vegetation indices to determine nitrogen application and yield prediction in maize (zea mays l.) from a standard uav service. **Remote Sensing**, v. 8, n. 12, 2016.

MATESE, A. et al. Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. **Remote Sensing**, v. 7, n. 3, p. 2971–2990, 13 mar. 2015.

MENNIS, J.; GUO, D. Spatial data mining and geographic knowledge discovery—An introduction. **Computers, Environment and Urban Systems**, v. 33, n. 6, p. 403–408, nov. 2009.

MERZ, T.; CHAPMAN, S. AUTONOMOUS UNMANNED HELICOPTER SYSTEM FOR REMOTE SENSING MISSIONS IN UNKNOWN ENVIRONMENTS. **ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XXXVIII-1/, p. 143–148, 6 set. 2012.

MONTEZANO, Z. F.; CORAZZA, E. J.; MURAOKA, T. Variabilidade de nutrientes em plantas de milho cultivado em talhão manejado homogeneamente. **Bragantia**, v. 67, n. 4, p. 969–976, dez. 2008.

MORAN, M. S.; INOUE, Y.; BARNES, E. M. **Opportunities and limitations for image-based remote sensing in precision crop management** *Remote Sensing of Environment*, 1997.

MUCHERINO, A.; PAPAJOJGI, P.; PARDALOS, P. M. A survey of data mining techniques applied to agriculture. **Operational Research**, v. 9, n. 2, p. 121–140, 6 ago. 2009.

NONATO, R. T.; DE, S. R.; OLIVEIRA, M. TÉCNICAS DE MINERAÇÃO DE DADOS PARA IDENTIFICAÇÃO DE ÁREAS COM CANA-DE-AÇÚCAR EM IMAGENS LANDSAT 5. v. 33, n. 6, p. 1268–1280, 2013.

PERIN, G. et al. Análise de Acurácia de Georrefereciamento de Mosaicos de Imagens Obtidas por RPA. **Anais do EATI - Encontro Anual de Tecnologia da Informação e STIN – Simpósio de Tecnologia da Informação da Região Noroeste do RS 193**, p. 193–199, 2016.

PICOLI, M. C. A. et al. Índice de vegetação do sensor MODIS na estimativa da produtividade agrícola da cana-de-açúcar. **Bragantia**, v. 68, n. 3, p. 789–795, 2009.

RICHARDS, J. A. **Remote Sensing Digital Image Analysis**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

RODDICK, J.; HORNSBY, K.; SPILIOPOULOU, M. **Temporal, Spatial, and Spatio-Temporal Data Mining**. [s.l.: s.n.]. v. 2007

SALAMÍ, E.; BARRADO, C.; PASTOR, E. UAV Flight Experiments Applied to the Remote Sensing of Vegetated Areas. **Remote Sensing**, v. 6, n. 11, p. 11051–11081, 11 nov. 2014.

SANKARAN, S. et al. **Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review** *European Journal of Agronomy*, 2015.

SENSEFLY. **The Professional's Mapping - Drone of Choice**. Disponível em: <<https://www.sensefly.com/home.html>>. Acesso em: 11 jan. 2018.

SHIMABUKURO, Y. E.; MAEDA, E. E.; FORMAGGIO, A. R. Sensoriamento Remoto e Sistemas de Informações Geográficas aplicados ao estudo dos recursos agronômicos e florestais. **Ceres**, v. 56, n. 4, p. 399–409, 2015.

SILVESTRO, P. et al. Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and Simple Algorithm for Yield (SAFY) Models. **Remote Sensing**, v. 9, n. 6, p. 509, 22 maio 2017.

SONG, C. et al. Classification and Change Detection Using Landsat TM Data. **Remote Sensing of Environment**, v. 75, n. 2, p. 230–244, fev. 2001.

SOUZA, Z. M. DE et al. Análise dos atributos do solo e da produtividade da cultura de cana-de-açúcar com o uso da geoestatística e árvore de decisão. **Ciência Rural**, v. 40, n. 4, p. 840–847, abr. 2010.

TASSINARI, W. S.; LORENZON, M. C.; PEIXOTO, E. L. Spatial regression methods to evaluate beekeeping production in the state of Rio de Janeiro. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 65, n. 2, p. 553–558, 2013.

VAPNIK, V.; GOLOWICH, S. E.; SMOLA, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. **Advances in Neural Information Processing Systems**, v. 9, p. 281–287, 1997.

VATSAVAI, R. R. et al. **Spatiotemporal data mining in the era of big spatial data**. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data - BigSpatial '12. **Anais...**New York, New York, USA: ACM Press, 2012

WEI, Z. et al. Modelling transpiration, soil evaporation and yield prediction of soybean in North China Plain. **Agricultural Water Management**, v. 147, p. 43–53, jan. 2015.

WERE, K. et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. **Ecological Indicators**, v. 52, p. 394–403, maio 2015.

WÓJTOWICZ, M.; WÓJTOWICZ, A.; PIEKARCZYK, J. Application of remote sensing methods in agriculture. **Communications in Biometry and Crop Science**, v. 11, p. 31–50, 2016.

WOOD, S. N. Generalized additive models: an introduction with R. **Wiley Interdisciplinary Reviews: Computational ...**, v. 16, n. July, p. xvii, 392 , 2006.

XIN, Q. et al. A production efficiency model-based method for satellite estimates of corn and soybean yields in the midwestern US. **Remote Sensing**, v. 5, n. 11, p. 5926–5943, 2013.

YI, J. L. R.; SHIMABUKURO, Y. E.; QUINTANILHA, J. A. Identificação e mapeamento de áreas de milho na região sul do Brasil utilizando imagens MODIS. **Engenharia Agrícola**, v. 27, n. 3, p. 753–763, dez. 2007.

YU, N. et al. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. 2016.

ZARCO-TEJADA, P. J.; GONZÁLEZ-DUGO, V.; BERNI, J. A. J. Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. **Remote Sensing of Environment**, v. 117, p. 322–337, fev. 2012.

ZARCO-TEJADA, P. J.; USTIN, S. L.; WHITING, M. L. Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. **Agronomy Journal**, v. 97, n. 3, p. 641–653, 2005.