

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

CHRISTOPHER DJONNY PEREIRA PRESTES

**PREDIÇÃO DE PRODUTIVIDADE DE TRIGO POR MEIO DE DADOS
ESPECTRAIS E ALTURA ESTIMADA DA PLANTA OBTIDOS POR MEIO DE
AERONAVE REMOTAMENTE PILOTADA**

PONTA GROSSA

2020

CHRISTOPHER DJONNY PEREIRA PRESTES

**PREDIÇÃO DE PRODUTIVIDADE DE TRIGO POR MEIO DE DADOS
ESPECTRAIS E ALTURA ESTIMADA DA PLANTA OBTIDOS 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. Dr. Eduardo Fávero Caires

Coorientação: Prof^ª Dr^ª Alaine Margarete Guimarães

PONTA GROSSA

2020

P936 Prestes, Christopher Djonny Pereira
Predição de produtividade de trigo por meio de dados espectrais e altura
estimada da planta obtidos por meio de aeronave remotamente pilotada /
Christopher Djonny Pereira Prestes. Ponta Grossa, 2020.
61 f.

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

Orientador: Prof. Dr. Eduardo Fávero Caires.
Coorientadora: Profa. Dra. Alaine Margarete Guimarães.

1. Triticum aestivum L.. 2. Predição de rendimento de grãos. 3.
Sensoriamento remoto. 4. Bandas espectrais. 5. Mineração de dados. I. Caires,
Eduardo Fávero. II. Guimarães, Alaine Margarete. III. Universidade Estadual de
Ponta Grossa. Computação para Tecnologias em Agricultura. IV.T.

CDD: 004



UNIVERSIDADE ESTADUAL DE PONTA GROSSA
Av. General Carlos Cavalcanti, 4748 - Bairro Uvaranas - CEP 84030-900 - Ponta Grossa - PR - <https://uepg.br>

TERMO

TERMO DE APROVAÇÃO

Christopher Djonny Pereira Prestes

PREDIÇÃO DE PRODUTIVIDADE DE TRIGO POR MEIO DE DADOS ESPECTRAIS E DE ALTURA ESTIMADA DA PLANTA OBTIDOS 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. Eduardo Fávero Caires - UEPG

Prof. Dr. Arion de Campos Junior - UEPG

Dr. Adriano Haliski - Fundação ABC

Ponta Grossa, 14 de agosto de 2020.



Documento assinado eletronicamente por **Arion de Campos Junior, Professor(a)**, em 18/08/2020, às 11:03, conforme art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Adriano Haliski, Usuário Externo**, em 18/08/2020, às 17:21, conforme art. 1º, III, "b", da Lei 11.419/2006.

Documento assinado eletronicamente por **Eduardo Favero Caires, Professor(a)**, em 18/08/2020, às 22:43, conforme art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida no site <https://sei.uepg.br/autenticidade> informando o código verificador **0255419** e o código CRC **386B3D42**.

20.000022143-3

0255419v6

AGRADECIMENTOS

Aos meus pais, que me apoiaram constantemente durante mais essa etapa da minha vida.

Ao meu orientador Eduardo Fávero Caires e a minha coorientadora Alaine Margarete Guimarães, pela ajuda e pelos ensinamentos durante toda a realização desse trabalho.

A todos os colegas dos laboratórios InfoAgro, Fertilidade do Solo e Geoprocessamento, da Universidade Estadual de Ponta Grossa, pela companhia e auxílio em muitas tarefas em campo.

Aos meus orientadores de Iniciação Científica, Tiago Gerke e Valter Luís Estevam Junior, pelo incentivo e recomendações que me ajudaram a entrar neste programa.

A Capes e a Fundação Araucária, pelo auxílio financeiro.

“O que sabemos é uma gota, o que não sabemos é um oceano.”

Isaac Newton

RESUMO

O acompanhamento da cultura de trigo no campo, desde sua semeadura até a colheita, é indispensável para o controle de eventuais problemas que podem ocorrer durante seu desenvolvimento. A predição de produtividade de trigo é uma forma de fornecer ao produtor informações sobre o quanto será o rendimento da sua lavoura, ajudando a identificar possíveis problemas antecipadamente. A altura da planta também pode ser um importante indicador de rendimento, uma vez que pode estar relacionada ao desempenho da cultura. A extração dessas informações em campo pode ser inviável quando realizada manualmente e o uso de tecnologias que implementam os conceitos de Sensoriamento Remoto (SR), como as aeronaves remotamente pilotadas (RPA), pode ser uma alternativa para suprir essa necessidade. Considerando o recente surgimento de RPA em aplicações agrícolas e o limitado número de trabalhos envolvendo a estimativa de altura das plantas por meio de modelos digitais, essa pesquisa teve o objetivo de obter um modelo numérico para a predição de produtividade de trigo baseado em valores digitais de bandas espectrais e valores estimados de alturas de plantas extraídos de imagens de RPA. As áreas experimentais desse estudo estão localizadas no município de Ponta Grossa – PR, e foram semeadas com trigo nos anos de 2018 e 2019. Os voos sobre as áreas em 2018 foram realizados por uma RPA modelo eBee, utilizando as câmeras RGB e NIR em diferentes resoluções. O voo sobre a área em 2019 foi realizado por uma RPA modelo Phantom utilizando uma câmera RGB com resolução de 5cm/px. As imagens obtidas foram processadas no software Pix4D, que faz uso de algoritmos baseados em *Structure from Motion* (SfM) e permite a criação dos modelos digitais de superfície (DSM) e elevação (DEM). A partir do cálculo da diferença entre esses modelos foi possível estimar a altura dos objetos acima do solo. Também foram realizadas medidas manuais de altura das plantas e de produtividade de grãos no campo a fim de validar os modelos. Para a criação dos modelos de predição foram selecionados três algoritmos de Mineração de Dados (MD) a serem utilizados em testes preliminares. O SMOReg, um algoritmo baseado em Máquina de vetores de suporte (SVM) para aplicações que envolvem regressão, apresentou os melhores resultados e foi selecionado para dar prosseguimento aos testes. Os resultados obtidos demonstraram que a altura da planta pode contribuir consideravelmente para o desempenho dos modelos de predição de produtividade de trigo. Os resultados envolvendo a estimativa de altura das plantas indicaram que a câmera NIR se apresentou como a opção mais precisa para esse objetivo, obtendo um erro médio de estimativa consideravelmente menor do que o da câmera RGB. Os testes utilizando a combinação dos dados digitais da câmera RGB e a altura estimada das plantas pelos modelos alcançaram índices de correlação (r) de até 0,97. Portanto, foi possível obter um modelo de predição que apresentou alta correlação com a produtividade de grãos de trigo, utilizando apenas dados de SR.

Palavras-chave: *Triticum aestivum* L., predição de rendimento de grãos, sensoriamento remoto, bandas espectrais, modelos digitais de superfície e elevação, mineração de dados.

ABSTRACT

The field monitoring of wheat crop, from sowing to harvest, is essential to control any problems that may occur during crop development. The wheat yield prediction is a way to provide the farmers information about how the crop will perform, helping to identify possible problems in advance. The plant height can also be an important indicator of yield, since it can be related to the crop performance. The extraction of this information in the field can be impracticable when performed manually and the use of technologies that implement the Remote Sensing (SR) concepts, such as remotely piloted aircraft (RPA), can be an alternative to supply this need. Considering the recent appearance of RPA in agricultural applications and the limited number of studies involving plant height estimation using digital models, this study aimed to obtain a numerical model for the wheat yield prediction based on digital values of spectral bands and estimated plant height values extracted from RPA images. The experimental areas of this study were located in Ponta Grossa, Parana State, Brazil, and were sown with wheat in 2018 and 2019. The flights over the areas in 2018 were carried out by an RPA model eBee, using RGB and NIR cameras in different resolutions. The flight over the area in 2019 was carried out by an RPA model Phantom using an RGB camera with 5 cm/px resolution. The images obtained were processed in the Pix4D software, which uses algorithms based on Structure from Motion (SfM) and allows the creation of digital surface (DSM) and elevation (DEM) models. From the calculation of the difference between these models it was possible to estimate the height of the objects above the ground. Manual measurements of plant height and grain yield in the field were also performed in order to validate the models. To create the prediction models, three Data Mining (MD) algorithms were selected for use in preliminary tests. SMOReg, an algorithm based on Support vector machine (SVM) for applications involving regression, presented the best results and was selected to continue the tests. The results based on the plant height estimation indicated that the NIR camera proved to be more accurate for this purpose, obtaining a considerably lower average error of estimate compared to that of the RGB camera. Tests performed with the combination of digital RGB data and plant height estimated by the models reached a correlation index (r) of up to 0.97. Therefore, a prediction model that showed a high correlation with wheat grain yield was obtained in our study, using only SR data.

Keywords: *Triticum aestivum* L., grain yield prediction, remote sensing, spectral bands, digital surface and elevation models, data mining.

LISTA DE FIGURAS

Figura 1 – Escala Feekes-Large de desenvolvimento de trigo	17
Figura 2 – Refletância de plantas saudáveis, plantas estressadas e do solo em diferentes comprimentos de onda de um sensor.	19
Figura 3 – Imagem aérea de um carro e os modelos DSM e DEM gerados para uma mesma área.....	20
Figura 4 – Altura estimada do objeto obtida por meio da diferença entre os modelos digitais DSM e DEM.....	21
Figura 5 – Diagrama de dispersão de dados de uma relação linear entre duas variáveis	23
Figura 6 – Fluxograma apresentando as principais etapas do projeto	25
Figura 7 – Área experimental 1 (Fazenda Santa Cruz).....	26
Figura 8 – Área experimental 2 (Fazenda Escola da UEPG).....	27
Figura 9 – Área experimental 3 (Fazenda Escola da UEPG).....	27
Figura 10 – RPA eBee, câmera RGB e câmera NIR	28
Figura 11 – RPA Phantom 4.....	29
Figura 12 – Receptor GPS Trimble ProXT e coletor de dados Juno SB e marcadores CGP... 30	
Figura 13 – Distância entre o centro de um GCP e o ponto coletado pelo GPS em solo nas imagens em que ele é visualizado na Área 3.....	32
Figura 14 – Ortomosaico e mapa de refletância gerados para a Área 1.....	33
Figura 15 – DSM e DEM gerados para a Área 1.....	33
Figura 16 – Reconstrução 3D da Área 1 utilizando imagens da câmera RGB.....	34
Figura 17 – Divisões e polígonos desenhados para cada parcela da Área 2	35
Figura 18 – Pontos de coleta indicando os locais de colheita em cada parcela na Área 3.....	36
Figura 19 – Exemplo de um polígono com diferentes quantidades de pontos por pixel	36
Figura 20 – Mapa de altura das plantas de trigo obtido pelos dados da câmera RGB	37
Figura 21 – Exemplo de dois <i>outliers</i> identificados em uma das bases de dados	40
Figura 22 – Mapas de Refletância, altura das plantas e produtividade de grãos da Área 1	44
Figura 23 – Mapas de Refletância, altura das plantas e produtividade de grãos da Área 2	45
Figura 24 – Dispersão dos dados entre altura das plantas e produtividade de grãos de trigo da Área 1	46
Figura 25 – Dispersão dos dados entre altura das plantas e produtividade de grãos de trigo da Área 2	46

Figura 26 – Mapa de erro de estimativa de altura das plantas de trigo da Área 1 para cada parcela de acordo com as câmeras RGB e NIR utilizando resolução de 10 cm/pixel..... 52

LISTA DE QUADROS

Quadro 1 – Estudos de predição de produtividade de grãos em cultura anuais	18
Quadro 2 – Descrição dos atributos das bases de dados.....	39
Quadro 3 – Comando utilizado para a execução do pacote wekaServer... ..	41
Quadro 4 – Algoritmos utilizados para os testes iniciais	42

LISTA DE TABELAS

Tabela 1 – Descrição dos dados de produtividade das áreas de estudo	28
Tabela 2 – Datas, resoluções e dias após a semeadura dos voos realizados.	31
Tabela 3 – Tamanho em instâncias de cada uma das bases criadas.	38
Tabela 4 – Exemplo mostrando a composição de uma base de dados do estudo.....	40
Tabela 5 – Desempenho dos algoritmos de acordo com a proporção dos dados da base A1_RGB_10 (RGB) correlacionados com a produtividade de trigo.....	47
Tabela 6 – Desempenho dos algoritmos de acordo a proporção dos dados da base A1_RGB_10 (RGB+Altura real) correlacionados com a produtividade de trigo.	48
Tabela 7 – Comparação entre as resoluções para as estimavas de altura das plantas de trigo na Área 1 utilizando dados da câmera RGB.....	49
Tabela 8 – Comparação entre as resoluções para as estimavas de altura das plantas de trigo na Área 1 utilizando dados da câmera NIR.....	50
Tabela 9 – Comparação entre as resoluções para as estimavas de altura das plantas de trigo na Área 2 utilizando dados da câmera RGB.....	51
Tabela 10 – Resultados dos testes de predição de produtividade de grãos de trigo de acordo com as resoluções das imagens da Área 1.....	53
Tabela 11 – Resultados dos testes de predição de produtividade de grãos de trigo de acordo com as resoluções das imagens das Áreas 2 e 3	54

LISTA DE SIGLAS

AP	Agricultura de precisão
CONAB	Companhia Nacional de Abastecimento
DAS	Dias após a semeadura
DEM	Modelo digital de elevação (do inglês <i>Digital Elevation Model</i>)
DP	Desvio padrão
DSM	Modelo digital de superfície (do inglês <i>Digital Surface Model</i>)
GCP	Pontos de controle terrestre (do inglês <i>Ground Control Point</i>)
GPS	Sistema de Posicionamento Global (do inglês <i>Global Positioning System</i>)
GNSS	Sistema Global de Navegação por Satélite (do inglês <i>Global Navigation Satellite System</i>)
IBGE	Instituto Brasileiro de Geografia e Estatística
MD	Mineração de dados
NIR	Infravermelho próximo (do inglês <i>Near Infrared</i>)
PPP	Posicionamento por ponto preciso (do inglês <i>Precise Point Positioning</i>)
RGB	Vermelho, verde e azul (do inglês <i>Red, Green e Blue</i>)
RPA	Aeronave remotamente pilotada (do inglês <i>Remotely Piloted Aircraft</i>)
SfM	Estrutura do movimento (do inglês <i>Structure from Motion</i>)
SIRGAS	Sistema de Referência Geocêntrico para as Américas.
SR	Sensoriamento remoto
SVLT	Sistemas de Varredura Laser Terrestre
SVM	Máquina de Vetores de Suporte (do inglês <i>Support Vector Machine</i>)
VANT	Veículo aéreo não tripulado
UAV	Veículo aéreo não tripulado (do inglês <i>Unmanned aerial vehicle</i>).
UEPG	Universidade Estadual de Ponta Grossa

SUMÁRIO

1	INTRODUÇÃO	14
2	OBJETIVOS.....	16
2.1	OBJETIVO PRINCIPAL	16
2.2	OBJETIVOS ESPECÍFICOS	16
3	REVISÃO DE LITERATURA	17
3.1	CULTIVO DO TRIGO E PREDIÇÃO DE PRODUTIVIDADE.....	17
3.2	SENSORIAMENTO REMOTO E AERONAVES REMOTAMENTE PILOTADAS	19
3.3	PROCESSAMENTO DIGITAL DE IMAGENS E MODELOS 3D	20
3.4	MINERAÇÃO DE DADOS (MD)	22
4	MATERIAL E MÉTODOS	25
4.1	ÁREAS DE ESTUDO E EQUIPAMENTOS UTILIZADOS	25
4.2	COLETA DOS DADOS.....	30
4.3	PROCESSAMENTO DAS IMAGENS	31
4.4	EXTRAÇÃO DOS DADOS	34
4.4.1	Extração dos valores digitais	34
4.4.2	Cálculo e obtenção das alturas das plantas	37
4.5	MINERAÇÃO DOS DADOS.....	38
4.5.1	Criação e pré-processamento das bases de dados	38
4.5.2	Análise dos dados e técnicas de aprendizado de máquina.....	41
5	RESULTADOS E DISCUSSÃO	44
5.1	AVALIAÇÃO DOS DADOS	44
5.2	TESTES INICIAIS DE REGRESSÃO	47
5.3	ESTIMATIVAS DE ALTURA	49
5.4	PREDIÇÃO DE PRODUTIVIDADE	53
6	CONCLUSÕES	56
	REFERÊNCIAS	58

1 INTRODUÇÃO

O acompanhamento da cultura de trigo no campo, desde sua semeadura até a colheita, é importante para o controle de eventuais problemas que podem ocorrer durante seu desenvolvimento. Dentro dos conceitos da Agricultura de Precisão (AP), a predição de produtividade de grãos é uma forma de permitir ao produtor identificar esses problemas e agir a tempo de corrigir eventuais falhas de desempenho ainda durante o estágio de desenvolvimento vegetativo da cultura. A altura da planta também pode ser um importante indicador de rendimento de grãos, uma vez que ela pode estar relacionada ao desempenho da cultura. Essa informação também pode servir como parâmetro para práticas de manejo específicas em um local identificado como de menor crescimento (LATI et al., 2013).

Considerando que o monitoramento manual de um campo de cultivo pode ser muitas vezes inviável dependendo de sua extensão, o uso de tecnologias que implementam conceitos de Sensoriamento Remoto (SR) pode ser uma alternativa ao agricultor que necessite de informações antecipadas de sua cultura para proporcionar a tomada de decisões visando a obtenção de maior rendimento de grãos. Essa metodologia já foi utilizada em diversos trabalhos nos últimos anos, tanto em trigo (RODRIGES et al., 2015; SILVESTRO et al. 2017) quanto em outras culturas, como soja (YU et al., 2016; RICHETTI et al., 2018) e milho (PANDA et al., 2010; BATTUDE et al., 2016).

As tecnologias baseadas em SR possuem a vantagem de monitorar grandes áreas em um menor período de tempo, o que pode otimizar e diminuir o custo desses processos. Xue et al. (2017) destacam que, embora os satélites tenham desempenhado um importante papel no SR durante muitos anos, recentemente houve o crescimento da utilização de Aeronaves Remotamente Pilotadas (RPAs – do inglês *Remotely Piloted Aircraft*), também conhecidas como *drone* (do inglês *zangão*), VANT (Veículo Aéreo Não Tripulado), ou UAV (do inglês *Unmanned Aerial Vehicle*).

A utilização de RPA apresenta vantagem em relação aos satélites pelo fato de, geralmente, não terem suas imagens comprometidas por condições atmosféricas em dias nublados e não dependerem do período orbital, que é o tempo em que um satélite volta a visitar a mesma área. Essas interferências podem dificultar o sensoriamento em aplicações agrícolas, especificamente aquelas relacionadas ao manejo de água e nutrientes. Outra vantagem é que essas plataformas podem voar em baixas altitudes, possibilitando maior resolução espectral e, com isso, fornecendo mais detalhes sobre a área imageada, como demonstrado nos trabalhos de Homan et al. (2016) e Makanza et al. (2018), os quais utilizaram resoluções de até 1 cm/pixel.

Após o sensoriamento de uma determinada área, as imagens geradas podem ser submetidas a técnicas de Processamento Digital de Imagens (PDI) e, posteriormente, podem ser extraídos dados sobre a cultura, como por exemplo valores digitais de refletância espectral. O PDI também permite a geração dos modelos digitais de superfície (DSM – do inglês *Digital Surface Model*) e modelos digitais do terreno (DTM – do inglês *Digital Terrain Model*). A partir desses modelos é possível obter informações como a altitude da área de cultivo e a altura das plantas.

Tendo em vista a grande quantidade de dados que podem ser obtidos e que eles podem se correlacionar com a produtividade de grãos, as técnicas de Mineração de Dados (MD) podem ser utilizadas para o processamento e a extração de informações que sejam úteis ao produtor. Tais técnicas permitem a combinação dos dados obtidos pelo SR e a elaboração de um modelo numérico que possa ter forte correlação com o rendimento de grãos de trigo.

Considerando que ainda é recente a utilização de RPA em aplicações agrícolas e o limitado número de trabalhos envolvendo a estimativa de altura das plantas por meio de modelos digitais, é importante também determinar quais os melhores métodos e parâmetros a serem utilizados como objetivo de prever a produtividade de grãos de trigo utilizando essa metodologia.

2 OBJETIVOS

2.1 OBJETIVO PRINCIPAL

O objetivo principal desse estudo foi obter um modelo numérico para a predição de produtividade de trigo baseado em valores digitais de bandas espectrais e valores estimados de altura de plantas extraídos de imagens de RPA.

2.2 OBJETIVOS ESPECÍFICOS

- Avaliar em que magnitude os valores de altura das plantas contribuem para a predição de produtividade de grãos de trigo em relação ao uso exclusivo dos valores digitais das bandas.
- Avaliar qual câmera (RGB ou NIR) e resolução espectral seriam melhores para a estimativa de altura das plantas de trigo.
- Aplicar técnicas de MD na determinação do modelo numérico de predição de produtividade de grãos de trigo.

3 REVISÃO DE LITERATURA

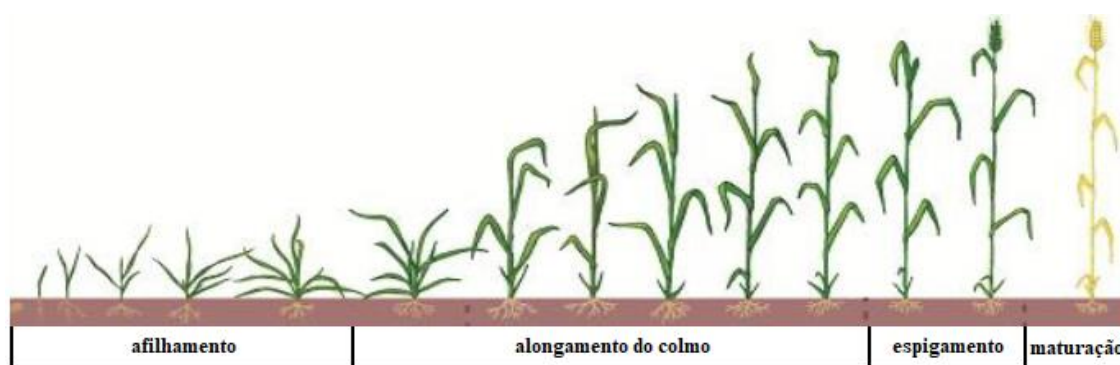
3. 1 CULTURA DO TRIGO E PREDIÇÃO DE PRODUTIVIDADE

O trigo é uma gramínea de ciclo anual cultivada em todo o mundo. Recentemente, o consumo desse cereal vem aumentando globalmente, inclusive em países com climas inadequados para a sua produção, principalmente porque alimentos à base de trigo fornecem uma variedade de componentes essenciais e benéficos para a dieta humana, incluindo proteínas e vitaminas (SHEWRY et al., 2015).

Com uma produção de 5,48 milhões de toneladas na safra de 2019, o Brasil está entre os maiores produtores de trigo do mundo. O cultivo desse cereal no país se concentra principalmente nos estados da região Sul, por possuírem clima mais favorável à essa cultura. Essa limitação geográfica é um dos motivos da necessidade brasileira de importar trigo de outros países para atender a demanda nacional, o que acarreta em custos que chegaram a até quatro bilhões de reais no período 2016/2017 (CONAB, 2017).

O trigo é cultivado no Brasil durante o inverno e o seu ciclo tem uma duração que pode variar de 90 a 180 dias, dependendo do genótipo, das condições climáticas e do solo. Em termos gerais, o ciclo de desenvolvimento da cultura do trigo pode ser dividido nas fases de afilamento, alongamento do colmo, espigamento e maturação. A Figura 1 ilustra as fases do ciclo de desenvolvimento do trigo, de acordo com a escala Feekes-Large (LARGE, 1954).

Figura 1. Escala Feekes-Large de desenvolvimento da cultura de trigo.



Fonte: adaptado de Pires et al. (2011).

Tendo em vista sua grande importância econômica e nutricional em diversos países, a cultura do trigo tem sido objeto de constantes pesquisas que procuram aumentar sua produtividade. Em muitos estudos com trigo podem ser aplicados os conceitos de AP, desde a

coleta de dados em uma determinada área, suas análises e a atuação com base nas mesmas (BERNARDI et al., 2013). As contribuições da AP para o aumento da produtividade de trigo podem ser significativas em diferentes objetivos de pesquisa, como em análises de solos (KRAMER et al., 2014), monitoramento de parâmetros agronômicos (SCHIRRMANN et al., 2016) e acompanhamento do crescimento das plantas (HOLMAN et al., 2016).

A predição de produtividade de grãos também se apresenta como opção ao se utilizar os conceitos da AP para potencializar o rendimento de uma determinada cultura. Esse objetivo de pesquisa vem sendo bastante utilizado recentemente em diversos estudos, sobretudo para a predição da produtividade de grãos de culturas anuais. O Quadro 1 apresenta uma relação de estudos realizados nos últimos anos com o objetivo de prever a produtividade de grãos em diferentes culturas.

Quadro 1 – Estudos de predição de produtividade de grãos em culturas anuais.

Autores	Cultura	Objetivo
Rodrigues et al. (2015)	Trigo	Correlacionar índices de vegetação e a condutividade elétrica do solo com rendimentos de grãos.
Yu et al (2016)	Soja	Correlacionar o rendimento dos grãos com medições geométricas das plantas obtidas por meio de imagens aéreas.
Battude et al. (2016)	Milho	Desenvolver uma metodologia para fornecer estimativas precisas da biomassa e do rendimento do milho em grandes áreas.
Gerke (2017)	Trigo	Estimativa de produtividade de trigo por meio de dados espectrais de alta resolução.
Silvestro et al. (2017)	Trigo	Determinar a produtividade por meio da obtenção dos índices de área foliar e de cobertura do dossel estimados por meio de imagens de satélite.
Richetti et al. (2018)	Soja	Estimar o rendimento da soja por meio do índice de vegetação aprimorado (EVI).

Fonte: O autor.

A relação dos trabalhos apresentados no Quadro 1 demonstra que existem diferentes formas de se obter uma predição de produtividade. Esse processo pode utilizar apenas dados espectrais (GERKE, 2017) e suas combinações (RICHETTI et al., 2018) ou fazer uso de outras variáveis agronômicas, como dados de solo (RODRIGUES et al., 2015) e informações a

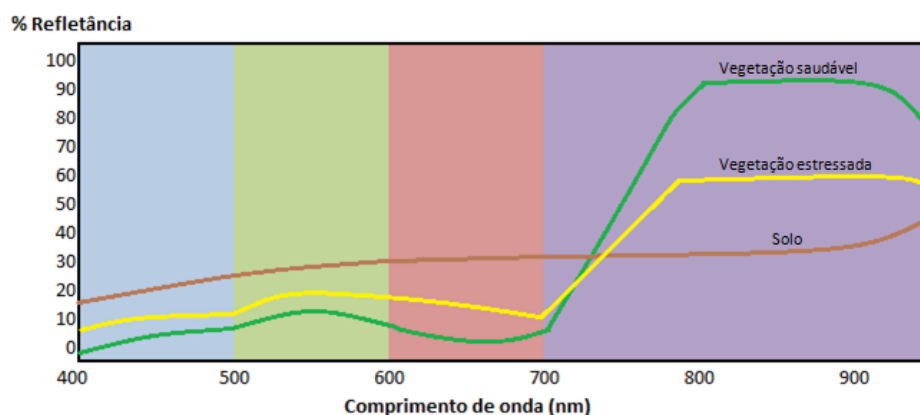
respeito do dossel (SILVESTRO et al., 2017) ou mesmo da altitude da área (YU et al., 2016). Um fator em comum entre esses estudos e que pode auxiliar em aplicações de AP ao facilitar a etapa de coleta de dados é a utilização de tecnologias que aplicam os conceitos de SR.

3.2 SENSORIAMENTO REMOTO E AERONAVES REMOTAMENTE PILOTADAS

O SR de uma área permite que sejam obtidas informações sobre objetos na superfície terrestre sem que haja contato físico entre eles e o sensor (FIGUEIREDO, 2005). Quando uma fonte de energia como a luz solar, por exemplo, é inserida sobre um objeto, parte dela pode atravessá-lo, parte pode ser absorvida e outra parte pode ser refletida e capturada por um sensor. O sensor, então, mede a radiação refletida pelo objeto, a qual pode corresponder a diferentes comprimentos de uma de suas bandas espectrais (MENESES; ALMEIDA, 2012). A identificação do alvo é possível uma vez que cada tipo de objeto apresenta diferentes padrões de refletância da radiação inserida sobre si.

Dentro da agricultura, a quantidade de radiação refletida por uma planta pode indicar sua presente condição de desenvolvimento. Uma planta saudável, geralmente, têm menor refletância nas bandas visíveis e maior refletância nas bandas não visíveis em relação à uma planta não saudável (CHANG et al., 2013). A Figura 2 apresenta a diferença da refletância de plantas saudáveis, plantas estressadas (não saudáveis) e do solo exposto em diferentes comprimentos de onda abrangendo o espectro visível (comprimentos de onda entre 400 e 700 nm) e não visível (a partir de 700 nm).

Figura 2. Refletância de plantas saudáveis, plantas estressadas e do solo em diferentes comprimentos de onda de um sensor.



Fonte: adaptado de Chang et al. (2013).

Os tipos de sensores utilizados para a medição da refletância podem ser divididos em ativos, os quais emitem radiação artificial e registram a resposta (ou reflexão) do alvo, ou passivos, que apenas fazem o registro de uma radiação de origem natural (RAY, 2016). Os Sistemas de Varredura Laser Terrestre (SVLTs) são exemplos de sensores ativos que utilizam a tecnologia LiDAR (obtenção de medidas por meio da luz– do inglês *Light Detection and Ranging*) para medir coordenadas ou mapear a superfície de uma determinada área. Os SVLTs passaram a ser recentemente utilizados em estudos a fim de se obter estimativas de altura de indivíduos em cultivos agrícolas e apresentaram resultados significativos em diferentes objetos de pesquisa, como cevada (TILLY et al., 2015), plantas daninhas (KOENIG et al., 2015), milho e soja (FRIEDLI et al., 2016).

Entre os sensores passivos, embora os satélites tenham sido bastante utilizados durante anos em diversas aplicações agrícolas, recentemente houve o crescimento da utilização de RPA, sobretudo por não dependerem do período orbital e por apresentarem custos financeiros e de tempo mais acessíveis para estudos em ambientes agrícolas (MOECKEL et al., 2018).

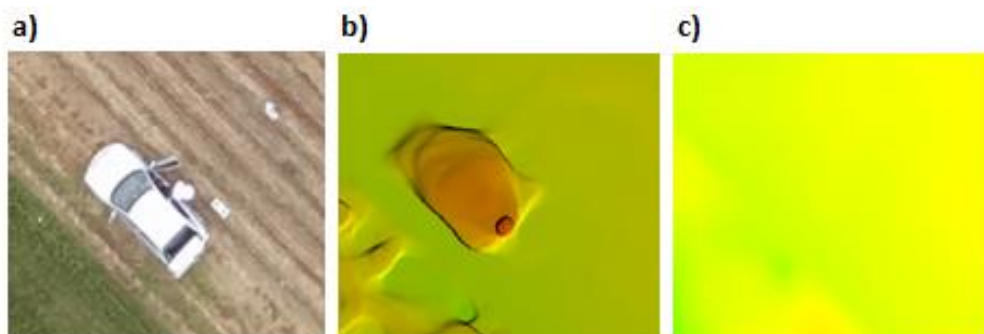
Holman et al. (2016) fizeram uso de uma RPA Octocopter equipada com uma câmera RGB para coletar dados de altura de plantas de trigo visando produzir um método de medição preciso em substituição aos métodos manuais. Também foi utilizada uma SVLT a fim de comparar o desempenho de ambos os sensores. O estudo concluiu que, embora o método utilizando SVLT tenha apresentado alta exatidão, a sua aplicação é problemática devido ao grande número de varreduras necessárias na área de estudo e, conseqüentemente, impactando no tempo gasto com as mesmas. A RPA apresentou maior potencial para o monitoramento em alta precisão de altura de plantas. O estudo também apontou como uma possibilidade de pesquisa a utilização de imagens NIR, por aumentarem o contraste entre a planta e o solo possibilitando aumentar a eficiência do modelo.

3. 3 PROCESSAMENTO DIGITAL DE IMAGENS E MODELOS 3D

A utilização de softwares para processamento de imagens obtidas por SR pode trazer informações significativas a respeito da área imageada. Os softwares baseados em algoritmos *Structure from Motion* (SfM) utilizam essa técnica fotogramétrica que permite a reconstrução tridimensional de uma área a partir de imagens bidimensionais sobrepostas (WESTOBY et al., 2012). Por meio desse processamento é também possível a criação dos modelos 3D denominados Modelo Digital de Superfície (DSM – do inglês *Digital Surface Model*) e Modelo Digital de Elevação (DEM - do inglês *Digital Elevation Model*). A Figura 3 apresenta um

exemplo após o processamento de um voo que gerou o DSM e o DEM e suas diferenças em relação à imagem real da área.

Figura 3. Imagem aérea de um carro (a) e os modelos DSM (b) e DEM (c) gerados para a mesma área.



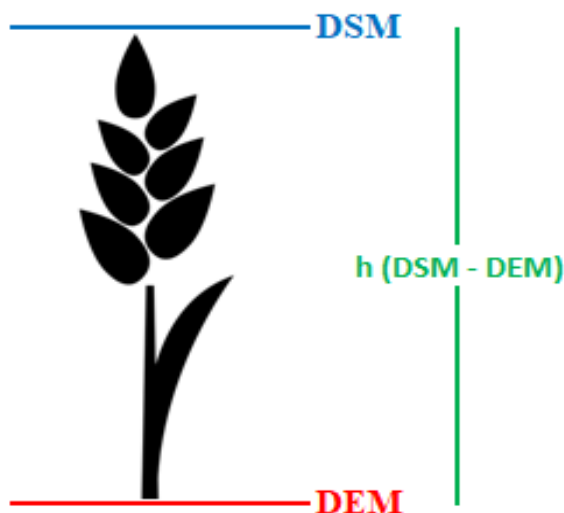
Fonte: O autor .

A principal diferença entre os modelos DSM e DEM é que o primeiro contempla toda a superfície em uma área imageada, incluindo objetos naturais e artificiais, enquanto o segundo contabiliza apenas a altura do terreno, ignorando quaisquer objetos acima do solo (MARTHA et al., 2010). Na Figura 3, é possível verificar as diferenças entre os modelos. A Figura 3a é uma imagem aérea de um carro recortada do ortomosaico. Na Figura 3b, o carro é representado por sua altura em uma escala de cor em que o ponto mais alto em vermelho representa o porta-malas aberto, o que significa que o DSM gerado por essa imagem foi capaz de detectar essa diferença. A Figura 3c é o resultado do DEM da mesma área, onde o carro é ignorado.

As principais aplicações do DSM envolvem a visualização de objetos em áreas civis e reconstrução de paisagens em 3D (RAYBURG et al., 2009), enquanto o DEM é principalmente utilizado para aplicações nas áreas de topografia e geomorfologia (ERASMI et al., 2016). Em áreas onde não há objetos acima da superfície, o DSM apresenta os mesmos valores do DEM.

Ambos os modelos (DSM e DEM), quando calculados por meio de dados de SR em uma área, apresentam uma camada *raster* (matriz) contendo valores de coordenadas e altura correspondentes a cada pixel da imagem dessa área. Essas informações permitem que esses modelos sejam utilizados em estudos que visam a estimativa da altura de objetos acima do solo, uma vez que é possível calcular esse valor por meio da diferença entre os valores gerados para cada um dos modelos (Figura 4).

Figura 4. Altura estimada do objeto (h) obtida por meio da diferença entre os modelos digitais DSM e DEM.



Fonte: adaptado de Holman et al. (2016)

Na agricultura, estudos que utilizam esses modelos vêm sendo realizados para calcular a altura de plantas de diferentes culturas e, com isso, identificar possíveis problemas no seu desenvolvimento. Holman et al. (2016) calcularam os valores dos modelos DSM e DEM com o objetivo de produzir um método preciso e de fácil replicabilidade para coletar dados de altura de plantas de trigo por meio de RPA. Os resultados mostraram que as técnicas SfM foram totalmente capazes de recriar modelos 3D precisos e coletar dados em campo de alta qualidade, embora ainda haja espaço para melhorar sua eficiência ao testar diferentes parâmetros de processamento.

Moeckel et al. (2018) utilizaram técnicas UfM para capturar informações de altura de três vegetais (beringela, tomate e repolho) durante um ciclo completo de desenvolvimento dessas culturas. Os resultados do estudo demonstraram que foi possível estimar a altura das plantas utilizando abordagens de SR e SfM e, com isso, substituir os métodos manuais. Os autores recomendaram mais estudos com outras culturas em diferentes períodos com a finalidade de avaliar qual o melhor estágio de desenvolvimento para realizar a estimativa da altura das plantas.

3. 4 MINERAÇÃO DE DADOS (MD)

A MD possibilita uma forma automática para descobrir padrões de dados sem apresentar as limitações baseadas na ótica humana (BRAGA, 2005). Sua utilização é

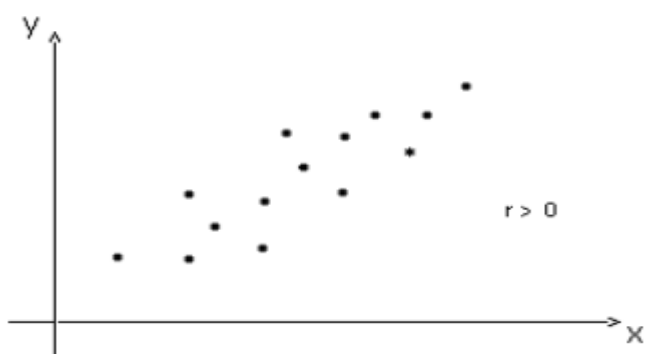
importante principalmente em bases com grandes quantidades de dados, em que a visão humana é incapaz de extrair padrões. Entre as principais funções da MD estão a classificação, cuja a tarefa é categorizar uma ocorrência em um determinado grupo (classe), e a regressão, um método estatístico utilizado para predição numérica (HAN; KAMBER, 2012).

Ao se trabalhar com dados de SR, é comum a geração de grandes bases de dados e a aplicação de MD é, muitas vezes, necessária. Quando se tem como objetivo prever um valor de produtividade, é necessária a utilização de algoritmos de regressão, uma vez que o atributo meta é numérico (quando não discretizado). A regressão é uma técnica estatística que permite correlacionar variáveis independentes com uma variável de resposta (ZBICKI ; DOS SANTOS, 2019). No exemplo do presente estudo, o modelo é gerado a partir da correlação entre os valores de produtividade (variável de resposta ou atributo meta) e os valores digitais das bandas e da altura das plantas (variáveis independentes ou atributos descritivos).

Para a geração dos modelos, são aplicadas técnicas de regressão nas bases de dados pré-processadas, com o atributo meta alocado ao final da base. A regressão linear é uma técnica que pode ser utilizada neste cenário, uma vez que tanto a cor como o tamanho das plantas pode estar entre os indicativos de seu desempenho e, conseqüentemente, de sua produtividade. Existe uma relação linear entre as variáveis quando os dados se aproximam de uma linha reta em um diagrama de dispersão de dados (Figura 5). A qualidade do modelo será dependente da relação entre os atributos e da intensidade com que se relacionam (RODRIGUES, 2012).

Aboelghar et al. (2014) fizeram uso da técnica de regressão linear com o objetivo de gerar modelos de previsão para a cultura de trigo por meio de dados espectrais de imagens de satélite. Os resultados demonstraram que a regressão linear foi a melhor técnica para representar os dados de entrada e a produtividade do trigo, sobretudo para a bandas RED, NIR e suas combinações. Os valores de correlação (r) obtidos foram de até 0,93.

Figura 5. Diagrama de dispersão de dados de uma relação linear entre duas variáveis.



Fonte: adaptado de Santos (2007).

O modelo gerado pela regressão linear quando utilizado para a predição da produtividade de uma cultura pode ter seu desempenho prejudicado se houver fatores que alterem a produtividade da planta, mas que não sejam explicados quantitativamente pelos atributos independentes. Uma alternativa para isso é a utilização de regressores que são capazes de gerar modelos tanto lineares quanto não lineares. Um exemplo de regressor que pode ser aplicado neste cenário é o algoritmo Máquina de Vetores de Suporte (SVM – Do inglês *Support Vector Machines*), o qual também possui aplicação em classificações e procura encontrar uma separação ótima linear ou um limite que separa os dados de uma classe para outra (classificação) ou um limite que se ajuste aos dados numéricos (regressão) (HAN; KAMBER, 2012).

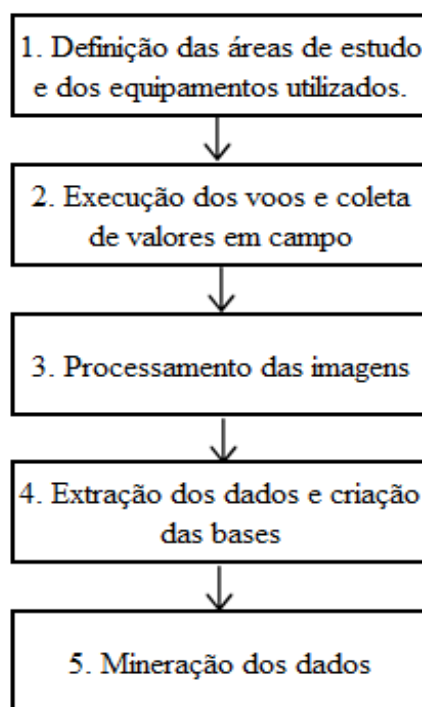
Gerke (2017) utilizou o algoritmo baseado em SVR SMOReg com o objetivo de estimar a produtividade de trigo por meio de dados de RPA e obteve valores de correlação (r) de até 0,61. Considerando que, nesse estudo, foram utilizados como atributos apenas os valores digitais das câmeras e suas combinações, é possível que uma variável auxiliar, como a altura das plantas, ajude a construção de um melhor modelo de estimativa.

Outra técnica de aprendizado de máquina que pode servir tanto para classificação quanto para regressão são as redes neurais artificiais (RNA). Os algoritmos que utilizam RNA possibilitam encontrar limiares para dados por meio de treinamento e ajuste dos pesos de suas conexões sinápticas, simulando um neurônio humano (RUSSEL; NORVIG, 2013). A utilização de RNA pode ser útil para problemas que envolvem dados lineares e não lineares. Panda et al. (2010) utilizaram RNA para predizer a produtividade de milho utilizando índices de vegetação combinados de dados espectrais por meio de imagens de satélite. O valor do coeficiente de determinação (R^2) obtido entre o Índice de Vegetação Perpendicular (PVI) e a produtividade de milho foi de até 0,72.

4 MATERIAL E MÉTODOS

A metodologia utilizada nesse trabalho se iniciou com a definição das áreas de estudo e do material utilizado. As etapas seguintes corresponderam a todo o processo necessário para a elaboração das bases de dados. Por fim, a etapa de mineração de dados teve como objetivo pré-processar e processar os dados obtidos e correlacioná-los com os valores reais coletados em campo para a criação dos modelos de predição de produtividade de trigo. A Figura 6 apresenta o fluxograma contemplando as etapas do trabalho.

Figura 6. Fluxograma apresentando as principais etapas do projeto.



Fonte: O autor.

Foram planejadas as execuções das etapas 2, 3, 4 e 5 em três áreas experimentais envolvendo dois anos diferentes (2018 e 2019) para melhor validação das técnicas utilizadas e dos modelos obtidos.

4.1 ÁREAS DE ESTUDO E EQUIPAMENTOS UTILIZADOS

As áreas experimentais utilizadas nesse estudo foram cultivadas com trigo na região dos Campos Gerais do Paraná, no município de Ponta Grossa. A primeira área de estudo,

denominada Área 1, estava localizada na Fazenda Santa Cruz (Figura 7) e situada nas coordenadas geográficas 25°13'31,3" de latitude Sul e 50°07'07,4" de longitude Oeste. Nessa área, as parcelas foram primeiramente divididas de acordo com a inoculação e não-inoculação de sementes de trigo com a bactéria *Azospirillum brasilense*. As parcelas foram subdivididas em dois tratamentos, sem e com aplicação de gesso agrícola, e as subsubparcelas receberam quatro diferentes doses de nitrogênio (N) em cobertura. O experimento contou com três repetições, totalizando 48 unidades experimentais de 52m² cada uma. O trigo, cultivar TBIO Sossego, foi semeado no dia 14 de junho de 2018.

Figura 7. Área experimental 1 (Fazenda Santa Cruz).



Fonte: O autor.

A segunda área experimental, denominada Área 2, estava localizada na Fazenda Escola “Capão da Onça” da Universidade Estadual de Ponta Grossa (UEPG) (Figura 8), situada nas coordenadas geográficas 25°05'58.7" de latitude Sul e 50°02'34.4" de longitude Oeste. Um experimento foi realizado nessa área por meio de aplicações de duas fontes de potássio (KCl e Polialita), gesso e suas combinações. O experimento contou com oito tratamentos e quatro repetições, totalizando 32 parcelas com 48 m² cada uma. O trigo, cultivar Quartzo, foi semeado no dia 04 de julho de 2018.

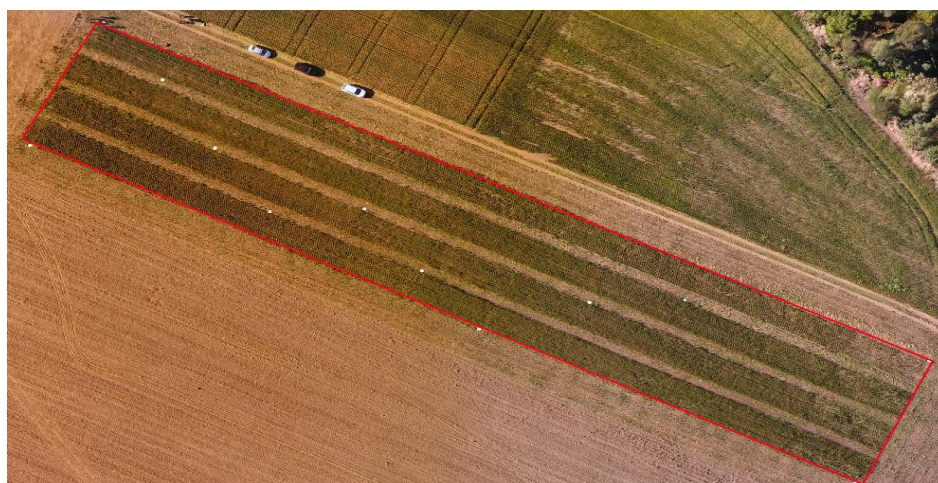
Figura 8. Área experimental 2 (Fazenda Escola da UEPG).



Fonte: O autor.

A terceira área de estudo, denominada Área 3, também localizada na Fazenda Escola da UEPG (Figura 9), está situada nas coordenadas geográficas 25°05'36.1" de latitude Sul e 50°02'46.3" de longitude Oeste. Um experimento foi realizado nessa área com aplicações de calcário, fontes de fósforo (P) e modos de aplicação de P (à lanço e no sulco de semeadura). O experimento contou com quatro repetições, totalizando 44 parcelas de 90 m² cada uma. O trigo, cultivar TBIO Toruk, foi semeado no dia 28 de junho de 2019.

Figura 9. Área experimental 3 (Fazenda Escola da UEPG).



Fonte: O autor.

A colheita na Área 1 foi realizada no dia 06 de novembro de 2018 e as produtividades de grãos variaram de 1172 a 3342 kg ha⁻¹. A colheita dos grãos na Área 2 foi realizada no dia 22 de novembro de 2018 e as produtividades de grãos ficaram entre 2196 e 4431 kg ha⁻¹. A colheita na Área 3, na safra seguinte, foi realizada no dia 14 de novembro de 2019 e as produtividades de grãos variaram de 1328 a 3892 kg ha⁻¹. Os dados de produtividade mínima (Min), máxima (Max), média de grãos e o desvio padrão (DP) entre as parcelas estão apresentados na Tabela 1.

Tabela 1 – Descrição dos dados de produtividade das áreas de estudo.

Áreas	Produtividade (kg ha ⁻¹)			
	Min	Max	Média	DP
Área 1 (Fazenda Santa Cruz)	1172,1	3341,7	2159,8	403,7
Área 2 (Fazenda Escola da UEPG)	2196,0	4430,6	3536,5	552,7
Área 3 (Fazenda Escola da UEPG)	1328,0	3892,1	2414,5	476,4

Fonte: O autor.

Os voos nas áreas 1 e 2 foram realizados utilizando uma RPA de asa fixa modelo eBee, fabricada pela empresa SenseFly (Figura 10a). Essa RPA tem como principais vantagens a autonomia e a facilidade de uso, sendo ideal para a captura de fotos de alta resolução (SenseFly, 2019). Para o controle desse equipamento, foi utilizado o software Emotion 2, o qual permite realizar todo o processo de planejamento dos voos, ajuste de configurações e o pré-processamento das imagens através da inserção das coordenadas em cada foto, de acordo com o GPS embarcado na RPA.

Figura 10. RPA eBee (a), câmera RGB (b) e câmera NIR (b).



Fonte: SenseFly (2019)

Acopladas à RPA, foram utilizadas duas câmeras, uma Sony Cybershot WX RGB com 18.2 megapixels (Figura 10b) e uma Canon S110 NIR de 12 megapixels (Figura 10c). A câmera RGB captura os dados no espectro visível através das bandas RED (vermelho), GREEN (verde) e BLUE (azul). Os valores numéricos de cada uma das bandas RGB variam entre 0 a 255 (8 bits). A câmera NIR também possui as bandas RED e GREEN e apresenta um filtro infravermelho na banda BLUE que a permite capturar valores na região do NIR (infravermelho próximo). Os valores numéricos de cada uma de suas bandas variam entre 0 a 65.535 (16 bits).

Em 2019, devido a problemas técnicos na RPA eBee, o voo sobre a Área 3 foi realizado utilizando uma RPA Phantom 4, fabricada pela empresa DJI (Figura 11). A câmera utilizada pela aeronave foi uma RGB com 12.0 megapixels.

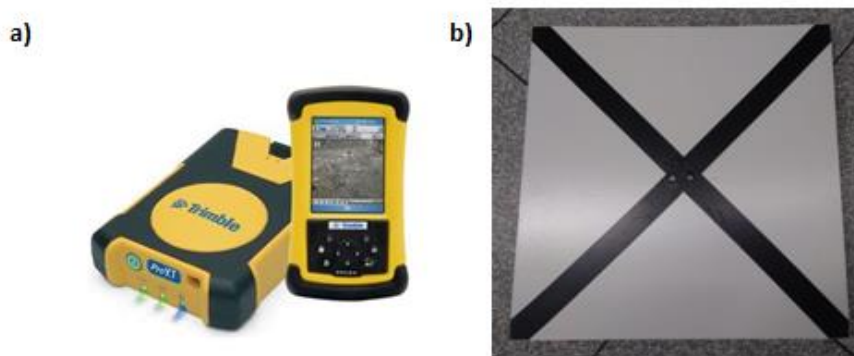
Figura 11. RPA Phantom 4.



Fonte: DJI (2019).

Foram utilizados um receptor GPS modelo Trimble ProXT juntamente com o coletor de dados Juno SB (Figura 12a) para uma melhor precisão no georreferenciamento durante e após o processamento das imagens. As coordenadas foram coletadas em diferentes posições das áreas de acordo com os pontos de controle GCPs (do inglês *Ground Control Points*) (Figura 12b), cujas estruturas facilitam a observação aérea do ponto de coleta. Oito GCPs foram utilizados para a Área 1, dez para a Área 2 e doze para a Área 3. O objetivo foi fazer com que em cada imagem fossem visualizados pelo menos 2 GCPs, que é o número ideal recomendado pelo software de processamento utilizado.

Figura 12. Receptor GPS Trimble ProXT e coletor de dados Juno SB (a) e marcadores GCP (b).



Fonte: O autor.

Na coleta dos pontos, a antena do GPS foi posicionada no centro do GCP durante alguns minutos e cerca de 500 pontos, em média, foram coletados em cada GCP. Os pontos foram extraídos do GPS utilizando o software Trimble GPS Pathfinder e exportados em arquivos RINEX para serem enviados ao Instituto Brasileiro de Geografia e Estatística (IBGE) através do sistema IBGE-PPP (Posicionamento por Ponto Preciso), um serviço online gratuito para o pós-processamento de dados GNSS (Global Navigation Satellite System). Esse serviço permitiu obter coordenadas referenciadas ao SIRGAS2000 que é o Sistema de Referência Geocêntrico para as Américas (IBGE, 2018). Após o pós-processamento, os dados contendo latitude, longitude e altitude de cada ponto foram salvos em arquivos .xyz para serem utilizados posteriormente nas etapas de processamento das imagens (sessão 4.3) e extração dos dados (sessão 4.4).

4.2 COLETA DOS DADOS

Na etapa de coleta dos dados, os voos foram realizados sobre as áreas durante o desenvolvimento da cultura do trigo utilizando diferentes resoluções e câmeras. O voo sobre a Área 1 foi realizado no dia 11 de setembro de 2018, aos 89 dias após a semeadura (DAS), utilizando as câmeras RGB e NIR em três resoluções diferentes (3,4, 10 e 26 cm/pixel). O voo na Área 2 foi realizado no dia 10 de outubro de 2018, aos 98 DAS, utilizando a câmera RGB em duas resoluções (3,4 e 10 cm/pixel). O voo planejado com a câmera NIR sobre a Área 2 não foi realizado devido às constantes chuvas durante todo período de crescimento da cultura. O voo sobre a Área 3 foi realizado no dia 09 de outubro de 2019, aos 103 DAS, e apenas a câmera RGB foi utilizada, por ser a única disponível na RPA, com resolução de 5 cm/pixel. Em todos

os voos realizados o trigo estava em fase de espigamento. As informações referentes aos voos nas três áreas experimentais são apresentadas na Tabela 2.

Tabela 2 – Datas, resoluções e dias após a semeadura (DAS) dos voos realizados.

Área experimental	Data	Resolução	DAS
Área 1 (Fazenda Santa Cruz)	11/09/2018	3.4 cm/pixel	89
		10 cm/pixel	
		26 cm/pixel	
Área 2 (Fazenda Escola da UEPG)	10/10/2018	3.4 cm/pixel 10 cm/pixel	98
Área 3 (Fazenda Escola da UEPG)	09/10/2019	5 cm/pixel	103

Fonte: O autor.

Realizou-se também a medida manual da altura das plantas nas áreas 1 e 2 nos mesmos períodos dos voos, com o objetivo de validar o processo de estimativa da altura de plantas. A altura foi medida a partir da base da planta até a extremidade da sua espiga mais alta. Foram coletadas 10 plantas de cada parcela das áreas de estudo e a altura foi definida pelo valor médio das medidas.

4.3 PROCESSAMENTO DAS IMAGENS

Para o processamento das imagens foi utilizado o software Pix4DMapper PRO. Esse software recebe como entrada as imagens pré-processadas pelo software Emotion 2 e tem como saída diferentes produtos a respeito da área mapeada, como o ortomosaico (Figura 14a) e o mapa de refletância (Figura 14b) (PIX4D, 2017).

O processamento das imagens pelo software foi dividido em quatro etapas: na primeira etapa o software realizou a busca por pontos iguais (*matching*) entre as imagens, utilizando-se de informações sobre suas localizações no mapa; na segunda etapa foi realizada a correção georreferencial das imagens, utilizando as coordenadas coletadas pelo GPS nos GCPs distribuídos nas áreas (sessão 4.1). O arquivo de texto .xyz foi importado para o software e foi possível visualizar as localizações de cada GCP na área e os erros georreferenciais em relação aos pontos coletados. Esses erros são explicados pela diferença de precisão entre o GPS embarcado na RPA utilizado para a junção das imagens na primeira etapa e o GPS utilizado em solo com coordenadas pós-processadas. A correção desses erros foi realizada identificando e

marcando manualmente o centro de cada GCP da área em todas as imagens em que o mesmo foi visualizado. Na Figura 13 é possível visualizar a distância entre o ponto coletado pelo GPS (em azul) e o centro de um dos GCP (em verde) em cada imagem que ele aparece.

Figura 13. Distância entre o centro de um GCP e o ponto coletado pelo GPS em solo nas 4 imagens em que ele é visualizado na Área 3.



Fonte: O autor.

Na terceira etapa foi gerada a nuvem de pontos (*point cloud*), sendo que cada pixel da nuvem tem uma coordenada que foi associada com uma imagem original da área. Na quarta e última etapa foram gerados os produtos finais do processamento, sendo eles o ortomosaico (uma única imagem abrangendo toda a área sobrevoada com balanceamentos de cores realizados pelo próprio software para melhor visualização) (Figura 14a), o mapa de refletância (um ortomosaico sem ajustes de cores) (Figura 14b) e os modelos DSM (Figura 15a) e DEM (Figura 15b). Para se trabalhar com os valores digitais das imagens, o ideal é utilizar o mapa de refletância por ele não apresentar balanceamento de cores durante o processamento e permanecer com os valores digitais originais (PIX4D, 2017). Os modelos DSM e DEM podem ser observados no próprio software e salvos em arquivos com formato .tif e lidos como arquivos *raster* em softwares SIGs (Sistemas de Informação Geográfica) contendo, para cada

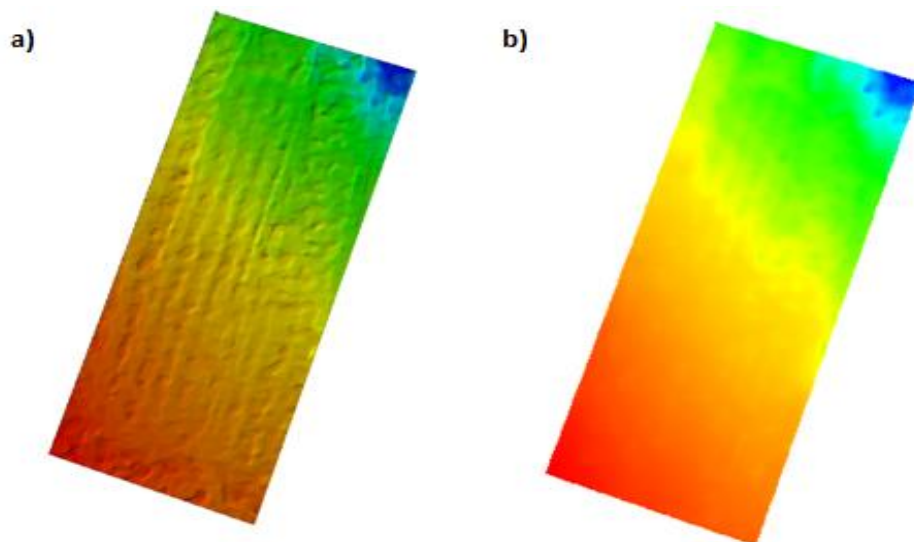
ponto, as coordenadas e a altitude do objeto mais alto acima do solo (DSM) ou do próprio solo ignorando os objetos (DEM).

Figura 14. Ortomosaico (a) e mapa de refletância (b) gerados para a Área 1.



Fonte: O autor.

Figura 15. DSM (a) e DEM (b) gerados sobre a Área 1.



Fonte: O autor.

Por ser baseado em algoritmos *Structure from Motion* (SfM), o software Pix4D é capaz de criar um mapa 3D da área baseando-se em imagens 2D. A reconstrução 3D de uma região pode ser observada no próprio software na aba *Volumes*. A Figura 16 apresenta uma reconstrução 3D da Fazenda Santa Cruz (Área 1), a partir do voo realizado sobre a área. É

possível observar que a reconstrução foi capaz de recriar a diferença entre o corredor da área e as plantas.

Figura 16. Reconstrução 3D da Área 1 utilizando imagens da câmera RGB.



Fonte: O autor.

4.4 EXTRAÇÃO DOS DADOS

O processo de extração dos dados das imagens e dos modelos foi realizado por meio do Quantum GIS (ou QGIS), um software livre que fornece ferramentas para visualização, edição e análise de dados georreferenciados (QGIS, 2019). O QGIS recebe como entrada os produtos gerados pelo Pix4D lidos como camadas *raster* contendo pontos de coordenadas para cada pixel das imagens.

4.4.1 Extração dos valores digitais

A extração dos valores digitais das câmeras RGB e NIR foi realizada sobre o mapa de refletância, o qual não tem seus valores digitais alterados durante o processamento. Tendo em vista que o valor da produtividade de trigo obtido no campo foi contabilizado para cada parcela, foi necessário separar cada uma delas na imagem antes de extrair os dados. Esse processo pode ser feito coletando pontos de GPS nas divisões das parcelas ou analisando visualmente a imagem (quando há algum tipo de marcador identificando cada divisa). Se ambos os métodos estiverem comprometidos, é possível também separar cada parcela por meio de métricas digitais em editores de imagem, considerando o croqui da área, suas medições em campo e a

resolução da imagem. Após essa identificação, são desenhados polígonos em cada parcela considerando o local em que foi realizada a colheita e a sua área de extensão (Figura 17).

Nas áreas 1 e 2, o local colhido contabilizado no estudo foi o centro das parcelas, desconsiderando as bordaduras, com o objetivo principal de minimizar interferências de tratamentos vizinhos. Cada polígono desenhado tinha um identificador *id*, e o seu número foi dado de acordo com o número da parcela definido no croqui da área.

Figura 17. Divisões e polígonos desenhados para cada parcela da Área 2.



Fonte: O autor.

Diferentemente das outras áreas, a colheita da Área 3 não foi realizada nos centros das parcelas e sim em locais onde as plantas foram menos prejudicadas em decorrência de uma infestação de azevém na área. Portanto, a localização em que os polígonos deveriam ser desenhados era variável para cada parcela. Dentro dessas condições, foi necessária a utilização de GPS para obter as coordenadas nos locais onde a colheita foi realizada. Os locais colhidos eram retangulares e contínuos, com 2 m de largura e 3 m de comprimento. A coleta dos pontos foi realizada antes da colheita do restante da área, e em cada uma das parcelas o GPS foi posicionado no canto superior direito da área colhida. Essas coordenadas foram pós-processadas e adicionadas ao QGIS como uma nova camada, permitindo a localização desses pontos no ortomosaico da área (Figura 18).

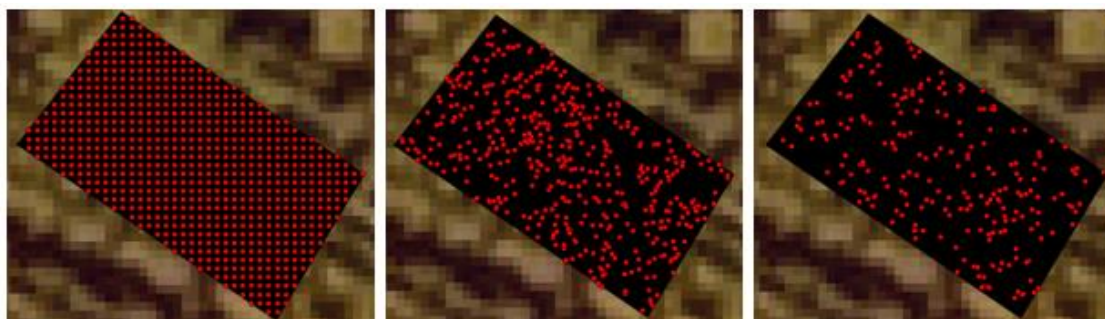
Figura 18. Pontos de coleta indicando os locais de colheita em cada parcela na Área 3.



Fonte: O autor.

Para a etapa de geração dos pontos foi utilizada a ferramenta *Pontos aleatórios no interior de polígonos*. Essa ferramenta permite que seja definida previamente a quantidade de pontos gerados dentro de cada polígono, possibilitando que menores amostras de dados sejam criadas. Para cada imagem foram gerados, primeiramente, um ponto para cada um dos pixels no interior de cada polígono; posteriormente foram gerados pontos para 66% e 33% dos pixels. Com isso, foram criadas três amostras de dados para cada imagem gerada. A Figura 19 apresenta um exemplo de polígono desenhado na Área 3 e as diferentes quantidades de pontos geradas.

Figura 19. Exemplo de um polígono com diferentes quantidades de pontos por pixel (100%, 66% e 33%).



Fonte: O autor.

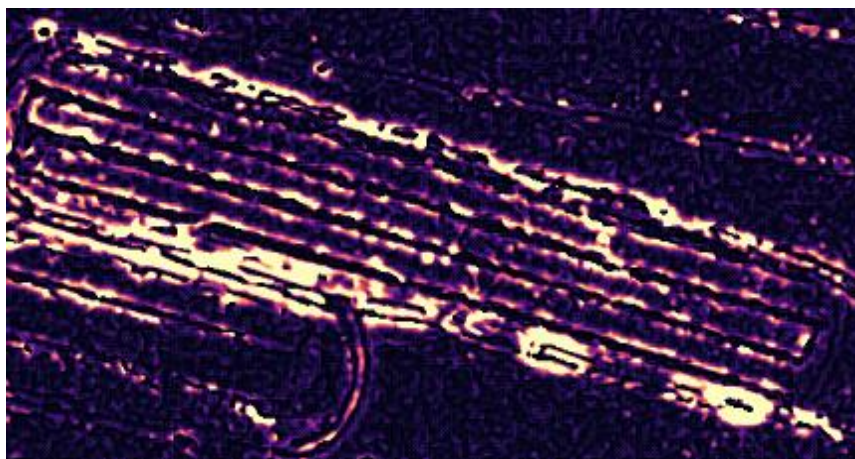
Dentro dos polígonos desenhados foram extraídos os valores digitais de cada banda utilizando o *plugin Point Sampling Tool*, também disponível para instalação no software QGIS.

Os valores digitais extraídos foram salvos em arquivos com formato .csv contendo também a identificação das parcelas definidas pelo *id* dos polígonos.

4.4.2 Cálculo e obtenção das alturas das plantas

Os valores de altura das plantas foram calculados a partir dos modelos DSM e DEM gerados pelo software Pix4D. Uma vez que no processamento de cada voo foram gerados ambos os modelos, os dois possuem a mesma coordenada para cada pixel. Isso permite que o QGIS os considere como duas camadas *rasters* diferentes na mesma área. A ferramenta *calculadora raster*, disponível no software, permite a realização de cálculos dos atributos entre duas camadas diferentes que possuem as mesmas coordenadas. Utilizando a operação de subtração entre o DSM e DEM, foi criada uma nova camada *raster*, a qual pode ser visualizada como um mapa de altura da área, contando com as mesmas coordenadas dos modelos, mas com os valores de diferença de altitude entre ambos que seriam as alturas dos objetos acima do solo. Na Figura 20, os pontos mais claros representam os objetos mais altos.

Figura 20. Mapa de altura das plantas de trigo obtido pelos dados da câmera RGB na Área 1.



Fonte: O autor.

Para a extração dos valores dos mapas de altura foram utilizados os mesmos arquivos contendo os polígonos desenhados e os pontos gerados no mapa de refletância. A ferramenta *Point Sampling Tool* foi novamente utilizada para extrair os valores de altura para cada pixel dentro de cada parcela. Assim como na extração dos valores digitais, os valores de altura também foram salvos em arquivos .csv contendo a identificação das parcelas definidas pelo *id* dos polígonos.

4.5 MINERAÇÃO DOS DADOS

A etapa de mineração de dados contemplou a junção dos dados obtidos pela etapa anterior, a análise e o pré-processamento desses dados e a aplicação de técnicas de aprendizado de máquina a fim de correlacionar os dados extraídos das imagens com os dados coletados em campo.

4.5.1 Criação e pré-processamento das bases de dados

O conjunto dos dados obtido nas etapas anteriores foi subdividido de acordo com a área em que foi realizado o voo, a câmera utilizada, a resolução da imagem e a proporção dos pontos extraídos em relação à sua totalidade (Tabela 3). Ao todo, foram geradas seis bases para a Área 1, duas bases para a Área 2 e uma base para a Área 3.

Os dados gerados pela câmera NIR utilizando 3.4 cm/pixel na Área 1 foram reduzidos devido a três parcelas terem sido comprometidas por falhas na junção das imagens. A causa disso pode ter sido o excesso de vento no momento do voo que alterou a posição da RPA e prejudicou a captura das imagens sobre as parcelas localizadas na extremidade da área. Também foram descartadas quatro parcelas da Área 3 por terem sido desconsideradas no momento da colheita.

Tabela 3 – Tamanho em instâncias de cada uma das bases criadas.

Base	Área	Câmera	Resolução (cm/pixel)	Instâncias		
				100%	66%	33%
A1_RGB_3	1	RGB	3,4	365.616	243.744	121.872
A1_NIR_3	1	NIR	3,4	342.765	228.510	114.255
A2_RGB_3	2	RGB	3,4	243.744	162.496	81.248
A1_RGB_10	1	RGB	10	33.600	22.400	11.200
A1_NIR_10	1	NIR	10	33.600	22.400	11.200
A2_RGB_10	2	RGB	10	22.400	14.934	7467
A1_RGB_26	1	RGB	26	6.240	4.160	2.080
A1_NIR_26	1	NIR	26	6.240	4.160	2.080
A3_RGB_5	3	RGB	5	92.400	61.600	30.800

Fonte: O autor.

O número de instâncias das bases é consideravelmente maior de acordo com a resolução da imagem e a proporção dos dados. Para cada base foram criadas outras duas com menores quantidades de amostras de dados (sessão 4.4.1). O objetivo da criação de diferentes quantidades de amostras foi verificar se existe a necessidade da utilização da totalidade de dados quando se trabalha com valores digitais. Isso é importante uma vez que o custo computacional aumenta consideravelmente de acordo com o tamanho da base.

Em relação às colunas das bases, cada uma é composta de um atributo de identificação (*id* da parcela), cinco atributos descritivos e um atributo meta (produtividade). O Quadro 2 apresenta o nome, o tipo e a descrição desses atributos. O atributo com os valores reais da altura das plantas também foi utilizado como meta em alguns testes.

Quadro 2 – Descrição dos atributos das bases de dados.

Atributo	Tipo	Descrição
Parcela	Numérico contínuo	<i>id</i> da parcela. Não é utilizado no momento da aplicação dos algoritmos de mineração de dados.
Banda_1 (RED)	Numérico contínuo	Valor digital da banda RED das câmeras RGB ou NIR. É contabilizado por pixel e pode variar para cada instância da base.
Banda_2 (GREEN)	Numérico contínuo	Valor digital da banda GREEN das câmeras RGB ou NIR. É contabilizado por pixel e pode variar para cada instância da base.
Banda_3 (BLUE ou NIR)	Numérico contínuo	Valor digital da banda BLUE (câmera RGB) ou NIR (câmera NIR). É contabilizado por pixel e pode variar para cada instância da base.
Altura_estimada_plantas	Numérico contínuo	Valor de altura estimada extraído dos modelos digitais. É contabilizado por parcela e pode variar de acordo com a quantidade delas.
Altura_real_plantas	Numérico contínuo	Valor de altura real medida em campo. É contabilizado por parcela e pode variar de acordo com a quantidade delas.
Produtividade	Numérico contínuo (meta)	Valor de produtividade coletada em campo. É o atributo meta e alocado na última coluna da base. É contabilizado por parcela e pode variar de acordo com quantidade delas.

Fonte: O autor.

Os valores estimados de altura das plantas, inicialmente contabilizados por pixel, foram transformados para serem contabilizados por parcela de acordo com os valores reais. Para cada parcela foram selecionados aleatoriamente 100 pontos e o valor médio entre eles foi contabilizado como a altura estimada da parcela. A Tabela 4 apresenta um fragmento obtido da base A1_NIR_10.

Tabela 4 – Exemplo mostrando a composição de uma base de dados do estudo (A1_NIR_10).

Instância	Parcela	R	G	NIR	Altura estimada (cm)	Altura real (cm)	Prod. de trigo
1	p1	3805,601	5064,881	27569,803	80,9	91,7	2540,4
2	p1	3849,365	5067,211	25190,053	80,9	91,7	2540,4
3	p1	3891,500	5069,097	25552,645	80,9	91,7	2540,4
4	p1	4061,265	5271,518	27486,207	80,9	91,7	2540,4
5	p1	3827,559	5130,599	26533,309	80,9	91,7	2540,4
6	p1	3773,554	5123,048	29513,279	80,9	91,7	2540,4
7	p1	4290,837	5498,931	26089,924	80,9	91,7	2540,4
8	p1	4099,203	5276,562	26121,141	80,9	91,7	2540,4
9	p1	4313,205	5498,842	26558,920	80,9	91,7	2540,4
...	...	...	...	...	...	...	...
33,600	p48	4604,828	5362,175	21539,652	72,3	93,3	2631,1

Fonte: O autor.

Na etapa de pré-processamento também foi realizada a identificação e o tratamento de possíveis *outliers* (Figura 21), geralmente causados por erros de softwares de manipulação dos dados.

Figura 21. Exemplo de dois *outliers* identificados em uma das bases de dados.

114.89101	118.31524	90.44618
116.49953	119.00747	91.07252
120.01322	119.03926	92.10751
115.61369	119.18722	91.17032
116842	119.19911	91.27845
115.24163	119.2167	92.81045
116.54706	119.53434	91.81485
116.5695	119.58554	91.82706
115.83917	119.68349	93.09587
116.12292	119.75783	91.9154
116.34128	119.9329	93.72724
116.90696	119.9931	92.21362
115.89342	120.01322	93.2625
114.79627	120.02415	92.08028
116.52233	120.03144	92.0229
116.78822	120.0361	91.72299
115.58761	120.08908	92.89226
117.05283	120.1159	93717
116.38116	120.16881	93.35788

Fonte: o autor.

Os *outliers* foram tratados de acordo com seu valor em relação aos das instâncias vizinhas. Para erros de fácil percepção, como a ausência de um ponto ou vírgula, eles foram corrigidos. No caso de erros na coleta dos dados, como o descrito na base A1_NIR_3, as parcelas ou instâncias foram desconsideradas.

4.5.2 Análise dos dados e técnicas de aprendizado de máquina

A análise visual dos dados e a aplicação de técnicas de aprendizado de máquina foram ambas realizadas utilizando o ambiente Weka, um software de código aberto que possui ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização (WEKA, 2019). Especificamente para a etapa de regressão, foi instalado o pacote wekaServer, o qual fornece um servidor simples baseado em *servlet* para execução de tarefas de mineração de dados de maneira distribuída, permitindo que todos os núcleos do processador sejam utilizados ao mesmo tempo, a fim de proporcionar um possível ganho computacional. A execução do servidor foi iniciada por meio da interface CLI do ambiente, e o comando utilizado está descrito no Quadro 3.

Quadro 3 – Comando utilizado para a execução do pacote wekaServer.

```
> java weka.Run WekaServer -host localhost -port 8085 -slots 16 -staleTime -1  
  
# -slots indica quantos núcleos do processador serão utilizados para a execução dos  
algoritmos. O Weka por padrão utiliza apenas 1.  
  
# -staleTime indica o tempo que o servidor aguarda até apagar uma execução parcial já  
realizada. Em execuções muito longas mostrou-se necessário desativar essa opção com o  
valor -1.
```

Fonte: O autor.

As bases de dados foram analisadas antes de serem processadas na ferramenta de visualização disponível no ambiente. O objetivo dessa análise foi verificar a possibilidade de encontrar padrões de comportamento entre os dados que auxiliassem na escolha dos algoritmos de mineração de dados a serem aplicados.

Para a etapa de aplicação de técnicas de aprendizado de máquina, além do algoritmo de regressão linear, foram selecionados para testes iniciais dois algoritmos mais robustos: o MultilayerPerceptron (RNA) e o SMOReg. O Quadro 4 apresenta a relação desses algoritmos, as configurações utilizadas (baseadas em resultados de testes preliminares) e uma breve descrição da própria ferramenta.

Quadro 4 – Algoritmos utilizados para os testes iniciais (fonte: Weka, 2019).

Algoritmo	Parâmetros	Descrição da ferramenta
LinearRegression	<i>default</i>	Utiliza regressão linear para previsão. Usa o critério de Akaike para seleção de modelo e é capaz de lidar com instâncias ponderadas.
MultilayerPerceptron	1 camada oculta com 5 nodos	Utiliza retropropagação para classificar instâncias. Quando a classe é numérica os nodos de saída se tornam unidades lineares sem limite.
SMOReg	Atributos estandarizados	Implementa a máquina de vetores de suporte para regressão.

Fonte: O autor.

Os testes iniciais foram todos realizados em uma mesma base (A1_RGB_10), com e sem a utilização do atributo de altura real das plantas e com as três diferentes quantidades de amostras de dados. Os objetivos desses testes foram: I - verificar qual algoritmo apresentava os melhores resultados para ser utilizado na sequência dos testes; II - verificar se a quantidade de dados utilizados afetava o desempenho dos algoritmos; e III - verificar se a altura real das plantas contribuía para a geração de um modelo melhor de predição de produtividade de grãos. Uma vez definidas essas questões, os testes seguintes em todas as bases tiveram os seguintes objetivos: I - verificar qual câmera e resolução apresentavam o melhor desempenho para estimar a altura das plantas e II - analisar se seria possível gerar um modelo de predição utilizando apenas dados de SR com forte correlação com a produtividade de trigo.

A metodologia utilizada para todos os testes foi de dez *folds* de validação cruzada e as métricas consideradas foram os coeficientes de correlação (r) e de determinação (R^2). Para avaliar o desempenho da estimativa da altura das plantas foi calculado o erro médio, em cm, entre a altura estimada e a altura real obtida por parcela. Também foram correlacionados ambos os valores considerando o coeficiente r . Tendo em vista que o objetivo principal desse trabalho era a predição de produtividade de grãos, a estimativa da altura das plantas foi considerada apenas como uma variável auxiliar para proporcionar ganho de desempenho ao modelo de predição. Em vista disso, a estimativa correta de plantas mais altas que outras em determinada parcela (altura relativa) tem a mesma importância que a estimativa exata da altura de plantas em cada parcela (altura real).

Para calcular as diferenças estatísticas entre os resultados foram realizados testes de Tukey utilizando o PAST, um software livre para análise de dados científicos, com funções para manipulação de dados e estatísticas (PAST, 2020).

5 RESULTADOS E DISCUSSÃO

Nessa sessão serão apresentados os resultados obtidos primeiramente da análise preliminar dos dados e, posteriormente, dos testes descritos na sessão 4.5.2.

5.1 AVALIAÇÃO DOS DADOS

Nas Figuras 22 e 23 são apresentados os comportamentos dos dados das Áreas 1 e 2, respectivamente. As bases foram divididas em três partes iguais e a coloração de cada uma representa a classe de acordo com o seu valor relativo ao restante da base: amarelo=alto (1/3), laranja=médio (1/3) e vermelho=baixo (1/3). Analisando os valores apresentados, calculados pela média entre as bandas RGB, e os valores de altura das plantas e produtividade de grãos, foi possível identificar que, geralmente, os dados se comportavam da seguinte forma:

se RGB=alto e Altura=baixa então Produtividade=baixa,

ou

se RGB=baixo e Altura=alta então Produtividade=alta.

Figura 22. Mapas de refletância RGB (a), altura das plantas (b) e produtividade de grãos (c) da Área 1.

Refletância (RGB)						Altura (cm)						Produtividade (kg ha ⁻¹)								
a)	16	104,3	32	110,4	48	114,2	b)	16	99,1	32	94,6	48	93,3	c)	16	2220,7	32	2468,0	48	2631,1
	15	122,9	31	98,5	47	105,4		15	82,1	31	99,2	47	94,7		15	1619,7	31	1918,1	47	2033,2
	14	107,9	30	124,4	46	110,5		14	93,9	30	83,8	46	93,7		14	2335,8	30	1696,5	46	1811,5
	13	101,8	29	101,6	45	98,2		13	94,3	29	96,1	45	96,2		13	2549,0	29	2497,8	45	1892,5
	12	124,9	28	106,5	44	104,4		12	80,2	28	93,3	44	92,1		12	1777,4	28	2536,2	44	2293,2
	11	98,1	27	98,4	43	112,5		11	94,2	27	96,9	43	89,1		11	2161,1	27	2233,5	43	1709,2
	10	94,4	26	102,8	42	95,7		10	99,4	26	95,3	42	98,7		10	2340,1	26	2278,5	42	2348,1
	9	103,6	25	116,7	41	95,5		9	90,1	25	87,3	41	100,2		9	2348,6	25	1815,8	41	2544,7
	8	100,5	24	106,2	40	110,1		8	94,8	24	92,0	40	90,3		8	1947,9	24	2194,3	40	1628,3
	7	95,4	23	119,7	39	104,4		7	95,5	23	83,6	39	91,4		7	3341,8	23	1773,2	39	2020,4
	6	103,0	22	103,7	38	97,7		6	94,8	22	95,5	38	93,4		6	2681,1	22	2668,3	38	2208,0
	5	118,4	21	100,1	37	105,2		5	82,2	21	96,9	37	90,3		5	1474,8	21	2118,4	37	2037,5
	4	118,6	20	116,3	36	99,4		4	82,4	20	87,3	36	94,7		4	1892,5	20	1568,6	36	2297,5
	3	102,6	19	100,2	35	105,2		3	89,3	19	93,8	35	91,7		3	2924,0	19	1798,8	35	2237,8
	2	102,2	18	124,9	34	111,5		2	88,5	18	82,2	34	89,7		2	2237,8	18	1172,2	34	2714,3
	1	102,5	17	115,6	33	138,6		1	91,7	17	93,1	33	83,7		1	2540,4	17	2139,8	33	1996,5

Fonte: O autor.

Figura 23. Mapas de Refletância RGB (a), altura das plantas (b) e produtividade de grãos (c) da Área 2.

Refletância (RGB)								Altura (cm)									
a)	1	97,9	9	95,3	17	99,8	25	101,8	b)	1	76,5	9	76,5	17	76,4	25	68,6
	2	103,7	10	92,3	18	93,6	26	101,0		2	76,5	10	80,5	18	70,2	26	80,6
	3	93,6	11	88,1	19	95,8	27	90,1		3	84,75	11	75,7	19	78,5	27	83,4
	4	93,3	12	91,4	20	97,7	28	87,3		4	82,85	12	84,6	20	72,7	28	86,9
	5	93,8	13	88,8	21	87,4	29	87,8		5	83,75	13	77,2	21	84,1	29	89,7
	6	90,2	14	89,7	22	92,2	30	92,3		6	86,9	14	82,5	22	83,6	30	88,8
	7	87,4	15	91,9	23	101,9	31	89,4		7	87,5	15	87,5	23	79,0	31	89,0
	8	91,8	16	87,8	24	89,9	32	85,9		8	85,85	16	84,0	24	81,0	32	94,6

Produtividade (kg ha-1)								
c)	1	2808,7	9	4200,2	17	2196,0	25	3011,1
	2	2436,8	10	3113,8	18	2966,0	26	3661,6
	3	3183,5	11	3228,5	19	3330,7	27	3619,5
	4	4094,9	12	3837,6	20	2931,8	28	3630,9
	5	3948,8	13	3871,2	21	3608,1	29	4218,3
	6	4430,7	14	4047,8	22	3646,1	30	3082,2
	7	4016,0	15	4175,1	23	2821,8	31	3418,4
	8	4090,0	16	3966,8	24	4058,0	32	3517,1

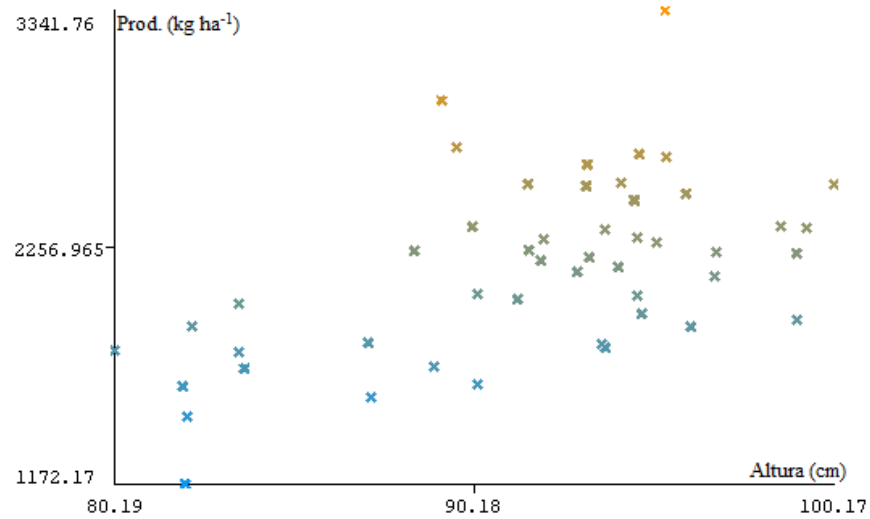
Fonte: O autor.

O valor alto de RGB significa que a planta está refletindo mais a luz (mais amarelada). Isso, somado a um baixo crescimento, demonstrou ser forte indicador de baixo rendimento de grãos. Porém, esse comportamento não se apresentou como uma regra geral. Foi possível observar que, especificamente nas parcelas 34 e 48 da Área 1, e na parcela 9 da Área 2, as plantas refletiram mais e eram menores, porém apresentaram alto rendimento de grãos. A explicação para tais valores conflitantes pode estar associada à possíveis interferências de fatores externos que podem ter prejudicado o rendimento de grãos das parcelas após a realização do sensoriamento da área. Comportamentos inversos aos citados também foram identificados, como nas parcelas 31 e 45 da Área 1, nas quais as plantas refletiram menos (mais esverdeadas) e eram maiores, porém produziram menos. Uma possível explicação agrônômica para isso é o fato dessas duas parcelas terem recebido uma dose mais alta de N em cobertura (120 kg ha^{-1}), o que pode ter favorecido o crescimento vegetativo, sem apresentar reflexos positivos no rendimento de grãos.

Nas Figuras 24 e 25 são apresentadas as dispersões dos dados entre a altura real das plantas medida no campo e a produtividade de grãos de trigo, de acordo com os dados coletados em cada parcela das Áreas 1 e 2. As relações obtidas entre altura das plantas e a

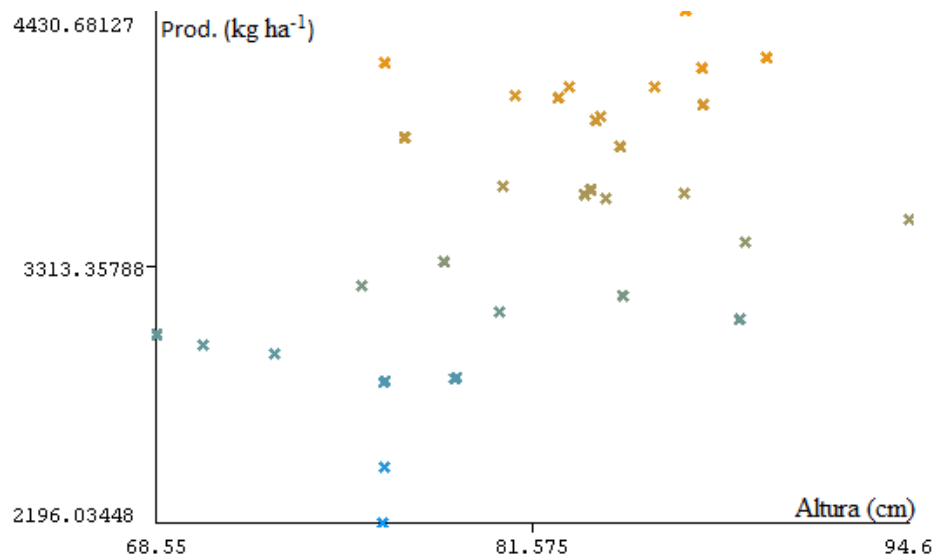
produtividade de grãos podem indicar um comportamento que se assemelha ao da linearidade, ainda que alguns pontos estejam fora desse padrão.

Figura 24. Dispersão dos dados entre altura das plantas e produtividade de grãos de trigo da Área 1.



Fonte: O autor.

Figura 25. Dispersão dos dados entre altura das plantas e produtividade de grãos de trigo da Área 2.



Fonte: O autor.

5.2 TESTES INICIAIS DE REGRESSÃO

Na Tabela 5 são apresentados os resultados dos processamentos realizados utilizando apenas as bandas RGB, desconsiderando a altura das plantas. Considerando os valores de r obtidos, não foram observadas diferenças significativas com a utilização de 100%, 66% ou 33% do volume total de dados.

Tabela 5. Desempenho dos algoritmos de acordo com a proporção dos dados da base A1_RGB_10 (RGB) correlacionados com a produtividade de trigo.

Algoritmo	%	r	R^2
LinearRegression	100	0,55	0,30
	66	0,55	0,30
	33	0,55	0,30
MultilayerPerceptron	100	0,61	0,37
	66	0,67	0,44
	33	0,66	0,43
SMOReg	100	0,79	0,62
	66	0,78	0,60
	33	0,77	0,59

Fonte: O autor.

Na Tabela 6 são mostrados os resultados dos processamentos utilizando as bandas RGB e a altura das plantas (RGB + Altura real). Nesse caso, o melhor resultado foi obtido pelo algoritmo SMOReg, com r de 0,95 e R^2 de 0,90, para as amostras contendo 66% e 100% dos dados. Porém, os resultados mostraram novamente que não houve diferenças significativas com a utilização de diferentes quantidades de amostras para nenhum dos algoritmos testados.

O atributo contendo os valores de altura real das plantas não foi capaz de melhorar o desempenho dos algoritmos LinearRegression e MultilayerPerceptron. Porém, a inserção da altura real das plantas proporcionou um ganho considerável para o algoritmo SMOReg, aumentando a correlação média (r) entre as três amostras de dados de 0,78 (Tabela 5) para 0,95 (Tabela 6). Isso indica que a utilização da altura das plantas pode trazer um considerável ganho aos modelos de estimativa de produtividade de grãos de trigo.

Tabela 6. Desempenho dos algoritmos de acordo a proporção dos dados da base A1_RGB_10 (RGB+Altura real) correlacionados com a produtividade de trigo.

Algoritmo	%	r	R ²
LinearRegression	100	0,56	0,31
	66	0,57	0,32
	33	0,57	0,32
MultilayerPerceptron	100	0,62	0,38
	66	0,68	0,46
	33	0,67	0,44
SMOreg	100	0,95	0,90
	66	0,95	0,90
	33	0,94	0,88

Fonte: O autor.

Os índices de correlação do SMOReg obtidos nesses testes foram superiores aos encontrados no estudo de Gerke (2017), o qual utilizou o mesmo algoritmo também com a finalidade de prever a produtividade de trigo. A comparação entre esses trabalhos é relevante, uma vez que ambos foram realizados na mesma área experimental e com o mesmo objetivo, porém em safras diferentes. No estudo de Gerke (2017), o maior valor de r obtido foi 0,61. Essa diferença na correlação obtida pode ser explicada pelo fato de o autor ter utilizado apenas os valores das bandas espectrais e suas combinações, enquanto que no atual estudo foram inseridos também os valores de altura das plantas.

Outro ponto observado por Gerke (2017) foi que as diferentes quantidades de dados utilizados por base não afetaram a eficiência dos modelos de predição, algo que se confirmou no presente estudo. Isso mostra que quando se trabalha com dados espectrais é recomendável utilizar uma menor quantidade de amostras de dados em relação à sua totalidade, uma vez que isso não prejudica o desempenho dos algoritmos. Em relação ao custo computacional, o tempo de execução do algoritmo SMOReg na base com 33% dos dados foi 30 vezes menor que o da base com 66%, e 5.500 vezes menor que o da base com 100%. Portanto, a utilização da base com menor quantidade de dados proporcionou uma redução considerável no custo computacional.

Considerando os resultados obtidos com os testes iniciais, o algoritmo SMOReg apresentou o melhor desempenho e foi determinado como a técnica de aprendizado de máquina

utilizada para a realização dos testes seguintes do trabalho. Os altos valores de correlação obtidos por esse algoritmo podem ser justificados pela maior robustez do SMOReg e a sua capacidade de resolver problemas mais complexos que podem apresentar comportamentos não lineares. Também foram consideradas para as etapas seguintes somente as bases contendo 33% do total dos dados.

5.3 ESTIMATIVAS DE ALTURA

Os resultados de estimativa da altura das plantas de trigo da Área 1, de acordo com as resoluções das câmeras RGB (Tabela 7) e NIR (Tabela 8), mostraram que a câmera NIR apresentou desempenho significativamente superior para estimar a altura das plantas em relação à câmera RGB.

Tabela 7. Comparação entre as resoluções para as estimativas de altura das plantas de trigo na Área 1 utilizando dados da câmera RGB.

Métrica	Altura Estimada (RGB)		
	Resolução		
	3,4	10	26
Altura média por parcela (real = 91.8 cm)	25,2	21,1	10,9
Parcela mais baixa (real = 80.1 cm)	1,9	3,3	1,1
Parcela mais alta (real = 100.1 cm)	67,0	41,2	30,0
Parcelas com alturas subestimadas	48	48	48
Parcelas com alturas superestimadas	0	0	0
Erro médio por parcela (cm)	66,5	70,7	80,9
Erro médio por parcela (%)	72,7	77,0	88,2
Correlação com altura real (r)	0,50	0,50	0,49
Correlação com a produtividade (r)	0,55	0,60	0,49

Fonte: O autor.

Tabela 8. Comparação entre as resoluções para as estimativas de altura das plantas de trigo na Área 1 utilizando dados da câmera NIR.

Métrica	Altura Estimada (NIR)		
	Resolução		
	3,4	10	26
Altura média por parcela (real = 91.8 cm)	34,2	78,6	100,7
Parcela mais baixa (real = 80.1 cm)	4	25	15,8
Parcela mais alta (real = 100.1 cm)	74,6	116,8	180,0
Parcelas com alturas subestimadas	48	33	23
Parcelas com alturas superestimadas	0	15	25
Erro médio por parcela (cm)	57,3	21,1	31,2
Erro médio por parcela (%)	62,1	23,0	35,0
Correlação com altura real (r)	0,51	0,67	0,66
Correlação com a produtividade (r)	0,44	0,62	0,70

Fonte: O autor.

O erro médio por parcela da câmera RGB na resolução de 10 cm/pixel foi de 70,7 cm (Tabela 7), enquanto que o da NIR foi de 21,1 cm (Tabela 8), mais de três vezes menor. Na resolução de 26 cm/pixel, o desempenho da câmera NIR também foi superior ao da RGB, apresentando erro médio de 31,2 cm, quase 50 cm menor.

Considerando a mesma câmera com diferentes resoluções, a câmera NIR apresentou erro médio significativamente inferior na resolução de 26 cm/pixel em relação à resolução de 3,4 cm/pixel, e a correlação da câmera NIR na resolução de 26 cm/pixel com a produtividade de grãos foi de 0,70, valor estatisticamente superior ao obtido pela resolução de 3,4 cm/pixel (0,44).

A Tabela 9 apresenta as estimativas obtidas para os dados da Área 2.

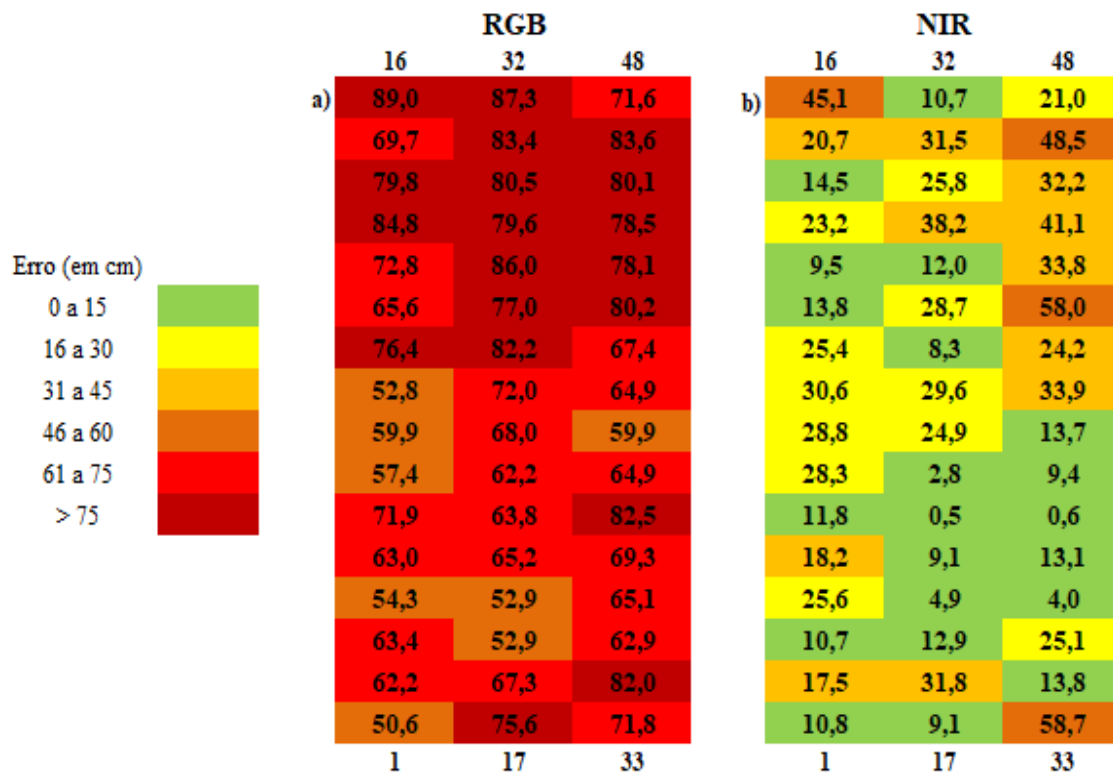
Tabela 9. Comparação entre as resoluções para as estimativas de altura das plantas de trigo na Área 2 utilizando dados da câmera RGB.

Métrica	Altura Estimada (RGB)	
	3,4	10
Altura média por parcela (real = 81.8 cm)	33.4	58,7
Parcela mais baixa (real = 68.5 cm)	4.2	6,9
Parcela mais alta (real = 94.6 cm)	88.7	167,2
Parcelas com alturas subestimadas	32	24
Parcelas com alturas superestimadas	0	8
Erro médio por parcela (cm)	48.2	44,1
Erro médio por parcela (%)	60.4	54,4
Correlação com altura real (r)	0.53	0,65
Correlação com a produtividade (r)	0.57	0,68

Fonte: O autor.

A câmera RGB não apresentou resultados satisfatórios para estimar a altura das plantas em nenhuma das resoluções utilizadas. Ainda que os testes realizados com ela na Área 2 (Tabela 9) tenham apresentado resultados significativamente melhores em relação aos da Área 1 (Tabela 7), a precisão continuou baixa, levando em consideração principalmente o erro médio por parcela (em cm). Os dados da RGB apresentaram uma tendência a subestimar a altura real das plantas, sendo que a altura estimada foi menor que a real em 200 das 208 das parcelas contabilizadas das Áreas 1 e 2, e nas três resoluções diferentes, representando 96,1% dos casos. O mapa de erros apresentado na Figura 26 mostra exatamente em quais parcelas o erro de cada câmera foi maior, utilizando dados com resolução de 10 cm/pixel da Área 1.

Figura 26. Mapa de erro de estimativa de altura das plantas de trigo da Área 1 para cada parcela de acordo com as câmeras RGB (a) e NIR (b) utilizando resolução de 10 cm/pixel.



Fonte: O autor.

Foi possível identificar que as parcelas que apresentaram os maiores erros de estimativas se concentraram nas áreas mais superiores do mapa (Figura 26). Uma possível explicação para isso é que existe uma declividade de 4 m de uma ponta à outra da área, e na parte mais baixa desse declive é justamente onde se concentram os maiores erros. Esses dados apontam que as câmeras tendem a subestimar ainda mais a altura dos objetos em regiões mais baixas. O erro de estimativa da câmera RGB nessa área foi aumentando cerca de 8,6 cm, em média, para cada metro de declive.

Os resultados obtidos demonstraram que a utilização da câmera NIR se apresentou como uma melhor opção em relação à câmera RGB para trabalhos que envolvem a estimativa da altura das plantas do trigo. Uma explicação para esses resultados pode ser o fato de a NIR ter como característica o maior realce do solo em relação à vegetação, o que pode ter contribuído para uma melhor geração dos modelos 3D no momento da separação entre solo e objetos.

As resoluções de 10 e 26 cm/pixel utilizadas pela NIR apresentaram maior potencial em relação à resolução 3,4 cm/pixel, demonstrando que resoluções mais altas podem não contribuir para uma melhor estimativa de altura das plantas de trigo.

5.4 PREDIÇÃO DE PRODUTIVIDADE

Na Tabela 10 são mostrados os resultados obtidos utilizando os dados das câmeras RGB e NIR combinados com as alturas estimadas das plantas para prever a produtividade de grãos de trigo nas bases da Área 1. Os resultados dos testes com os dados da câmera RGB apresentaram maiores índices de correlação com a produtividade de trigo em relação à câmera NIR. Por outro lado, considerando as três resoluções, os testes estatísticos não apontaram diferenças significativas entre os resultados das duas câmeras.

Tabela 10 - Resultados dos testes de predição de produtividade de grão de trigo de acordo com as resoluções das imagens da Área 1.

Resolução	Atributos	r	R²
3,4	RGB	0,78	0,60
	RGB + Altura estimada	0,94	0,88
	NIR	0,71	0,50
	NIR + Altura estimada	0,88	0,77
10	RGB	0,77	0,59
	RGB + Altura estimada	0,97	0,94
	NIR	0,73	0,53
	NIR + Altura estimada	0,91	0,82
26	RGB	0,74	0,55
	RGB + Altura estimada	0,97	0,94
	NIR	0,68	0,46
	NIR + Altura estimada	0,84	0,70

Fonte: O autor.

Na Tabela 11 estão apresentados os resultados dos testes realizados com os dados das Áreas 2 e 3. Os resultados obtidos na Área 3 foram significativamente inferiores em relação às outras áreas, sendo que os máximos valores de r e de R² obtidos na Área 3 foram de 0,77 e

0,59, respectivamente (RGB + Altura Estimada). Entre as explicações para esse desempenho inferior pode ser o fato de a Área 3 ter apresentado menor variação entre os tratamentos. Também é possível que a câmera acoplada à RPA Phantom não tenha sido capaz que capturar o detalhamento necessário para prever a produtividade de grãos.

Tabela 11 – Resultados dos testes de predição de produtividade de grãos de trigo de acordo com as resoluções das imagens das Áreas 2 e 3.

Área	Resolução	Atributos	r	R ²
2	3,4	RGB	0,74	0,54
		RGB + Altura estimada	0,90	0,81
2	10	RGB	0,77	0,59
		RGB + Altura estimada	0,92	0,85
3	5	RGB	0,63	0,39
		RGB + Altura estimada	0,77	0,59

Fonte: O autor.

Avaliando os dados das três áreas, a resolução de 26 cm/pixel apresentou maior potencial para prever a produtividade de grãos de trigo, por ter apresentado valor de R² de até 0,94 (RGB + Altura Estimada) e por ter gerado bases de dados menores relação às outras três resoluções (sessão 4.5.1), proporcionando menor custo computacional no processamento.

Considerando todos os resultados dessa etapa, foi verificado que a câmera RGB apresentou maiores valores de R² em relação aos dados da câmera NIR, indicando que a utilização dessa câmera se mostrou mais adequada quando se tem o objetivo de prever a produtividade de trigo. Os modelos gerados pelas bases A1_RGB_10, A1_RGB_26 e A2_RGB_10, utilizando os atributos RGB+Altura Estimada, apresentaram os melhores resultados entre todos os testes, com valores de R² de 0,94 para os modelos da Área 1 e de 0,85 para o modelo da Área 2. Esses resultados mostraram que foi possível a geração de modelos de predição com altos índices de correlação com produtividade de trigo.

Os valores de R² obtidos nesse estudo foram iguais ou superiores aos encontrados em outros trabalhos que realizaram estimativas da produtividade de grãos de trigo, como os de Bose et al. (2016) e de Aboelghar et al. (2014). Bose et al. (2016) obtiveram valores de r de até 0,80 (R² = 0,64) ao correlacionar a produtividade do trigo com dados de imagens espaço temporais. Aboelghar et al. (2012) conseguiram um valor de r de 0,93 (R²=0,86) por meio de técnicas de regressão linear com o objetivo de gerar modelos de previsão de produtividade de

trigo utilizando dados espectrais. É possível que a utilização da altura das plantas como atributo descritivo dos dados tenha sido determinante para o ganho de desempenho do atual estudo em relação aos demais trabalhos realizados.

6 CONCLUSÕES

Com base nos resultados obtidos nessa pesquisa, pode-se concluir que a altura das plantas, seja por meio de medida coletada no campo ou de medida estimada remotamente, contribui significativamente para a geração de modelos mais precisos para a predição da produtividade de trigo.

Considerando os testes nas diferentes proporções dos dados, foi verificado que ao se trabalhar com dados espectrais não é necessário utilizar a quantidade total de amostras para se obter ganho de desempenho, tendo em vista que apenas 1/3 dos dados demonstrou ser suficiente nesse estudo. Com isso foi possível obter uma considerável economia de custo computacional.

A câmera NIR apresentou maior precisão em relação à câmera RGB para estimar a altura das plantas de trigo, obtendo menor erro médio por parcela em todas as resoluções testadas (3,4, 10 e 26 cm/pixel). A resolução de 10 cm/pixel se apresentou como uma boa opção para se realizar essa estimativa.

Em relação à produtividade de grãos de trigo, os modelos de predição gerados, com valores de r de até 0,97, demonstraram que é possível se obter alta correlação com rendimento de trigo utilizando apenas valores digitais de bandas espectrais combinados com a altura das plantas também estimada remotamente. A câmera RGB explicou melhor a relação entre esses dados e a variação da produtividade de grãos, se mostrando como a melhor opção para esse objetivo. As resoluções de 10 e 26 cm/pixel apresentaram desempenho semelhante à de 3,4 cm/pixel, revelando que um maior detalhamento da imagem não necessariamente implica em maior correlação com produtividade de trigo.

O RPA eBee demonstrou eficiência no trabalho de captura das imagens por sensoriamento remoto, sendo de fácil manuseio e capaz de cobrir grandes áreas rapidamente. O Software Pix4D, além de gerar os mapas de refletância das áreas de estudo, foi também uma excelente ferramenta para a criação dos modelos digitais, assim como para a correção georreferencial das imagens. O algoritmo SMOReg se mostrou como uma técnica promissora de aprendizado de máquina para prever a produtividade de grãos de trigo, e a utilização do pacote wekaServer, disponível na ferramenta Weka, proporcionou considerável ganho computacional nas execuções de seus algoritmos.

Considerando possíveis trabalhos futuros, seria recomendável a execução de uma maior quantidade de voos, preferencialmente em períodos mais curtos após a semeadura, a fim de verificar a possibilidade de mapear o crescimento das plantas durante todo o ciclo de

desenvolvimento da cultura do trigo. Considerando que a altura das plantas do trigo foi subestimada na resolução de 10 cm/pixel e superestimada na resolução de 26 cm/pixel, poderia ser interessante em trabalhos que visem essa estimativa a utilização de uma resolução intermediária, como por exemplo de 15 cm/pixel. Dentre as configurações do software Pix4D, é possível que a opção de classificação da nuvem de pontos (*point cloud classification*), não disponível na versão utilizada para esse estudo, proporcione a geração de modelos DSM mais detalhados e, consequentemente, estimativas de altura de plantas mais precisas

REFERÊNCIAS

- ABOELGHAR, M.; ALI, A. R.; ARAFAT, S. Spectral wheat yield prediction modeling using SPOT satellite imagery and leaf area index. **Arabian Journal of Geosciences**, v. 7, p. 465-474, 2014.
- BATTUDE, M.; AL BITAR, A.; MORIN, D.; CROS, J.; HUC, M., SICRE, C. M.; LE DANTEC, V.; DEMAREZ, V. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. **Remote Sensing of Environment**, v. 184, p. 668-681, 2016.
- BERNARDI, A. C. D. C.; NAIME, J. D. M.; DE RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. **Agricultura de Precisão: resultados de um novo olhar**. Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. Brasília: Embrapa; Cubo, 2014.
- BOSE, P.; KASABOV, N. K.; BRUZZONE, L.; HARTONO, R. N. Spiking neural networks for crop yield estimation based on spatio temporal analysis of image time series. **IEEE Transactions on Geoscience and Remote Sensing**. v. 54, p. 6563-6573, 2016.
- BRAGA, L. P. V. B. **Introdução à Mineração de Dados** - 2a edição: Edição ampliada e revisada. Rio de Janeiro: E-papers, 2005.
- CHANG, J.; CLAY, D. E.; CLAY, S. A.; REESE, C. L. Using field scouting or remote sensing technique to assess soybean yield limiting factors. Em: CLAY, D.E.; CARLSON, C.G.; CLAY, S.A.; WAGNER, L.; DENEKE, D.; Hay, C. **iGrow Soybean: Best Management Practices for Soybean Production**. Dakota do Sul: Igrow – SDSU Extension, 2013.
- CONAB, **Acompanhamento da Safra Brasileira**, V. 4 Safra 2016/2017 Décimo Segundo Levantamento. Companhia Nacional de Abastecimento - CONAB. 2017.
- DJI. **Phantom 4**. Especificações. Disponível em: <https://www.dji.com/br/phantom-4/>. Acesso em: 07 nov. 2019.
- ERASMI, S.; ROSENBAUER, R.; BUCHBACH, R.; BUSCHE, T.; RUTISHAUSER, S. Evaluating the quality and accuracy of TanDEM-X digital elevation models at archaeological sites in the Cilician Plain, Turkey. **Remote Sensing**, v. 6, p. 9475-9493, 2014.
- FIGUEIREDO, D. **Conceitos básicos de sensoriamento remoto**. 2005. Disponível em: <https://www.academia.edu/download/50009361/conceitos_sm.pdf>. Acesso em: 19 mar. 2019.
- FRIEDLI, M.; KIRCHGESSNER, N.; GRIEDER, C.; LIEBISCH, F.; MANNALE, M.; WALTER, A. Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. **Plant Methods**, v. 12, p. 9, 2016.
- GERKE, T. **Mineração de dados de imagens obtidas com aeronave remotamente pilotada para estimativa de produtividade do trigo**. Dissertação (Mestrado em Computação Aplicada). Universidade Estadual de Ponta Grossa, 2017.

HAN, J.; PEI, J.; KAMBER, M. **Data mining**: concepts and techniques. Amsterdã: Elsevier, 2011.

HOLMAN, F. H.; RICHE, A. B.; MICHALSKI, A.; CASTLE, M.; WOOSTER, M. J.; HAWKESFORD, M. J. High Throughput Field Phenotyping of Wheat Plant Height and Growth Rate in Field Plot Trials Using UAV Based Remote Sensing. **Remote Sensing**, v. 8, p. 1031, 2016.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Serviço online para pós-processamento de dados GNSS - IBGE-PPP**. Disponível em: <https://www.ibge.gov.br/geociencias/informacoes-sobre-posicionamento-geodesico/servicos-para-posicionamento-geodesico/16334-servico-online-para-pos-processamento-de-dados-gnss-ibge-ppp.html>. Acesso em: 12 set. 2018.

KOENIG, K.; HÖFLE, B.; HÄMMERLE, M.; JARMER, T.; SIEGMANN, B.; LILIENTHAL, H. Comparative classification analysis of post-harvest growth detection from terrestrial LiDAR point clouds in precision agriculture. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 104, p. 112-125, 2015.

KRAMER, L. F. M.; MULLER, M. M. L.; TORMENA, C. A.; GENÚ, A. M.; MICHALOVICZ, L.; VICENSI, M.; Atributos químicos do solo associados à produtividade do trigo em um talhão com diferentes potenciais produtivos. **Revista Brasileira de Ciência e Solo**, v. 38, p. 1190-1199, 2014.

LARGE, E. C. Growth stages in cereals illustration of the Feekes scale. **Plant pathology**, v. 3, p. 128-129, 1954.

LATI, R.N.; FILIN, S.; EIZENBERG, H. Estimating plant growth parameters using an energy minimization-based stereovision model. **Computers and Electronics in Agriculture**, v. 98, p. 260-271, 2013.

MAKANZA, R.; ZAMAN-ALLAH, M.; CAIRNS, J. E.; MAGOROKOSHO, C., TAREKEGNE, A.; OLSEN, M.; PRASANNA, B. M. High-throughput phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. **Remote Sensing**, v. 10, p. 330, 2018.

MARTHA, T. R.; KERLE, N.; JETTEN, V.; VAN WESTEN, C. J.; & KUMAR, K. V. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. **Geomorphology**, v. 116, p. 24-36, 2010.

MENESES, P. R., ALMEIDA, T. **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: UNB, 2012. Disponível em: <http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8>. Acesso em: 20 mar. 2019.

MOECKEL, T.; DAYANANDA, S.; NIDAMANURI, R. R.; NAUTIYAL, S., HANUMAIAH, N.; BUERKERT, A.; WACHENDORF, M. Estimation of vegetable crop parameter by multi-temporal UAV-borne images. **Remote Sensing**, v. 10, p. 805, 2018.

- PANDA, S. S.; AMES, D. P.; PANIGRAHI, S. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. **Remote Sensing**, v. 2, p. 673-696, 2010.
- PAST, **Past 4** - the Past of the Future. Disponível em: <http://folk.uio.no/ohammer/past>. Acesso em: 20 fev. 2020.
- PIRES, J. L. F.; DA CUNHA, G. R.; DALMAGO, G. A.; PASINATO, A.; SANTI, A., PEREIRA, P. D. S.; DOS SANTOS, H. P.; SANTI, A. L. Integração de práticas de manejo no sistema de produção de trigo. **Trigo no Brasil**. Brasília: Embrapa, 2011.
- PIX4D, **manual**. 2017. Disponível em: <https://support.pix4d.com/hc/en-us/sections/200591059-Manual>. Acesso em: 10 jan. 2019.
- QGIS, **userguide**. Disponível em: https://docs.qgis.org/2.18/en/docs/user_manual. Acesso em: 12 jan. 2019.
- RAY, A. S. Remote Sensing in Agriculture. **International Journal of Environment, Agriculture and Biotechnology**, v. 1, p. 362-367, 2016.
- RAYBURG, S.; THOMS, M.; NEAVE, M. A comparison of digital elevation models generated from different data sources. **Geomorphology**, v. 106, p. 261-270, 2009.
- RICHETTI, J.; JUDGE, J.; BOOTE, K. J.; JOHANN, J. A.; URIBE-OPAZO; M. A., BECKER; W. R; PALUDO; A.; SILVA, L. C. de A. Using phenology-based enhanced vegetation index and machine learning for soybean yield estimation in Paraná State, Brazil. **Journal of Applied Remote Sensing**, v. 12, p. 026029, 2018.
- RODRIGUES, F. A., JR.; ORTIZ-MONASTERIO, I.; ZARCO-TEJADA, P. J.; SCHULTHESS, U.; GÉRARD, B. High resolution remote and proximal sensing to assess low and high yield areas in a wheat field. **10th European Conference on Precision Agriculture, ECPA 2015**; Wageningen Academic Publishers: Wageningen, Netherlands, 2015.
- RODRIGUES, S. C. A. **Modelo de Regressão Linear e suas Aplicações**. Relatório de Estágio (Mestrado em Ensino de Matemática). Universidade da Beira Interior. 2012.
- RUSSEL, S.; NORVIG, P.; **Inteligência Artificial**. 3ª ed. Rio de Janeiro: Elsevier, 2013. 988 p.
- SANTOS, C. M. A.; **Estatística Descritiva** – Manual de auto-aprendizagem. Lisboa: Sílabo; 2007.
- SCHIRRMANN, M.; GIEBEL, A.; GLEINIGER, F.; PFLANZ, M.; LENTSCHKE, J., DAMMER, K. H. Monitoring agronomic parameters of winter wheat crops with low-cost UAV imagery. **Remote Sensing**, v. 8, p. 706, 2016.
- SENSEFLY. senseFly, **eBee Classic**, The Professional Mapping Drone. Disponível em: <https://www.sensefly.com/drone/ebec-mapping-drone>. Acesso em: 22 fev. 2019.

SHEWRY, P. R.; HEY, S. J. The contribution of wheat to human diet and health. **Food and energy security**, v. 4, p. 178-202, 2015.

SILVESTRO, P. C.; PIGNATTI, S.; PASCUCCI, S.; YANG, H.; LI, Z.; YANG, G.; HUANG, W.; CASA, R. 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, p. 509, 2017.

TILLY, N.; AASEN, H.; BARETH, G. Fusion of plant height and vegetation indices for the estimation of barley biomass. **Remote Sensing**, v. 7, p. 11449-11480, 2015.

WESTOBY, M. J.; BRASINGTON, J.; GLASSER, N. F.; HAMBREY, M. J.; REYNOLDS, J. M. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. **Geomorphology**, v. 179, p. 300-314, 2012.

XUE, J.; SU, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. **Journal of Sensors**, v. 2017, 2017.

WEKA, **Data mining software in java**. The University of Waikato. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka>. Acesso em: 20 fev. 2019

YU, N.; LI, L.; SCHMITZ, N.; TIAN, L. F.; GREENBERG, J. A.; DIERS, B. W. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. **Remote Sensing of Environment**, v. 187, p. 91-101, 2016.

ZBICKI, R.; DOS SANTOS, T. M. **Machine Learning sob a ótica estatística**. 2019. Disponível em: <http://www.rizbicki.ufscar.br/sml.pdf>. Acesso em: 04 fev. 2019.