

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
MESTRADO EM COMPUTAÇÃO APLICADA

TIAGO GERKE

**MINERAÇÃO DE DADOS DE IMAGENS OBTIDAS COM AERONAVE
REMOTAMENTE PILOTADA PARA ESTIMATIVA DE PRODUTIVIDADE DO
TRIGO**

PONTA GROSSA
2017

TIAGO GERKE

**MINERAÇÃO DE DADOS DE IMAGENS OBTIDAS COM AERONAVE
REMOTAMENTE PILOTADA PARA ESTIMATIVA DE PRODUTIVIDADE DO
TRIGO**

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.
Orientador(a): Prof^a Dr^a Alaine Margarete Guimarães
Co-orientador: Prof. Dr. Eduardo Fávero Caires

PONTA GROSSA

2017

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

Gerke, Tiago
G354 Mineração de dados de imagens obtidas
com aeronave remotamente pilotada para
estimativa de produtividade do trigo/
Tiago Gerke. Ponta Grossa, 2017.
81f.

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

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

Coorientador: Prof. Dr. Eduardo Fávero
Caires.

1.Aeronave remotamente pilotada.
2.Drone. 3.Veículo aéreo não tripulado.
4.VANT. 5.Sensoriamento remoto.
I.Guimarães, Elaine Margarete. II.
Caires, Eduardo Fávero. III. Universidade
Estadual de Ponta Grossa. Mestrado em
Computação Aplicada. IV. T.

CDD: 006.3

TERMO DE APROVAÇÃO

Tiago Gerke

“MINERAÇÃO DE DADOS DE IMAGENS OBTIDAS COM AERONAVE REMOTAMENTE PILOTADA PARA ESTIMATIVA DE PRODUTIVIDADE DO TRIGO”.

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:

Orientadora:


Dr.ª Alaine Margarete Guimarães
UEPG


Dr. Helio Antonio Wood Joris
UEPG


Dr.ª Deborah Ribeiro Carvalho
PUC/PR

Ponta Grossa, 03 de fevereiro de 2017.

“A ciência sem fé é loucura, e a fé sem ciência é fanatismo.”

Martinho Lutero

AGRADECIMENTOS

A Deus, autor e consumidor da minha fé, por me capacitar e dar forças a cada dia. A Ele toda Honra e toda a Glória!

A minha esposa Isabel, por me apoiar e incentivar incondicionalmente, e por não me deixar desistir em momentos em que essa parecia ser a única opção. Essa conquista é nossa!

Aos meus pais Horst e Dorlis, pois somente através dos sacrifícios que fizeram, fé e ensinamentos que me transmitiram e pelo constante incentivo foi possível chegar a tantas conquistas, assim como esta. Amo vocês.

Aos meus sogros Horst e Cristiane, que pelas constantes orações, apoio e incentivo, foram igualmente fundamentais para essa conquista. Vocês são muito importantes em minha vida.

A minha orientadora Prof^a. Dr^a Elaine M. Guimarães, por toda ajuda durante o mestrado e pelos ensinamentos que tornaram essa pesquisa possível. Também ao meu co-orientador Prof. Dr. Eduardo F. Caires, por disponibilizar os dados das áreas experimentais utilizadas nesse trabalho.

Ao Programa de Mestrado em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela oportunidade de aprender, realizar a pesquisa e aprimorar minha formação.

Ao meu amigo e colega Victor S. Lacerda, pela parceria durante todo o mestrado, e também aos seus pais Pedro e Suzane, por terem me acolhido em sua casa, à sua mesa, principalmente no período em que eu ainda não residia em Ponta Grossa.

Ao meu amigo e colega Roger Mazur e sua esposa Ingrid, irmãos na fé, pelas diversas vezes em que também me acolheram em sua casa, que Deus os abençoe.

Ao Instituto Federal do Paraná, por ter me permitido participar desse programa de mestrado.

RESUMO

O cultivo do trigo desempenha um papel importante no desenvolvimento econômico de várias regiões do Brasil e do mundo, bem como na dieta humana. A produção brasileira do trigo é insuficiente para atender à demanda nacional, tornando necessárias pesquisas com intuito de melhorar a produtividade desse cereal. O objetivo desse trabalho foi a estimativa de produtividade do trigo, a partir da criação de modelos preditivos por meio da mineração de dados obtidos em imagens de alta resolução espacial, coletadas por aeronave remotamente pilotada (RPA). O trabalho foi realizado em duas áreas experimentais na cidade de Ponta Grossa – PR, onde para cada área foram feitas oito coletas de imagens, em diferentes estádios de desenvolvimento da cultura, com resolução espacial de 3,4cm/px e duas coletas com resolução 10cm/px e 20cm/px, através de uma RPA eBee utilizando uma câmera RGB e outra NIR. O processamento das imagens foi feito a partir do software Pix4D, e resultou em um ortomosaico com os valores de refletância em diferentes comprimentos de onda: R, G e B da câmera RGB e R, G e NIR da câmera NIR, além de uma imagem com valores de NDVI obtidos a partir da aritmética das bandas Nir e Red (vermelho). A correção de georreferenciamento de cada ortomosaico e a extração dos valores de refletância foram feitas com auxílio do sistema de informação geográfica (SIG) Quantum GIS. A partir dos valores de refletância extraídos, foram criadas bases de dados em diferentes proporções (10%, 20%, 40%, 70% e 100%) para mineração de dados por meio do algoritmo SMOReg, baseado em máquina de vetor de suporte (SVM) para regressão (SVR). A correção de georreferenciamento utilizando 10 pontos de controle proporcionou ortomosaicos com erro médio quadrático (RSME) de distância de 0,35m, o qual não mostrou diferença significativa para a correção com 5 pontos de controle (RMSE = 0,38m). Os valores de refletância foram diferentes para cada área de estudo, tornando difícil a indicação de melhores períodos para a estimativa de produtividade do trigo. Os maiores índices de correlação da produtividade com os comprimentos de onda, foram obtidos com os dados das imagens da câmera RGB, seguido da câmera NIR e NDVI, com as correlações de 0,6168, 0,5423 e 0,5324, respectivamente. A quantidade de informação extraída das imagens, refletida na proporção das bases de dados, não se mostrou significativa nos modelos preditivos gerados, bem como nos índices de correlação, os quais foram estatisticamente iguais. Índices de correlação melhores foram obtidos a partir dos dados extraídos das imagens com resolução espacial de 20cm/px, o que sugere que imagens de alta resolução espacial podem não ser adequadas para estimativa de produtividade do trigo.

Palavras-chave: Aeronave remotamente pilotada, drone, veículo aéreo não tripulado, VANT, sensoriamento remoto, processamento digital de imagens, mineração de dados.

ABSTRACT / KEYWORDS

Wheat cultivation plays an important role to Brazil and the world economic development, as well as in the human diet. The wheat Brazilian production is insufficient to meet the national demand, making research needed in order to improve the yield of this cereal. The goal of this work was to estimate wheat yield, searching for a predictive model through the data mining techniques, with data obtained from high spatial resolution images collected by unmanned aerial vehicles (UAV). The work was carried out in two experimental areas at Ponta Grossa city, Parana state, where for each area eight images were taken, at different culture development stages, with spatial resolution of 3.4cm/px and two images with resolution 10cm/px and 20cm/px, using an eBee UAV with an RGB and a NIR camera. The image processing was done with the Pix4D software, and resulted in an orthomosaics with reflectance values at different wavelengths: Red, Green and Blue, from the RGB camera and Red, Green and NIR from the NIR camera, besides an image with NDVI values obtained from the arithmetic of NIR and Red wavelengths. The georeferencing correction of each orthomosaic and the extraction of the reflectance values were done with Quantum GIS geographic information system (GIS). From the extracted reflectance values, databases in different proportions (10%, 20%, 40%, 70% and 100%) were created for data mining, using the SMOReg algorithm, based on a support vector machine (SVM) for regression (SVR). The georeferencing correction using 10 control points provided orthomosaics with mean square error (RSME) of distance of 0.35m, which did not show significant difference compared to the correction with 5 control points (RMSE = 0.38m). The reflectance values were different for each study area, making it difficult to indicate better periods for estimating wheat yield. The highest correlation were obtained with data from RGB camera images, followed by the NIR and NDVI camera, with correlations of 0.6168, 0.5423 and 0.5324, respectively. The amount of information extracted from the images, reflected in the proportion of the databases, was not significant to generated predictive models, as well as in the correlation indexes, which were statistically the same. Better correlation indices were obtained from the data extracted from the images with spatial resolution of 20cm/px, which suggests that high spatial resolution images may not be adequate for wheat yield estimation.

Keywords: Remotely piloted aircraft, drone, unmanned aerial vehicle, UAV, remote sensing, digital image processing, data mining.

LISTA DE FIGURAS

Figura 1. Padrão de reflexão de diferentes materiais no decorrer do espectro.....	20
Figura 2. Imagens resultantes a partir do cálculo de índices de vegetação. (a) Imagem original; (b) RVI e (c) NDVI.....	23
Figura 3. Demonstração da separação dos dados que a SVM busca.....	30
Figura 4. Possibilidades de separação dos dados. (a) margem pequena e (b) margem grande	31
Figura 5. Fluxograma com etapas realizadas no decorrer do projeto	33
Figura 6. Área experimental da Fazenda Santa Cruz	34
Figura 7. Área experimental da Fazenda Escola UEPG	35
Figura 8. Produtividade por parcelas da Fazenda Santa Cruz.....	36
Figura 9. Produtividade por parcelas da Fazenda Escola.....	36
Figura 10. Produtividade das três repetições de cada um dos 16 tratamentos da Fazenda Santa Cruz	37
Figura 11. Produtividade das quatro repetições de cada um dos 10 tratamentos da Fazenda Escola.....	37
Figura 12. RPA eBee utilizada nas coletas de dados.....	38
Figura 13. (a) Câmera RGB e (b) Câmera NIR utilizadas pela RPA eBee.....	38
Figura 14. Comprimentos de onda filtrados pelas câmeras RGB e NIR	39
Figura 15. Tela do software Emotion 2 para controle da RPA eBee	40
Figura 16. Interface do software Pix4D	42
Figura 17. Nuvem de pontos gerada pelo Pix4D para geração do ortomosaico	43
Figura 18. Diferença entre o ortomosaico (a) e o mapa de refletância (b)	44
Figura 19. Marcadores utilizados como GCPs para a correção de georreferenciamento	47
Figura 20. (a) GPS Trimble R4 e (b) a disposição dos GCPs na área experimental.....	47
Figura 21. Pontos brancos para demarcação da divisão de parcelas da Fazenda Santa Cruz (a) e da Fazenda Escola (b)	49
Figura 22. Linhas de separação de parcelas da Fazenda Santa Cruz.....	49
Figura 23. Linhas de separação de parcelas da Fazenda Escola.....	50
Figura 24. Polígonos representando cada parcela da Fazenda Santa Cruz	50
Figura 25. Polígonos representando cada parcela da Fazenda Escola	51
Figura 26. Visualização dos pontos de uma parcela em uma imagem de 3,4cm/px.....	52

Figura 27. Visualização dos pontos de uma parcela em uma imagem de 10cm/px..	53
Figura 28. Visualização dos pontos de uma parcela em uma imagem de 20cm/px..	53
Figura 29. Demonstração do erro de georreferenciamento (tratamento Pix0)	55
Figura 30. Comportamento dos valores de NDVI ao longo das coletas. (a) Fazenda Santa Cruz e (b) Fazenda Escola	58
Figura 31. Comportamento dos valores digitais de R, G e B ao longo das coletas, obtidos com a câmera RGB para Fazenda Santa Cruz	58
Figura 32. Comportamento dos valores digitais de R, G e B ao longo das coletas, obtidos com a câmera RGB para Fazenda Escola.....	59
Figura 33. Comparativo de comportamento dos valores no canal R (Red) obtidos com a câmera (a) NIR e com a câmera (b) RGB, para Fazenda Santa Cruz.....	59
Figura 34. Comparativo de comportamento dos valores no canal G (Green) obtidos com a câmera NIR (a) e com a câmera RGB (b), para Fazenda Santa Cruz.....	60
Figura 35. Comparativo de comportamento dos valores no canal R (Red) obtidos com a câmera NIR (a) e com a câmera RGB(b), para Fazenda Escola.....	60
Figura 36. Comparativo de comportamento dos valores no canal G (Green) obtidos com a câmera NIR (a) e com a câmera RGB (b), para Fazenda Escola.....	60
Figura 37. Comportamento dos valores de NIR ao longo das coletas. (a) Fazenda Santa Cruz e (b) Fazenda Escola	61
Figura 38. Índices de correlação obtidos com bases na proporção de 10% para Fazenda Santa Cruz	65
Figura 39. Valores médios de refletância para o canal RGB_R nos diferentes períodos de coleta na Fazenda Santa Cruz, considerando as diferentes proporções das bases de dados	68
Figura 40. Valores médios de refletância para o canal RGB_R nos diferentes períodos de coleta na Fazenda Escola, considerando as diferentes proporções das bases de dados	68
Figura 41. Diferença nas características de uma imagem nas resoluções de 3,4cm/px (a), 10cm/px (b) e 20cm/px (c)	72

LISTA DE TABELAS

Tabela 1. Classificação de categorias de RPAs a partir da massa, alcance, altitude e autonomia.....	25
Tabela 2. Datas dos voos realizados e resolução das imagens.....	41
Tabela 3. Altitude de voo utilizada por resolução para cada câmera	42
Tabela 4. Número de pontos de acordo com a taxa de cobertura de cada parcela de acordo com a resolução da imagem	52
Tabela 5. Número total de pontos de acordo com a taxa de cobertura por área experimental com base na resolução das imagens.....	52
Tabela 6. Exemplo de valores extraídos a partir do <i>Point Sampling Tool</i>	54
Tabela 7. Distâncias em metros entre GCPs e a coordenada exata em solo e RMSE para cada mosaico	56
Tabela 8. Resultados do teste de Tukey na verificação da diferença entre as distâncias dos tratamentos ¹	57
Tabela 9. Representação de parte da base de dados utilizada para MD com a câmera RGB para estimativa de produtividade do trigo na Fazenda Escola no período de 84DAS.....	62
Tabela 10. Representação de parte da base de dados utilizada para MD com a câmera NIR para estimativa de produtividade do trigo na Fazenda Escola no período de 84DAS.....	63
Tabela 11. Representação de parte da base de dados utilizada para MD com os valores de NDVI para estimativa de produtividade do trigo na Fazenda Escola no período de 84DAS.....	63
Tabela 12. Correlações de produtividade com cada tipo diferente de dados (Câmera RGB, NIR e o NDVI) com a proporção de 10% para as bases da Fazenda Santa Cruz.....	63
Tabela 13. Correlações de produtividade com cada tipo diferente de dados (Câmera RGB, NIR e o NDVI) com a proporção de 10% para as bases da Fazenda Escola..	64
Tabela 14. Tempo médio para geração do modelo preditivo de acordo com a proporção da base	65
Tabela 15. Bases de dados executadas no algoritmo de MD para Fazenda Santa Cruz, onde o X representa as bases executadas, e – (traço) as bases não executadas.....	66

Tabela 16. Bases de dados executadas no algoritmo de MD para Fazenda Escola, onde o X representa as bases executadas, e – (traço) as bases não executadas ...	66
Tabela 17. Correlação em diferentes proporções para Fazenda Santa Cruz	66
Tabela 18. Correlação em diferentes proporções para Fazenda Escola.....	67
Tabela 19. Valores de RGB_R nos diferentes períodos de coleta da Fazenda Santa Cruz em diferentes proporções ¹	69
Tabela 20. Valores de RGB_R nos diferentes períodos de coleta da Fazenda Escola em diferentes proporções ¹	70
Tabela 21. Modelos de estimativa de produtividade do trigo gerados em diferentes proporções para a Fazenda Santa Cruz ¹	70
Tabela 22. Modelos de estimativa de produtividade do trigo gerados em diferentes proporções para a Fazenda Escola ¹	71
Tabela 23. Coeficientes de correlação em diferentes resoluções. Fazenda Escola a 87 DAS.....	72
Tabela 24. Coeficientes de correlação em diferentes resoluções. Fazenda Escola a 98 DAS.....	72
Tabela 25. Modelos gerados para os maiores coeficientes de correlação ¹	73

LISTA DE SIGLAS

AAT	<i>Automatic Aerial Triangulation</i>
AP	Agricultura de Precisão
BBA	<i>Bundle Block Adjustment</i>
CONAB	Companhia Nacional do Abastecimento
DAS	Dias Após Semeadura
DECEA	Departamento de Controle do Espaço Aéreo
DSM	<i>Digital Surface Model</i>
GCP	<i>Ground Control Point</i>
GNSS	<i>Global Navigation Sattelite System</i>
GPS	<i>Global Positioning System</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
MD	Mineração de Dados
ICA	Instrução do Comando da Aeronáutica
IV	Índice de Vegetação
PDI	Processamento Digital de Imagens
RMSE	<i>Root Mean Square Error</i>
RPA	<i>Remotely Piloted Aircraft</i>
RPAS	<i>Remotely Piloted Aircraft System</i>
RVI	<i>Ration Vegetation Index</i>
SAVI	<i>Soil-Adjusted Vegetation Index</i>
SIG	Sistema de Informação Geográfica
SR	Sensoriamento Remoto
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
VANT	Veículo Aéreo Não Tripulado

SUMÁRIO

1	INTRODUÇÃO.....	13
2	OBJETIVOS	16
2.1	OBJETIVO GERAL	16
2.2	OBJETIVOS ESPECÍFICOS.....	16
3	REVISÃO DA LITERATURA.....	17
3.1	TRIGO E A AGRICULTURA DE PRECISÃO.....	17
3.2	SENSORIAMENTO REMOTO.....	19
3.3	AERONAVES REMOTAMENTE PILOTADAS	24
3.4	PROCESSAMENTO DIGITAL DE IMAGENS DE SENSORIAMENTO REMOTO.....	26
3.5	MINERAÇÃO DE DADOS	29
4	MATERIAL E MÉTODOS	33
4.1	ÁREAS DE ESTUDO	34
4.2	EQUIPAMENTOS UTILIZADOS.....	37
4.3	COLETAS DE DADOS	40
4.4	PROCESSAMENTO DAS IMAGENS.....	42
4.5	CORREÇÃO RADIOMÉTRICA DAS IMAGENS	45
4.6	GEORREFERENCIAMENTO DOS MOSAICOS.....	46
4.7	EXTRAÇÃO DE DADOS DAS IMAGENS	48
4.8	MINERAÇÃO DOS DADOS	54
5	RESULTADOS E DISCUSSÃO.....	55
5.1	MOSAICOS GERADOS PARA EXTRAÇÃO DOS DADOS	55
5.2	COMPORTAMENTO DAS BASES DE DADOS.....	57
5.3	RESULTADOS DA REGRESSÃO E MODELOS PREDITIVOS.....	62
6	CONCLUSÃO E TRABALHOS FUTUROS.....	75
	REFERÊNCIAS.....	77

1 INTRODUÇÃO

Os cereais desempenham um papel importante no desenvolvimento econômico de várias regiões, bem como na dieta humana. O trigo é o segundo cereal mais produzido no mundo, com grande peso na economia global. Com praticamente 55% de toda área de trigo plantada no Brasil, o Paraná é o principal estado produtor do país (CONAB, 2017; MAPA, 2013).

Devido às condições climáticas desfavoráveis à cultura em vários estados do país, o Brasil só consegue atender cerca de 60% da demanda nacional do trigo, o que implica em sua importação, afetando a cadeia produtiva brasileira com reflexos nos supermercados e indústrias de massas, biscoitos e pães.

A estimativa de produtividade, bem como a detecção do estado nutricional da cultura em seus diferentes estádios de desenvolvimento, preferencialmente nos estádios mais iniciais podem contribuir para a aplicação em tempo hábil de insumos que resultam no melhor desenvolvimento da planta e, conseqüentemente, em maior produtividade do trigo. Este é um dos objetivos da agricultura de precisão (AP), cujas práticas de culturas são adaptadas de acordo com seu estado atual, o que requer uma análise aprofundada da saúde das plantas e a variabilidade no campo durante o ciclo de crescimento da cultura (VERGER et al., 2014).

Existem algumas formas de se estimar a produtividade de uma cultura. Segundo Dalposso et al. (2013), uma delas é por meio do Sensoriamento Remoto (SR), o qual permite que dados sejam coletados sem que haja contato entre o sensor e o objeto ou área de interesse.

Torres-Sánchez et al. (2014) afirmam que uma das questões mais importantes do SR é quantificar os índices de vegetação (IV) de uma cultura, sendo um primeiro passo crucial na definição de objetivos futuros. Além disso, monitorar a variabilidade temporal e espacial dos índices de vegetação de uma cultura em uma área específica tem muitas aplicações para agricultura de precisão, como por exemplo, a identificação de degradação e desertificação da terra (XIAO et al., 2005), a estimativa dos estados físicos e fenológicos da cultura (YU et al., 2013), assim como na estimativa de produtividade (BRAGAGNOLO et al., 2013).

Métodos baseados em sensoriamento remoto são úteis para monitorar grandes áreas com um baixo custo. Verger et al. (2014) e Torres-Sánchez et al. (2014) afirmam que estes métodos eram tradicionalmente realizados por meio de

aviões tripulados, satélites ou equipamentos acoplados a tratores, porém, os recentes avanços tecnológicos e a miniaturização de sensores permitiram a criação das Aeronaves Remotamente Pilotadas (RPAs – do inglês *Remotely Piloted Aircraft*).

As RPAs, também conhecidas como *drone* (do inglês zangão), VANT (Veículo Aéreo Não Tripulado), UAV (do inglês *Unmanned Aerial Vehicle*), superam as limitações tradicionalmente associadas a essas plataformas comuns, pois têm a capacidade de voar em baixa altitude, possibilitando o imageamento de alta resolução. Além disso, as RPAs podem ser utilizadas mesmo em dias nublados e possuem alta flexibilidade no agendamento do sensoriamento (VERGER et al., 2014; TORRES-SÁNCHEZ et al., 2014). A denominação RPA para esses equipamentos foi adotada a partir de novembro de 2015, de acordo com a nova regulamentação do Departamento de Controle do Espaço Aéreo (DECEA), a Instrução do Comando da Aeronáutica (ICA) 100-40 (DECEA, 2015).

O sensoriamento remoto proporciona uma grande quantidade de dados, os quais precisam ser analisados e processados. A Mineração de Dados (MD) constitui um conjunto de técnicas e ferramentas para o tratamento e análise de grandes volumes de dados, possibilitando extrair informações úteis para o planejamento do controle e manejo em campo.

A utilização de RPAs na agricultura ainda é recente, sendo assim, pesquisas utilizando tais equipamentos são necessárias, para auxiliar produtores no processo de controle e manejo de culturas, visando melhorias no apoio à tomada de decisão e, conseqüentemente a melhoria na capacidade produtiva com redução de impactos ao meio ambiente.

Existem várias questões relevantes ainda em aberto sobre o uso de RPA na agricultura, como a resolução de imagem mais adequada para propósitos específicos. Ao mesmo tempo em que as RPAs permitem altas resoluções, de 3,4 cm/px por exemplo, contribuindo para a obtenção de maiores detalhes das áreas imageadas, essa maior resolução impacta em bases de dados maiores e, conseqüentemente, maior tempo de processamento. Torna-se importante, então, avaliar se a imagem de alta resolução, até então não obtidas com o uso de satélites e aviões tripulados, é viável para determinada aplicação agrícola, como a estimativa de produtividade.

Outra questão a ser estudada é a melhor forma de gerar os ortomosaicos a partir das imagens coletadas durante o voo. O ortomosaico corresponde a uma

única imagem da área de interesse, obtido a partir da combinação das várias imagens tomadas ao longo do voo. A quantidade de imagens coletadas depende da resolução utilizada, e uma vez que é a partir do ortomosaico que são extraídos os dados para análise da área de estudo, sua precisão pode interferir na qualidade das análises. Existem diferentes ferramentas e métodos para a geração do ortomosaico, não havendo ainda uma forma padrão que já tenha sido definida como a mais adequada a ser adotada para propósitos agrícolas.

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo principal desse trabalho é a criação de um modelo preditivo que utilize valores de refletância espectral em diferentes comprimentos de onda na estimativa de produtividade do trigo, por meio da mineração de dados obtidos em imagens em diferentes resoluções espaciais, coletadas por aeronave remotamente pilotada.

2.2 OBJETIVOS ESPECÍFICOS

- Verificar a precisão de georreferenciamento nos mosaicos gerados a partir de imagens coletadas com RPA.
- Verificar a relação da resolução das imagens com a estimativa de produtividade do trigo.
- Verificar a influência da proporção do número de pontos, utilizados para extração das informações das imagens, na estimativa da produtividade do trigo.
- Avaliar o uso dos canais RGB e NIR e do índice de vegetação NDVI na estimativa de produtividade do trigo.

3 REVISÃO DA LITERATURA

3.1 TRIGO E A AGRICULTURA DE PRECISÃO

O trigo é uma gramínea com ciclo anual, geralmente cultivada durante o outono/inverno, cuja principal forma de consumo é através da farinha, na fabricação de pães, bolos e biscoitos, ou na aplicação em ração animal. No Brasil, cerca de 90% da produção total desse cereal ainda está concentrada na região sul (RS, SC e PR), porém o trigo começa a avançar para região central do país. (CONAB, 2017; BORÉM e SHEEREN, 2015).

A produção do trigo no Brasil geralmente oscila próxima a 6 milhões de toneladas. Em 2016, alcançou o recorde nacional de 6,7 milhões de toneladas (21,5% acima da produção na safra anterior) segundo dados da Companhia Nacional do Abastecimento – CONAB. Mesmo com a alta na produção, ainda não é possível atender a toda a demanda nacional, que se mantém próxima a 11 milhões de toneladas. Devido a isso, entre agosto de 2015 e julho de 2016, o país importou cerca de 5,5 milhões de toneladas de trigo a um custo aproximado de 1,1 bilhão de dólares (MAPA, 2013; CONAB, 2017; CAMPONOGARA et al., 2015; BORÉM e SCHEEREN, 2015).

As importações do trigo entre o ano 2000 e 2015 implicaram em um gasto de cerca de 67,7 bilhões de reais (RABELO, 2016). Nesse sentido, pode-se perceber que caso a quantidade de trigo importado seja menor, valores que seriam destinados para economia internacional permaneceriam no Brasil, aquecendo a economia interna. Com uma previsão de aumento anual de consumo de 1,3%, torna-se essencial a aplicação de técnicas com o objetivo de aumentar a produção, minimizando a utilização de recursos não renováveis (CONAB, 2017; CAMPONOGARA et al., 2015; BORÉM e SCHEEREN, 2015).

O trigo é considerado uma cultura de alto risco, pois demanda um grande investimento desde a preparação do solo até a colheita, apresentando baixo retorno aos produtores. Nesse sentido, a agricultura de precisão (AP) pode ajudar na otimização dos recursos utilizados no cultivo da cultura do trigo. A AP pode ser definida como “um conjunto de ferramentas e tecnologias aplicadas para permitir um sistema de gerenciamento agrícola baseado na variabilidade espacial e temporal da

unidade produtiva, visando o aumento de retorno econômico e a redução de impacto ao meio ambiente” (MAPA, 2013).

Os primeiros relatos do trabalho com AP surgiram no começo do século XX, sendo que indícios de sua prática são encontrados na Europa e Estados Unidos da América (EUA) desde 1980. Na Europa, com a criação do primeiro mapa de produtividade, e nos EUA, relatos mostram a primeira aplicação de doses variadas de adubo (MAPA, 2013).

Segundo Bernardi et al. (2014) e MAPA (2013), foi com o surgimento do GPS em 1978 (*Global Positioning System* – Sistema de Posicionamento Global), através de uma cadeia de satélites que se tornou operacional em 1995, que possibilitou o passo determinante em direção a implementação da AP, onde foi possível a instalação de receptores GPS em colhedoras, os quais permitiram a coleta de dados georreferenciados de produção de forma instantânea.

Bernardi et al. (2014) considera a AP como um ciclo de três etapas: (i) Leitura, que consiste no levantamento e coleta de dados; (ii) Interpretação dos dados, com foco na análise dos dados para o planejamento das operações em campo e (iii) atuação, que é execução do planejamento obtido através da segunda etapa.

Tais etapas têm sido muito aplicadas à cultura do trigo, mostrando retorno econômico aos produtores, como mostrado no trabalho de Shadeck e Cardoso (2016), em que através da AP, obteve-se uma redução de 43% no uso de calcário durante todo o ciclo da cultura, com um aumento de 28% na produtividade. Ainda, estudos têm sido feitos com intuito de analisar como atributos químicos do solo podem afetar a produtividade do trigo, conforme mostrado por Kramer et al. (2014), mostrando que a produtividade do trigo é positivamente associada a teores de carbono orgânico, PH do solo, Cálcio, Magnésio e Potássio, e negativamente com teores de Alumínio.

Além disso, estudos buscando encontrar formas de auxílio na identificação de pragas, plantas daninhas e doenças do trigo têm sido desenvolvidos baseados em técnicas de AP, com objetivo de controlar a aplicação de pesticidas e fertilizantes para evitar danos ao meio ambiente. Nesse sentido, Tartachnyk et al. (2006) mostraram que é possível identificar a fungos e deficiência de nitrogênio em plantas de trigo com base em fluorescência induzida por laser.

As etapas do ciclo da AP têm avançado constantemente, em grande parte devido ao apoio de técnicas e tecnologias que fornecem suporte à sua execução,

como a utilização de sensoriamento remoto na etapa de leitura. Na etapa de interpretação, o auxílio vem por parte dos sistemas de informação geográfica (SIG), geoestatística e técnicas de mineração de dados. Para atuação em campo, o maquinário agrícola, que conta hoje com tecnologias altamente sofisticadas, as quais permitem a real atuação da AP no campo (BERNARDI *et. al.*, 2014).

3.2 SENSORIAMENTO REMOTO

O termo Sensoriamento Remoto (SR) foi criado por volta do ano de 1960 por Evelyn L. Pruitt e sua equipe e é uma das tecnologias mais bem sucedidas para a coleta automática de dados para levantamento e monitoramento de recursos terrestres. SR pode ser definido como um conjunto de técnicas e ferramentas para obtenção de informações de objetos que compõem a superfície terrestre, sem que haja contato físico entre o sensor e o objeto (MENESES e ALMEIDA, 2012; FIGUEIREDO, 2007).

Segundo Figueiredo (2007) o SR teve início com a invenção da câmera fotográfica, que foi o primeiro instrumento usado para o sensoriamento e é utilizada até hoje. Existem dezenas de satélites em órbita pela terra, sensoriando todas as faixas espectrais possíveis, alguns com sensores hiperespectrais com centenas de bandas e outros com sensores de alta resolução espacial capazes de identificar objetos na superfície com menos de 50 centímetros (MENESES e ALMEIDA, 2012).

Alguns dos principais satélites utilizados para sensoriamento remoto são os do projeto Landsat, que se encontra atualmente na versão oito, e conta com um sensor multiespectral (com resolução de 30 metros), um pancromático (resolução de 15 metros) e um sensor infravermelho térmico (com resolução de 30 metros). Além dos satélites Landsat, existem outros como o IKONOS e o QuickBird, que possuem um sensor multiespectral com a resolução espacial de 4m e 2.4m, respectivamente (USGS, 2016).

Outros dois satélites em órbita que capturam imagens em resolução maior são o GeoEye-1 e o WorldView 2, com resoluções de 1.65m e 1,8m, respectivamente (YANG, 2012). Em satélites atuais, a resolução espacial alcançada é ainda menor, como no caso do WorldView 4, lançado em 2016, com capacidade de conseguir imagens pancromáticas e no espectro do visível em resoluções de até 31cm, e multiespectrais de até 1,24m (SIC, 2016).

Além de satélites, os dados de sensoriamento remoto podem ser obtidos a partir de outras plataformas, como aviões, sensores acoplados em máquinas agrícolas, sensores de mão e RPAs (BERNARDI et al., 2014). Todos os sensores, independente da plataforma de obtenção dos dados, trabalham com o conceito de refletância que segundo Meneses e Almeida (2012) pode ser entendido como a razão entre a quantidade de energia refletida por uma unidade do terreno/objeto (Radiância) pela quantidade de energia incidente na mesma unidade (Irradiância), no mesmo instante de tempo.

Como a irradiância e a radiância são densidades de fluxo, o valor da razão (refletância) torna-se adimensional, sendo expresso em porcentagem. A partir disso, cada alvo no terreno poderá refletir de 0 a 100% da radiação incidente, ou seja, cada material irá refletir e absorver a energia incidente de forma diferente, em comprimentos de onda diferentes. Muitos desses materiais possuem padrões de reflexão e absorção diferentes no decorrer do espectro (Figura 1).

O percentual de refletância de alguns materiais é expresso no eixo Y, considerando cada banda no espectro no eixo X. Como pode ser observado, cada um dos materiais possui uma assinatura de refletância diferente comparado uns aos outros, e segundo Meneses e Almeida (2012) isso é essencial para que haja a identificação e distinção de objetos alvo em uma imagem de sensoriamento remoto.

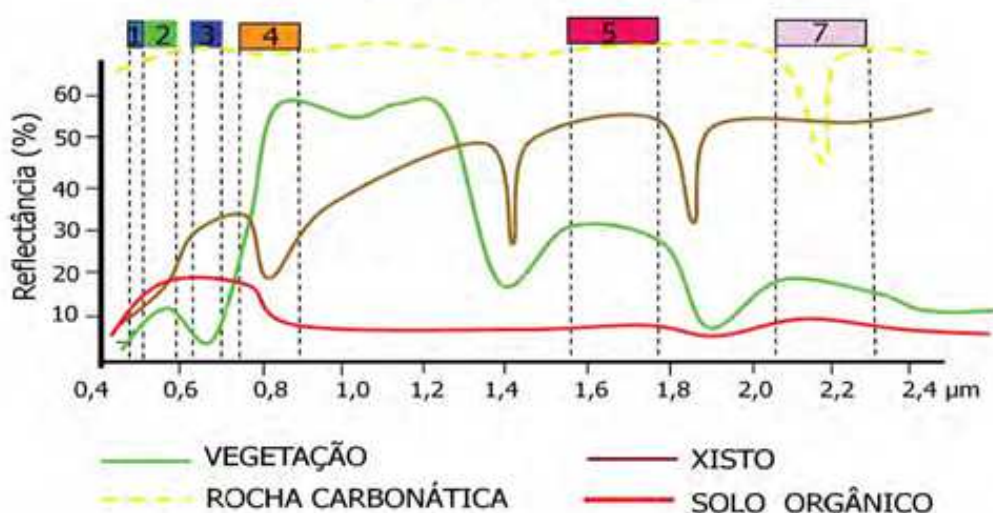


Figura 1. Padrão de reflexão de diferentes materiais no decorrer do espectro
Fonte: MENESES e ALMEIDA (2012).

A combinação de dados/imagens de diferentes bandas do espectro, obtida através de modelos matemáticos, pode fornecer uma melhor avaliação das

propriedades de refletância. Em geral, essas combinações utilizam as faixas do visível (RGB – *Red*, *Green* e *Blue*) e infravermelho próximo (NIR – *Near Infrared*), e são denominadas de Índices de Vegetação (IV) (BERNARDI et al., 2014).

3.2.1 Índices de Vegetação

De acordo com Santos (2005), foi a partir do momento em que se descobriu que os vegetais absorvem parte da radiação eletromagnética do sol para viver, que a relação entre vegetação e radiação começou a ser estudada mais a fundo. Sendo uma técnica do sensoriamento remoto, os índices de vegetação são combinações lineares de dados espectrais, que realçam a vegetação em imagens, o que facilita e permite o monitoramento sazonal, interanual e variações de longo prazo dos parâmetros estruturais, fenológicos e biofísicos da vegetação. Essa tarefa assume um papel importante no monitoramento e estimativa de diversos fenômenos meteorológicos e ambientais.

Conforme afirmado por Bernardi et al. (2014), na região espectral do infravermelho, que se divide em infravermelho próximo, infravermelho médio e infravermelho distante, é onde a reflexão na estrutura interna das folhas é muito forte. Sendo assim, grande parte dos trabalhos envolvendo índices de vegetação trabalha com imagens dentro dessas bandas.

Existem alguns sensores remotos que calculam esses índices automaticamente, um deles é o *Greenseeker*, que segundo Grohs et al. (2009), foi desenvolvido pela Universidade de Oklahoma na década de 1990, e utiliza diodos emissores de radiação e um microprocessador para analisar a radiação refletida. Porém, esse aparelho possui algumas limitações, e as três principais delas são: (i) precisa estar a uma pequena distância do solo, e por isso é geralmente acoplado a tratores ou máquinas agrícolas, (ii) a área de cobertura pode ser pequena e (iii) possui custo alto. Outro sensor utilizado nos cálculos de índice de vegetação é o *Crop Circle* e o *N-Sensor*, que também são sensores de solo, sendo muito utilizados por produtores e pesquisadores. Porém, possuem as mesmas limitações que o *Greenseeker*, em relação à altitude, cobertura do terreno e custo.

Outra forma de obtenção dos índices de vegetação é através do processamento digital de imagens (PDI), considerando diferentes bandas no espectro. Segundo Santos (2005), existem diversos tipos de índices de vegetação

que podem ser extraídos através do PDI, e Bernardi et al. (2014) afirma que já foram criados inúmeros IVs para diferentes finalidades.

O índice mais conhecido universalmente é o RVI (*Ratio Vegetation Index*), gerado através da razão entre a banda situada na região do infravermelho (NIR) e a banda do vermelho (RED), conforme apresentado na Equação 1. No infravermelho, as plantas mostram alta refletância, e na banda do vermelho, é onde as plantas apresentam maior absorção da radiação, devido à clorofila presente em sua constituição (MENESES e ALMEIDA, 2012).

$$RVI = \frac{NIR}{RED} \quad (1)$$

A partir de uma imagem original (Figura 2a), o resultado da divisão da banda de maior refletância pela banda de menor refletância é uma imagem monocromática (Figura 2b), em escala de cinza, onde tons de cinza mais claro indicam uma alta concentração de vegetação, e tons escuros mostram a ausência de vegetação.

Outro índice similar é o NDVI (*Normalize Difference Vegetation Index*), que é calculado através da razão entre a diferença e a soma das bandas do infravermelho e vermelho. Segundo Meneses e Almeida (2012), esse índice é mais vantajoso por ser mais proporcional à biomassa das plantas e mais indicado quando pretende-se fazer comparações ao longo do tempo em uma mesma área de estudo. O cálculo do NDVI é apresentado na Equação 2.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2)$$

O resultado dessa razão também é uma imagem monocromática em escala de cinza, com valores variando entre 1 e -1 (Figura 2c). Quanto mais próximo de 1, indica a presença de alta concentração de biomassa, quanto mais próximo de -1 mostra a ausência de biomassa (MENESES e ALMEIDA, 2012).

Existem outros índices, cada um com sua particularidade, fornecendo aspectos diferentes para os mais diversos tipos de aplicação. Porém, segundo

Meneses e Almeida (2012) e Bendig et al. (2015), o índice mais relevante para quantificação de biomassa, e o mais utilizado pela comunidade científica é o NDVI, considerado essencial para auxílio na estimativa de produtividade de culturas.

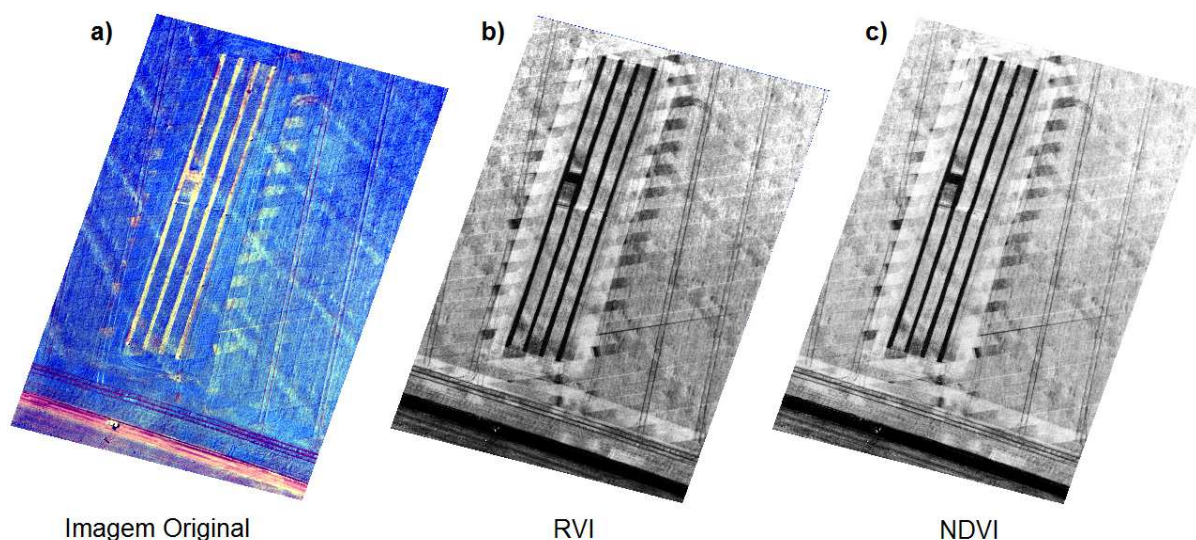


Figura 2. Imagens resultantes a partir do cálculo de índices de vegetação. (a) Imagem original; (b) RVI e (c) NDVI

Os únicos casos em que o NDVI não é indicado, como mostrado por Huete (1988), são em regiões áridas e semi-áridas, uma vez que a vegetação presente é pobre. Dessa maneira foi proposto o SAVI (*Soil-Adjusted Vegetation Index*), apresentado na Equação 3, um índice que é melhor para áreas com baixa cobertura vegetal.

$$SAVI = \left(\frac{NIR - RED}{NIR + RED} \right) (1 + L) \quad (3)$$

Nesse caso, a constante L é utilizada para minimizar a sensibilidade do índice com a refletância do solo, podendo variar de -1 a 1, de acordo com o nível de cobertura do solo (MENESES e ALMEIDA, 2012).

Com o advento da agricultura de precisão, juntamente com o sensoriamento remoto, tornou-se possível a análise de vigor da cultura e sua relação com a produtividade. Araújo et al. (2005) utilizaram imagens multiespectrais obtidas através de videografia aérea na análise e estimativa de produtividade do trigo. Nesse trabalho, através de regressão linear simples, os autores mostraram que o índice de

vegetação por diferença normalizada (NDVI) tem a tendência de ser mais eficiente na estimativa de produtividade do trigo, com uma correlação de 0,819.

Através de índices de vegetação derivados de imagens do satélite Landsat TM/ETM, Satir e Berberoglu (2016) utilizaram regressão linear e multilinear para estimar a produtividade do trigo, milho e algodão. Os autores concluíram que os períodos em que o NDVI é mais alto são mais sensíveis para a estimativa de rendimento do trigo, obtendo uma correlação de 0,67. Da mesma maneira Wang et al. (2014), utilizando regressão linear com dados de imagens dos satélites SPOT-5 e HJ-CCD obtiveram os maiores índices de correlação de 0,886 e 0,871 para o NDVI, na estimativa de produtividade do trigo em duas áreas de cultivo, respectivamente.

3.3 AERONAVES REMOTAMENTE PILOTADAS

Segundo Bernardi et al. (2014) as aeronaves remotamente pilotadas surgiram como uma importante opção na agricultura de precisão, pois sua aplicação é favorecida pelo crescente avanço e desenvolvimento tecnológico, principalmente pela redução de custos e tamanho dos equipamentos.

A definição de RPA, segundo o DECEA (2015) consiste em uma aeronave não tripulada pilotada a partir de uma estação de pilotagem remota. Associado a RPA está o Sistema de Aeronave Remotamente Pilotada (RPAS – do inglês *Remotely Piloted Aircraft System*) que consiste da RPA, sua(s) estação(ões) de pilotagem remota, o enlace de pilotagem e qualquer outro componente, como especificado no projeto da aeronave.

Existe uma variedade de RPAs, porém três componentes são comumente identificados: (i) A RPA, (ii) a estação de controle em terra e (iii) o canal de comunicação para troca de dados (COLOMINA, 2014).

Eisenbeiss (2009) afirma que a principal vantagem das RPAs comparadas com as aeronaves tripuladas é que a RPA pode ser utilizada em situações de alto risco, com necessidade de sobrevoar em baixa altitude, onde as aeronaves comuns não poderiam ser utilizadas, pois colocariam a vida de pessoas em perigo. Outras vantagens são citadas, como a capacidade de coleta rápida de dados mesmo com clima desfavorável e o baixo custo de operação.

Devido à grande quantidade de equipamentos e funções, não existe ainda uma classificação oficial para as RPAs, porém autores têm classificado esses sistemas com base em peso, alcance, altura de voo e autonomia, conforme mostra a Tabela 1.

Tabela 1. Classificação de categorias de RPAs a partir da massa, alcance, altitude e autonomia

Categoria	Massa (kg)	Alcance (km)	Altitude (m)	Autonomia (h)
Micro	Menos de 5	Menor que 10	250	1
Mini	Entre 25 e 150	Menor que 10	Entre 150 e 300	Até 2
Curto Alcance	Entre 25 e 100	Entre 10 e 30	3000	De 2 a 4
Médio Alcance	Entre 50 e 250	Entre 30 e 70	3000	De 3 a 6
HALE	Mais de 250	Maior que 70	Maior que 3000	Maior que 6

Adaptado de Eisenbeiss (2009) e Bernardi et al. (2014).

Na agricultura de precisão, o controle de vegetação e biodiversidade é tradicionalmente realizado com aviões tripulados e/ou satélites imageadores, o que implica em altos custos operacionais, principalmente quando a alta resolução de imagens é necessária. Já as RPAs têm sido utilizadas com foco nos mesmos tipos de sensores (baseados na refletância), pois introduzem uma plataforma pequena e com baixo custo de operação (BERNARDI et al., 2014; COLOMINA et al., 2014).

Para utilização de RPAs em agricultura de precisão é necessário que algumas etapas sejam seguidas, resumindo-se basicamente em: (i) planejamento de voo, (ii) voo com sobreposição (coleta de mais de uma imagem de uma mesma parcela da área de interesse), (iii) obtenção de imagens georreferenciadas, (iv) processamento das imagens, (v) geração de ortomosaico (criação de uma única imagem da área de interesse), (vi) análise em uma ferramenta de informação geográfica, e (vii) geração de relatórios (BERNARDI et al., 2014).

As pesquisas utilizando RPAs na agricultura ainda são recentes, porém nos últimos anos, trabalhos vêm sendo publicados mostrando a aplicabilidade desses equipamentos na agricultura, como por exemplo no trabalho de Silva et al. (2016), em cujo trabalho os autores utilizaram imagens obtidas por uma RPA na classificação do uso de cobertura de solo no ambiente de cerrado. Por meio de

fotográficas com resolução espacial de aproximadamente 10cm/px, foram utilizadas técnicas de classificação (agrupamento) supervisionada e não supervisionada que alcançaram correlações de 94%.

No trabalho de López-Granados et al. (2015), foi utilizado um microdrone operado através de controle remoto equipado com uma câmera RGB e outra multiespectral, para geração de um mosaico na plantação de oliveiras com objetivo de estimar a área, volume e a altura das árvores, parâmetros essenciais na verificação da saúde e produção da cultura. Tal metodologia apresentou resultados com 97% de precisão. Bendig et al. (2015) confirmaram a possibilidade de se extrair índices de vegetação a partir de imagens coletadas com uma RPA, utilizando uma câmera RGB, mostrando que esses índices têm grande potencial na estimativa de biomassa, na cultura da cevada.

Tattaris et al. (2016) utilizaram um veículo aéreo não tripulado na obtenção de imagens para estimativa de produtividade do trigo. Esse estudo foi conduzido no centro internacional e melhoramento do trigo e milho, no estado do Novo México nos Estados Unidos da América, onde os autores alcançaram um índice de correlação de 0,75 para o NDVI. Tais resultados foram melhores que a estimativa feita com base em imagens de satélite, que alcançou um índice de correlação de 0,53.

3.4 PROCESSAMENTO DIGITAL DE IMAGENS DE SENSORIAMENTO REMOTO

Segundo Gonzalez e Woods (2008) o campo do processamento digital de imagens se refere ao processamento de imagens digitais por meio de um computador digital, ou seja, PDI pode ser entendido como processos que extraem atributos através de técnicas computacionais, onde a entrada e saída são imagens.

O PDI é composto por várias etapas, onde as principais são (GONZALEZ e WOODS, 2008):

- Aquisição da imagem: processo de obtenção da imagem através de um sensor de imageamento;
- Realce da imagem: é o processo de manipulação de uma imagem de forma que o resultado é mais adequado do que o original para uma aplicação específica;
- Restauração de imagem: é o processo de aperfeiçoamento da qualidade de uma imagem;

- *Wavelets*: permite a representação de imagens em vários graus de resolução;
- Compressão: visa reduzir a quantidade de espaço necessária para armazenamento de uma imagem;
- Segmentação: consiste em separar uma imagem de entrada em várias partes;
- Descrição: também conhecida como seleção de características, onde características básicas são extraídas para diferenciação de classes de objetos;
- Reconhecimento/Classificação: é o processo de atribuir um rótulo a um objeto da imagem de acordo com as informações de descrição;

O PDI pode ser aplicado em diversas áreas, sendo que na agricultura é especificamente utilizado para imagens de sensoriamento remoto. O PDI de SR é orientado a cada tipo de problema e configurado por algoritmos especializados, disponibilizando para os usuários várias técnicas de processamento (MENESES e ALMEIDA, 2012).

Os *softwares* especializados em PDI de SR executam tarefas que podem ser divididas basicamente em duas etapas: realce e classificação, cujo objetivo principal é corrigir os erros que as imagens de entrada possam ter. Meneses e Almeida (2012) afirmam que os problemas identificados em fotografias de sensoriamento remoto são originados em sua grande maioria devido a defeitos e erros que os sensores apresentam durante a sua vida útil e por instabilidade na plataforma em que o sensor está acoplado. Além desses erros, as imagens de SR podem sofrer influência externa, que por sua vez geram distorções nos dados coletados.

A atmosfera é um dos fatores externos que pode causar o efeito chamado de Atenuação Atmosférica, que geralmente se apresenta na forma de deformações de escala, incorreções na posição espacial de cada pixel, perda de contraste/brilho e registros incorretos dos valores digitais em pixels. Tais distorções são observadas em imagens obtidas com satélites e aviões, devido a altitude. (SOUZA et al., 2013; MENESES e ALMEIDA, 2012).

O tipo de correção a ser feita depende da aplicação, mas algumas técnicas são comumente utilizadas, como a (i) Eliminação de Ruídos, que são considerados valores digitais errados em pixels isolados ou em um conjunto de pixels. (ii) Correção

Atmosférica, que tem por objetivo eliminar ruídos causados pela atmosfera, tendo em vista que esta pode influenciar imagens de diferentes formas em diferentes bandas no espectro;

(iii) Correção de Distorções Geométricas, que tem por objetivo remover distorções causadas devido à rotação, inclinação e curvatura da terra e pela visada cônica dos sensores, através de modelos matemáticos (MENESES e ALMEIDA, 2012).

Além das correções citadas, Meyer et al. (1993) e Teillet (1986) destacam que um outro processo de correção em imagens aéreas, denominado correção radiométrica, é uma prática comum e necessária, uma vez que os números digitais (ND) que representam o brilho de tais imagens não podem ser descritos apenas em relação às propriedades de reflexão e emissão dos objetos/áreas observados. Tais características podem ser influenciadas por outros efeitos, os quais podem ser divididos em efeitos induzidos por sensores e efeitos relacionados à cena.

Os efeitos induzidos por sensores são relativos aos sensores de coleta de dados, pois são influenciados por aspectos técnicos, como calibração, filtragem de dados, estabilidade no momento da coleta, etc., que impactam nos resultados das imagens. Já os efeitos relacionados à cena, são características da área/objeto de estudo que podem influenciar na qualidade dos dados, como topografia, atmosfera, ângulo de visão, posição do sol e as propriedades de refletância dos objetos (TEILLET, 1986).

Calderón et al. (2013) realizaram a correção radiométrica de imagens utilizadas em seu trabalho através de um sistema de fonte luminosa uniforme (CSTM-USS-2000C). No trabalho de Shi et al. (2016), imagens foram corrigidas radiometricamente com base na refletância de alvos (quadrados de concreto) pintados de preto (10% de refletância), cinza escuro (20% de refletância) e cinza claro (40% de refletância), obtidas por meio de um espectrômetro de mão (FieldSpec, ASD, Boulder, CO, USA).

Um processo similar foi utilizado por Berni et al. (2009), onde alvos brancos e escuros foram empregados para a correção radiométrica das imagens. Para leitura de refletância também foi utilizado o espectrômetro de mão (FieldSpec Handheld PRO, ASD, CO, USA) que foi previamente calibrado através de um painel branco Spectralon (SRT-99-180) com 99% de refletância, permitindo o cálculo exato da refletância dos alvos para posterior correção das imagens. Tattaris et al. (2016)

aplicaram correção radiométrica de vinhetamento e correção geométrica antes da utilização das imagens obtidas por uma RPA na estimativa de produtividade do trigo.

Além das técnicas de correção de imagens, outra parte do PDI importante para a agricultura de precisão é a aritmética de bandas, pois é a principal responsável pela obtenção dos índices de vegetação. Meneses e Almeida (2012) afirmam que essa é uma operação muito simples e rápida de ser executada, permitindo que muitas informações importantes sejam extraídas através da combinação (soma, subtração, multiplicação e divisão) dos valores digitais de imagens, com a geração uma imagem completamente nova.

3.5 MINERAÇÃO DE DADOS

Todo trabalho por meio de sensoriamento remoto resulta em uma grande quantidade de dados (e.g. tabelas de análise com milhares de linhas), os quais precisam ser processados e analisados, com objetivo de extrair informações úteis no processo de tomada de decisão na agricultura. É nesse sentido que a mineração de dados entra como ferramenta para o tratamento de tais dados.

A mineração de dados pode ser definida como uma etapa do processo de descoberta de conhecimento, que combina métodos tradicionais de análise de dados com algoritmos sofisticados para processar grandes volumes de dados. É entendida como o processo de descoberta de informações úteis em grandes repositórios de dados (TAN, STEINBACH e KUMAR, 2009).

Várias tecnologias são utilizadas para permitir que a mineração de dados aconteça, como por exemplo: (i) estatística, (ii) aprendizado de máquina, (iii) sistemas de banco de dados e (iv) recuperação de informações. A descoberta de conhecimento é baseada no aprendizado sobre os dados, podendo ser do tipo descritivo ou preditivo. (HAN e KAMBER, 2014; TAN, STEINBACH e KUMAR, 2009)

O aprendizado descritivo analisa os dados e identifica similaridades (agrupamentos) ou associações (regras de associação). O aprendizado preditivo analisa os dados que representam eventos passados, buscando relações que permitam prever situações em novos dados futuros, tais como a classificação, para previsões de valores discretos, e a regressão para previsões de valores contínuos (CARVALHO E DALAGASSA, 2014)

A análise preditiva pode ser efetuada por meio de diferentes técnicas, entre elas estão as Máquinas de Vetor de Suporte (SVM – Do inglês *Support Vector Machines*). SVM emprega um algoritmo altamente preciso, com capacidade de criar modelos lineares e não lineares complexos, utilizado para tarefas de classificação e regressão. Basicamente, ela busca transformar os dados para uma nova dimensão, onde procura achar uma separação ótima linear, ou seja, um limite que separa os dados de uma classe para outra (no caso da classificação) ou um limite que se ajuste aos dados (no caso da regressão) (HAN e KAMBER, 2012).

Como exemplo, pode ser observado na Figura 3, em que existem duas classes separadas de dados, cujas linhas pontilhadas representam margens para a separação entre classes, ou seja, existem várias formas de separação desses dados (HAN e KAMBER, 2012).

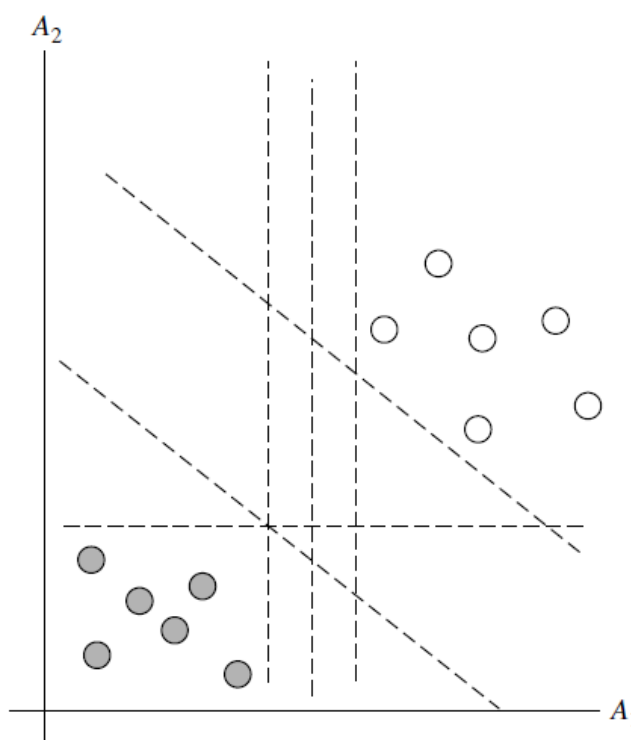


Figura 3. Demonstração da separação dos dados que a SVM busca
 FONTE: Han e Kamber (2012).

A SVM tenta achar os melhores limites para separação, ou seja, a separação que apresente o menor erro de classificação. Para isso o algoritmo procura pelo hiperplano com as maiores margens de separação entre as classes de dados. Na Figura 4 podem ser observadas duas possibilidades de separação: a primeira (Figura 4a) com uma margem pequena e a segunda (Figura 4b) com uma margem

maior. Ambas as possibilidades têm a capacidade de classificar os dados corretamente, porém as separações com margens maiores tendem a ser mais precisas (HAN e KAMBER, 2012).

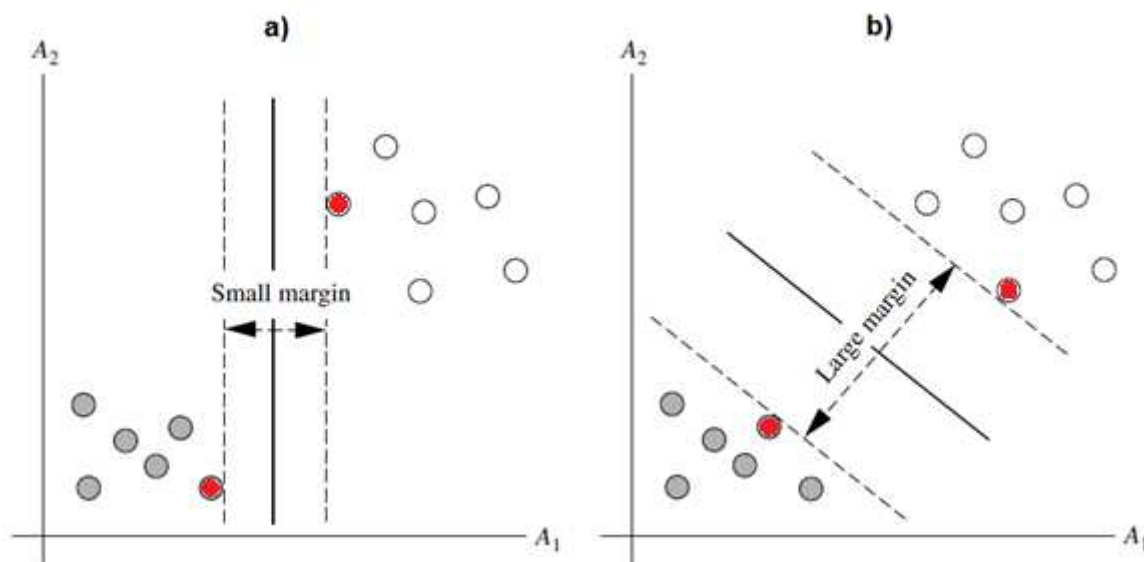


Figura 4. Possibilidades de separação dos dados. (a) margem pequena e (b) margem grande
 FONTE: Han e Kamber. (2012)

Para encontrar os limites de separação com as maiores margens, o algoritmo utiliza dos vetores de suporte, que é o conjunto de dados mais próximos das margens de separação (pontos em vermelho na Figura 4). Conforme afirmado por Han e Kamber, (2012), os vetores de suporte são os dados mais difíceis de classificar, porém proveem a maior quantidade de informação ao modelo, por isso são utilizados na descoberta do hiperplano de separação dos dados com o auxílio de vários mecanismos matemáticos.

As SVMs são muito menos propensas ao *overfitting* do que outros métodos, fornecendo uma descrição compacta do modelo aprendido. As SVMs são aplicadas a várias áreas, incluindo reconhecimento de dígitos manuscritos, reconhecimento de objeto e reconhecimento de voz, bem como testes de previsão de séries temporais de referência (HAN et al., 2012).

No caso das SVM aplicadas a regressão (SVR – do inglês *Support Vector Regression*), a mesma idéia das margens é mantida, porém o objetivo do algoritmo é produzir um classificador ajustado aos dados de entrada, também denominado de modelo de predição, capaz de prever novos dados com base em dados de entrada. Esse processo é denominado treinamento, onde o resultado é um

classificador (função $f(x)$) que tem a capacidade de receber um valor de entrada, fornecendo uma predição de saída. Um exemplo disso é o trabalho de Bose et al. (2016), cujos autores utilizaram vários algoritmos para estimativa de produtividade do trigo, dentre eles o SVR, onde os dados de entrada eram valores de NDVI e a saída era um valor estimado de produtividade. Com o SVR foi obtido um índice de correlação de 0,67.

Gill et al. (2006) apresentaram uma metodologia para estimativa de umidade de solo através de máquina de vetor de suporte, onde os coeficientes de correlação chegaram a 0,89, mostrando que o SVR tem muito potencial na estimativa de variáveis agronômicas.

4 MATERIAL E MÉTODOS

Nesta seção serão abordadas as principais etapas executadas no desenvolvimento desse trabalho, como pode ser observado na Figura 5, começando com a definição e descrição das áreas experimentais onde este estudo foi realizado (seção 4.1), definição e descrição dos equipamentos utilizados para o levantamento de dados (seção 4.2), descrição de como foram realizadas as coletas das imagens das áreas de estudo (seção 4.3), processamento das imagens coletadas (seção 4.4), correções realizadas nas imagens (seção 4.5 e 4.6), extração dos dados e criação das bases de dados (seção 4.7) e por fim a mineração dos dados e geração dos modelos preditivos (seção 4.8).

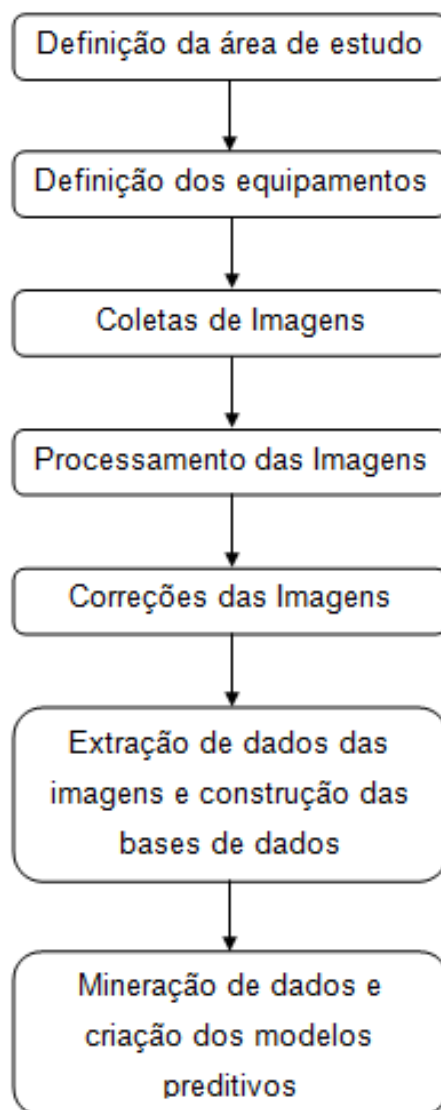


Figura 5. Fluxograma com etapas realizadas no decorrer do projeto

Tais etapas foram fundamentais na criação dos principais objetos de estudo desse trabalho, as bases de dados, as quais foram utilizadas para geração dos modelos preditivos a partir da aplicação da mineração de dados.

4.1 ÁREAS DE ESTUDO

O trabalho foi realizado em duas áreas experimentais cultivadas com trigo na cidade de Ponta Grossa – Paraná. A primeira encontra-se na Fazenda Santa Cruz (Figura 6), cuja área situada nas coordenadas geográficas 25°13'31,3" de latitude Sul, 50°7'7,3" de longitude oeste. De acordo com o sistema de classificação climática de Köppen baseado na vegetação, temperatura e pluviosidade, a região dos Campos Gerais pertence ao clima Cfb – Clima temperado; temperatura média no mês mais frio abaixo de 18°C, com verões frescos, temperatura média no mês mais quente abaixo de 22°C e sem estação seca definida, a precipitação média anual é de 1422 milímetros, sendo janeiro o mês mais chuvoso e agosto o mais seco (IAPAR, 2015).



Figura 6. Área experimental da Fazenda Santa Cruz

O trigo, cultivar TBIO Sinuelo foi semeado no dia 23/06/2016. Utilizou-se área experimental de um estudo com e sem tratamento de sementes com *Azospirillum* e aplicação de diferentes doses de N e gesso. Tal experimento foi dividido em 16 tratamentos com 3 repetições cada, totalizando 48 parcelas de aproximadamente 35m² cada.

O segundo experimento foi realizado na Fazenda Escola da Universidade Estadual de Ponta Grossa (UEPG) (Figura 7), situada nas coordenadas geográficas 25°13' de latitude Sul, 50°01' de longitude Oeste e 869 metros de altitude, mantendo os mesmos aspectos climáticos e pluviométricos da Fazenda Santa Cruz. Nessa área foi realizado um estudo envolvendo a aplicação de calcário e fertilizantes na cultura do trigo. A cultivar Turok foi semeada em 01/07/2016, com 11 tratamentos e 4 repetições cada, totalizando 44 parcelas com 90m² cada.



Figura 7. Área experimental da Fazenda Escola UEPG

Em relação a produtividade observada nas áreas, na Fazenda Santa Cruz (Figura 8), os valores ficaram entre 2300,3kg ha⁻¹ e 4236,4kg ha⁻¹, com uma média de 3309,4kg ha⁻¹. Já para os valores da Fazenda Escola (Figura 9), o valores obtidos foram maiores, entre 2359,4kg ha⁻¹ e 7012,7kg ha⁻¹, e uma média de 4191,8kg ha⁻¹.

No geral, a média de produtividade da Fazenda Santa Cruz foi mais baixa do que a média paranaense e a média brasileira, diferentemente da Fazenda Escola,

onde os valores foram mais altos e a média de produtividade foi mais alta que a média paranaense e brasileira.

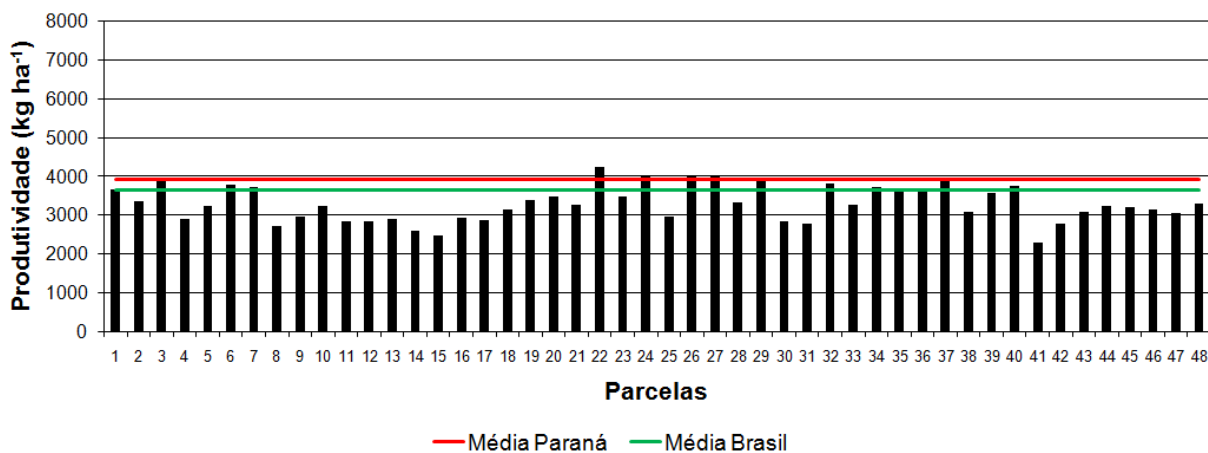


Figura 8. Produtividade por parcelas da Fazenda Santa Cruz

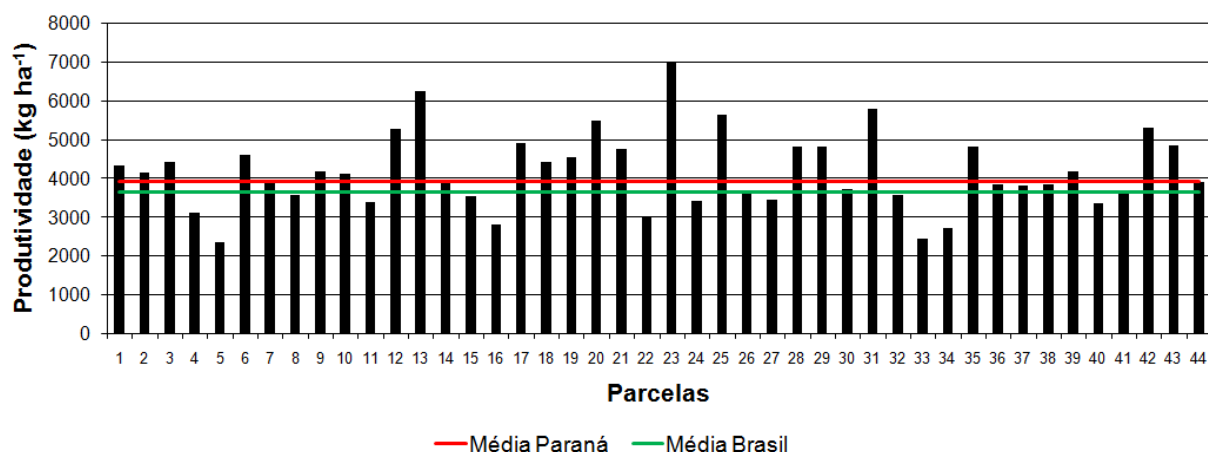


Figura 9. Produtividade por parcelas da Fazenda Escola

Outro aspecto importante é o comparativo entre a produtividade nas repetições de cada tratamento. Como pode ser visualizado, tanto na Fazenda Santa Cruz (Figura 10) quanto na Fazenda Escola (Figura 11), no geral, os valores variaram bastante. Com exceção de poucos tratamentos que mantiveram a produtividade próxima entre as repetições, como por exemplo o tratamento 7, 8 e 15 na Fazenda Santa Cruz e o tratamento 2 na Fazenda Escola .

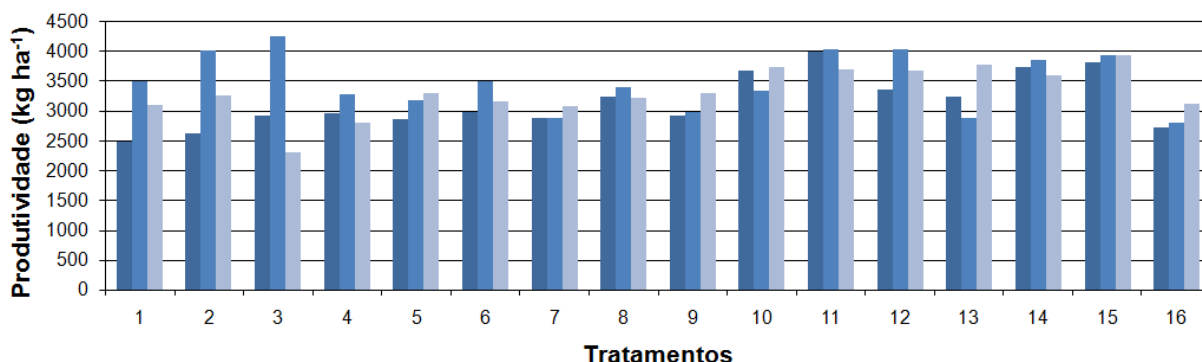


Figura 10. Produtividade das três repetições de cada um dos 16 tratamentos da Fazenda Santa Cruz

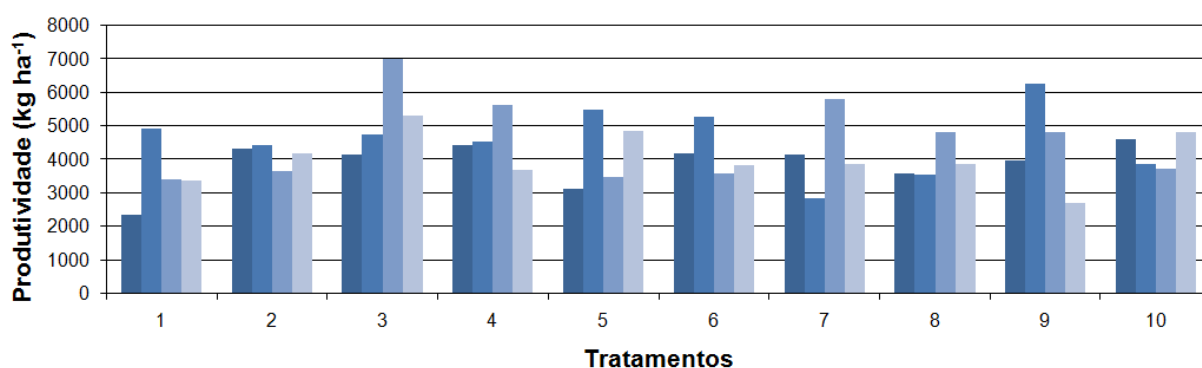


Figura 11. Produtividade das quatro repetições de cada um dos 10 tratamentos da Fazenda Escola

4.2 EQUIPAMENTOS UTILIZADOS

Para a coleta de dados e imageamento da área, foi utilizada uma RPA de asa fixa modelo eBee (Figura 12), fabricada pela SenseFly¹, adquirido pelo Programa de Pós Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa.

Essa RPA permite realizar a tomada de fotos de forma rápida e sob demanda, incluindo inteligência artificial para realizar a decolagem, o voo e a aterrissagem automaticamente. Além da aeronave foram utilizadas duas câmeras, uma Sony Cybershot WX RGB com 18.2 megapixels (Figura 13a) e uma Canon S110 NIR de 12 megapixels (Figura 13b).

As câmeras já vieram pela empresa fornecedora com um conjunto de parâmetros configurados, para garantir que as imagens obtidas sejam as mais precisas considerando a aplicação na agricultura. A configuração mais importante, principalmente na câmera NIR é a relação de velocidade do obturador, a qual deve

¹ <https://www.sensefly.com/drones/ebec.html>

ser a menor possível para evitar borrões nas imagens, uma vez que a câmeras estão em constante movimento na RPA. A velocidade configurada é de 1/2000s, aproximadamente 0.0005 segundos (SANTIAGO-CINTRA, 2016a).



Figura 12. RPA eBee utilizada nas coletas de dados



Figura 13. (a) Câmera RGB e (b) Câmera NIR utilizadas pela RPA eBee

Cada câmera possui a capacidade de filtrar informações em diferentes regiões no espectro de luz. A câmera RGB captura dados nas bandas do RED

(vermelho), GREEN (verde) e BLUE (azul), enquanto a NIR captura informações na região do NIR (infravermelho próximo), RED (vermelho) e GREEN (verde). Apesar das câmeras capturarem informações na região do vermelho e do verde, cada uma considera comprimentos de onda diferentes, como pode ser observado na Figura 14 (SANTIAGO-CINTRA, 2016a; SANTIAGO-CINTRA, 2016b).

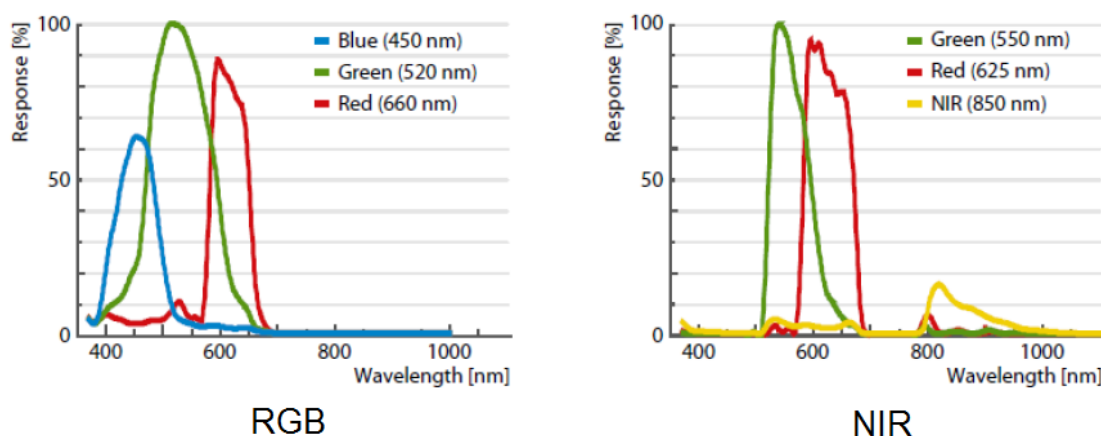


Figura 14. Comprimentos de onda filtrados pelas câmeras RGB e NIR
 Fonte: Adaptado de SANTIAGO-CINTRA, 2016a; SANTIAGO-CINTRA(2016b)

Além disso, as câmeras trabalham com uma intensidade de bits diferente. Na câmera RGB as imagens são feitas no padrão JPEG com 8 bits por canal, ou seja, os valores digitais que representam as cores de cada pixel podem variar de 0 a 255 em cada canal. Já na câmera NIR, as imagens são obtidas originalmente no formato CR2, e depois são convertidas no pré-processamento para o formato TIFF (mais preciso) de 16 bits, onde os valores digitais variam de 0 a 65.536 (SANTIAGO-CINTRA, 2016a).

O equipamento eBee vem acompanhado do software Emotion 2, que permite configuração prévia do plano de voo bem como sua alteração em tempo real (Figura 15). Com o software é possível definir a área da coleta de dados, a altitude, resolução das imagens e a sobreposição longitudinal e latitudinal da área a ser sobrevoada. O programa também permite realizar o monitoramento em tempo real do eBee através de informações de voo, tais como altitude, velocidade do vento, status de bateria, dentre outras.

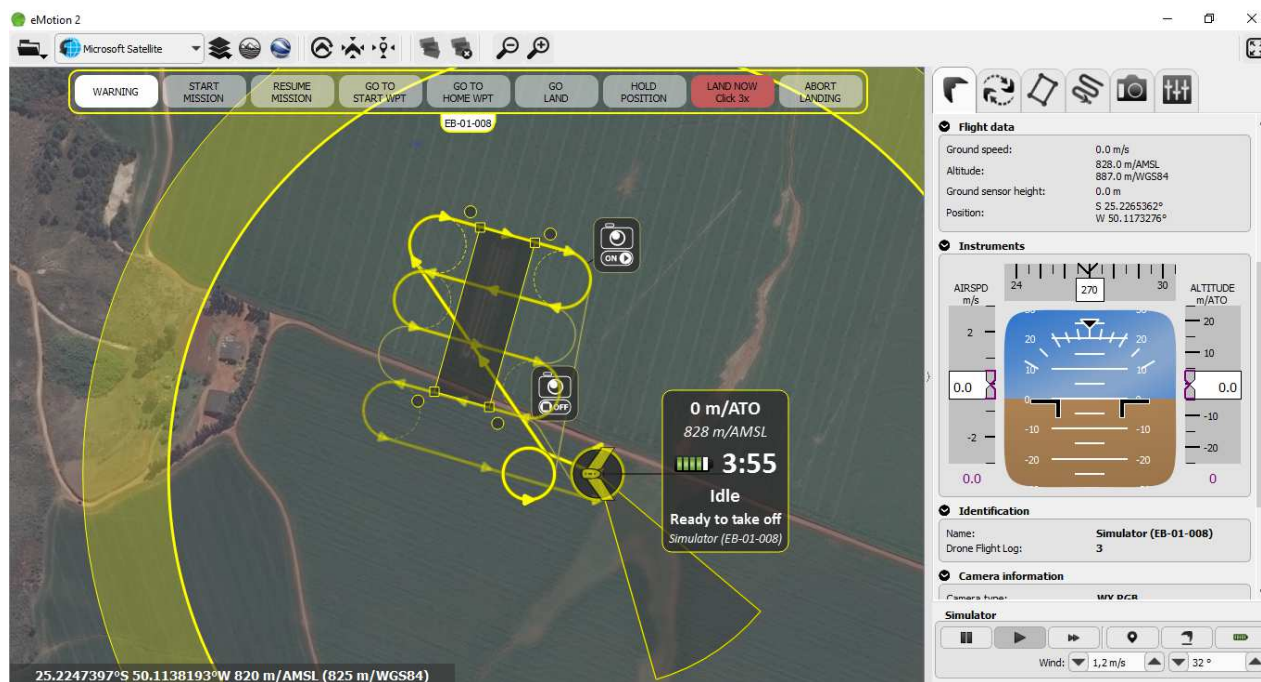


Figura 15. Tela do software Emotion 2 para controle da RPA eBee

O Emotion 2 é também responsável pela realização do pré-processamento das imagens, onde através do sistema inercial e do GPS embarcados na RPA, o software é capaz proporcionar as coordenadas geográficas de cada foto, além das conversões de formato citadas anteriormente.

4.3 COLETAS DE DADOS

Para a coleta de imagens, os voos foram realizados nas duas áreas de estudo entre os meses de agosto e outubro de 2016, considerando o período de emergência até o período de maior vigor das plantas.

Segundo Gómez-Candón et al. (2014), com o objetivo de evitar distorções geométricas nas imagens coletadas por RPAs, e também considerando que geralmente estes equipamentos voam a baixas altitudes, muitas imagens sobrepostas devem ser coletadas para que depois possam ser unidas e consolidadas em uma única imagem (conhecida por ortomosaico), que representa toda a área de estudo. Assim as imagens de ambas as áreas foram coletadas considerando uma sobreposição lateral de 60% e longitudinal de 75%. Foram feitas ao todo oito coletas tanto para a área da Fazenda Santa Cruz como da Fazenda Escola.

Em todas as datas foram obtidas imagens com resolução espacial de 3,4 centímetros por pixel, ou seja, cada pixel da imagem representa um quadrado de 3,4 por 3,4 centímetros da área real. Por ser a resolução mais alta possível de se alcançar com esse equipamento, é também a resolução que permite a extração da maior quantidade de informações nas imagens.

Como um dos objetivos era o de verificar a influência da resolução das imagens na estimativa da produtividade do trigo, em duas datas (08/10/2016 e 17/10/2016) foram realizadas coletas de imagens em diferentes resoluções no período de maior vigor do trigo, para que tal verificação fosse possível. Dessa maneira, nesses dias além de imagens de 3,4 cm/px foram obtidas imagens com resolução de 10 e 20 centímetros por pixel. Na Tabela 2 são apresentadas as datas de voo com as respectivas resoluções utilizadas, bem como a quantidade de dias após a semeadura (DAS), correspondentes a cada data.

Tabela 2. Datas dos voos realizados e resolução das imagens

Fazenda Santa Cruz			Fazenda Escola		
Data	DAS*	Resolução	Data	DAS*	Resolução
11/08/2016	49	3,4 cm/px	25/08/2016	55	3,4 cm/px
29/08/2016	67	3,4 cm/px	29/08/2016	59	3,4 cm/px
05/09/2016	74	3,4 cm/px	12/09/2016	73	3,4 cm/px
16/09/2016	85	3,4 cm/px	16/09/2016	76	3,4 cm/px
23/09/2016	92	3,4 cm/px	23/09/2016	83	3,4 cm/px
29/09/2016	98	3,4 cm/px	27/09/2016	87	3,4 cm/px
		3,4 cm/px			3,4 cm/px
08/10/2016	107	10 cm/px 20cm/px	08/10/2016	98	10 cm/px 20cm/px
		3,4 cm/px			3,4 cm/px
17/10/2016	116	10 cm/px 20cm/px	17/10/2016	107	10 cm/px 20cm/px

Como o eBee tem a possibilidade de voar somente com uma câmera por vez, para cada resolução foram necessários 2 voos, um para coleta de imagens RGB e outro para imagens NIR. A altitude de voo é calculada de forma automática pelo software Emotion 2 e varia de acordo com a câmera utilizada e a resolução espacial

escolhida. Na Tabela 3 são apresentadas as resoluções utilizadas e respectivas altitudes de voo, para cada câmera.

Tabela 3. Altitude de voo utilizada por resolução para cada câmera

Câmera RGB		Câmera NIR	
Resolução	Altitude de Voo	Resolução	Altitude de Voo
3,4 cm/px	120 m	3,4 cm/px	95 m
10 cm/px	350 m	10 cm/px	295 m
20 cm/px	700 m	20 cm/px	580 m

Procurou-se realizar todas as coletas em dias ensolarados e com pouca presença de nuvens, e os voos ocorreram sempre entre as 10 horas da manhã e 14 horas da tarde, período onde a radiação solar é mais intensa.

4.4 PROCESSAMENTO DAS IMAGENS

Para o processamento das imagens e geração dos ortomosaicos foi utilizado o software Pix4DMapper PRO, que pode ser adquirido a partir de uma licença para uso particular ou por contrato de serviço online.



Figura 16. Interface do software Pix4D

Por meio do pré-processamento das imagens realizado pelo software Emotion 2, um projeto para cada voo é criado para o processamento no Pix4D (PIX4D, 2016a). Dessa forma o Pix4D recebe todos os parâmetros de voo do momento em que as imagens foram coletadas, bem como os parâmetros de cada imagem, como orientação da foto e coordenada geográfica. A interface do software Pix4D é apresentada na Figura 16.

A partir disso, a primeira etapa foi o processamento inicial, onde três tarefas principais foram executadas. A primeira tarefa foi a geração de um relatório de qualidade, no qual pode ser visualizada uma prévia do ortomosaico (para verificação de possíveis falhas nas imagens), além de informações de posicionamento das imagens e geolocalização. A segunda tarefa realizada foi uma análise da correspondência (*matching*) das imagens para posterior geração do ortomosaico; a terceira tarefa consistiu na calibração das imagens (PIX4D, 2016a).

A segunda etapa no processamento das imagens pelo Pix4D tratou da geração de uma nuvem de pontos (Figura 17), utilizada na criação do ortomosaico. Essa nuvem é gerada com auxílio das técnicas de Ajuste de Feixes em Bloco (BBA – do inglês *Bundle Block Adjustment*) e de Triangulação Aérea (AAT – do inglês *Automatic Aerial Triangulation*), onde cada pixel da nuvem é associado às imagens originais da área, e possui uma coordenada geográfica associada (PIX4D, 2016a).

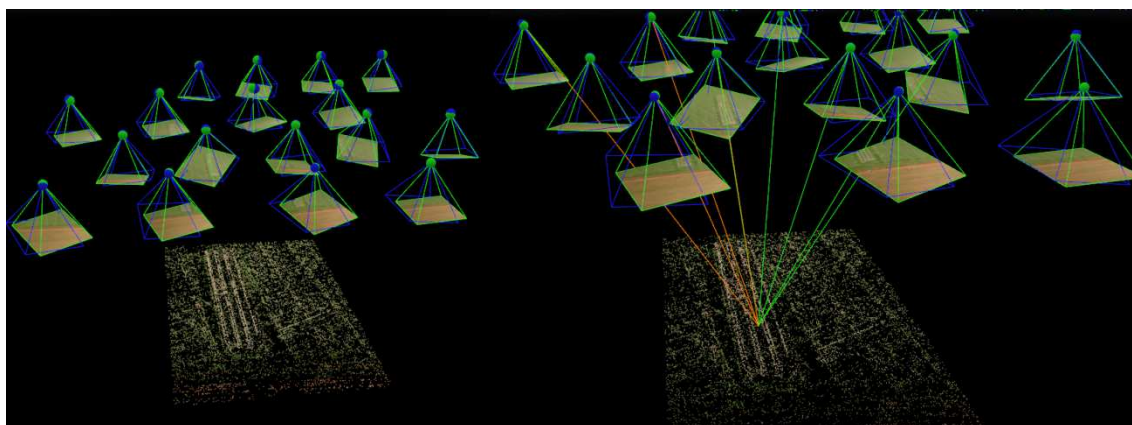


Figura 17. Nuvem de pontos gerada pelo Pix4D para geração do ortomosaico

A terceira e última etapa no processamento consistiu na geração de um Modelo Digital de Superfície (DSM – do inglês *Digital Surface Model*), que depois de combinado com a nuvem de pontos fornece o produto final do processamento com o PIX4D, que é o ortomosaico, o qual equivale a uma única imagem da área

sobrevoada. Além do ortomosaico, nessa etapa também é gerado o mapa de refletância (PIX4D, 2016a).

Tanto o mapa de refletância quanto o ortomosaico são obtidos com base no DSM, através da média ponderada dos pixels das imagens originais. No caso do ortomosaico, é aplicado balanceamento de cores, ou seja, o Pix4D procura ajustar a intensidade das cores dos pixels de forma que combinem melhor na imagem resultante, com objetivo de gerar um produto mais agradável visualmente (PIX4D, 2016b).

Já no mapa de refletância, nenhum balanceamento é aplicado, e a média do valor digital de cada pixel é calculada levando em conta outras informações como luz incidente, abertura e velocidade do obturador da câmera. O objetivo é produzir um mapa onde o valor de cada pixel indique fielmente a refletância da área (PIX4D, 2016b). Sendo assim, neste trabalho foram utilizados somente os mapas de refletância. Na Figura 18 pode-se perceber a diferença entre o ortomosaico (Figura 18a) e o mapa de refletância (Figura 18b) gerados para a área experimental da Fazenda Santa Cruz.

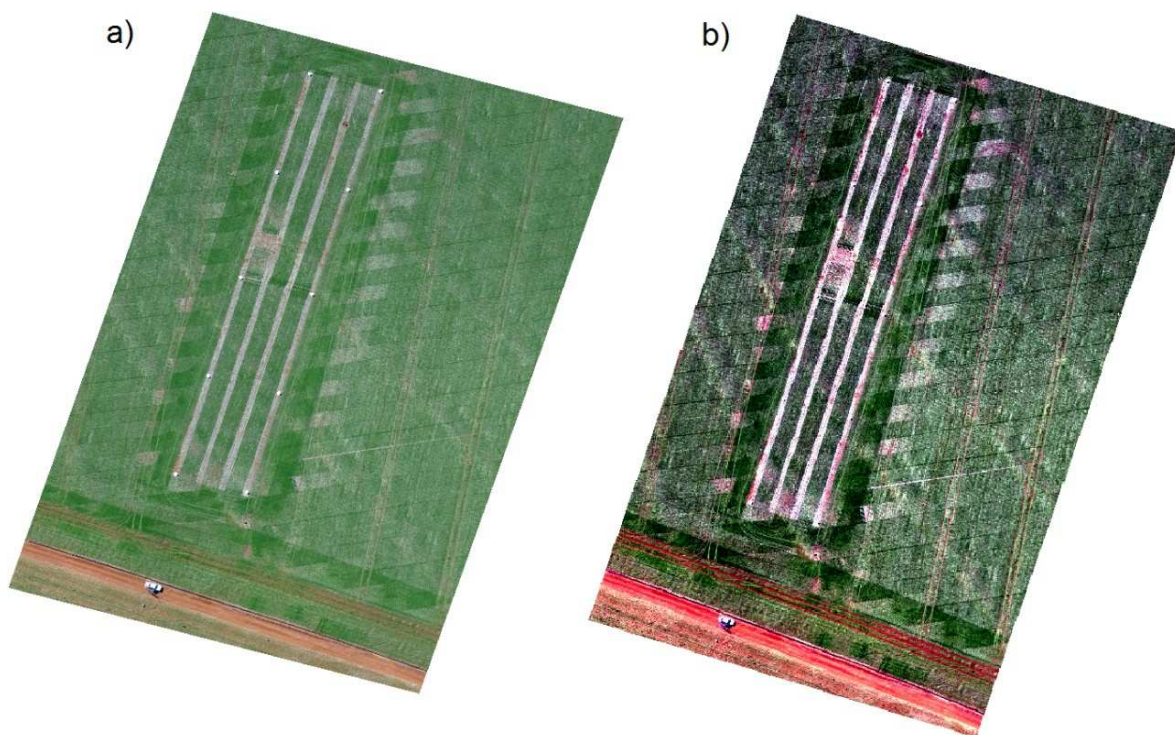


Figura 18. Diferença entre o ortomosaico (a) e o mapa de refletância (b)

4.5 CORREÇÃO RADIOMÉTRICA DAS IMAGENS

A correção radiométrica é um processo importante, conforme mencionado na seção 3.4. O software Pix4D fornece algumas opções de calibração radiométrica para as imagens, que levam em consideração a iluminação e os efeitos induzidos pelos sensores (câmeras). As opções disponíveis são as correções *Camera Only* (somente câmera), *Camera and Sun Irradiance* (câmera e irradiância solar) e a opção *Camera, Sun Irradiance and Sun Angle* (câmera, irradiância solar e ângulo do sol) (PIX4D, 2016a).

Na opção *Camera Only*, são realizadas correções baseadas somente em parâmetros da câmera, como *vignetting*, *dark current*, ISO, *aperture*, *exposure time* e *spectral overlap*. A correção de *vignetting* remove o efeito causado devido ao formato esférico da lente da câmera, que faz com que as bordas da imagem não recebam a mesma quantidade de luz que o centro da imagem. O problema de *dark current* é quando há acúmulo de elétrons nos sensores devido à temperatura, podendo gerar ruídos na imagem (PIX4D, 2016c).

A correção de *spectral overlap* (sobreposição espectral) ocorre em câmeras que utilizam filtros para captar dados em outras frequências, como a do infravermelho. Para isso, o filtro é aplicado em uma das bandas originais da câmera, como por exemplo na câmera Canon S110 NIR, que utiliza um filtro na banda do azul para captar o infravermelho. O resultado disso é que a banda cobre uma faixa grande no espectro, podendo ocasionar contaminações entre as bandas (PIX4D, 2016c).

No geral, as câmeras precisam ajustar continuamente suas configurações devido à diferença de brilho em cada imagem, ou seja, uma imagem escura não indica necessariamente que esta tenha menos luz, mas pode ser que o tempo de exposição tenha sido menor. Se as imagens são fruto de sensores lineares, pode-se combinar imagens com diferentes ISO, *aperture* (abertura) e *exposure time* (tempo de exposição) apenas corrigindo tais efeitos (PIX4D, 2016c).

Demais correções, que levam em consideração a irradiância solar, necessitam de um alvo de calibração radiométrica com refletância conhecida. Essa refletância é utilizada como parâmetro para correção das imagens, com intuito de garantir valores de refletância mais precisos (PIX4D, 2016c). Por não se ter os

equipamentos necessários (alvos e espectrômetro) para tal calibração, neste trabalho foi utilizada somente a opção de correção *camera only* do Pix4D.

4.6 GEORREFERENCIAMENTO DOS MOSAICOS

O processamento fotogramétrico de imagens obtidas por RPA pode ser realizado com as coordenadas geográficas centrais de cada imagem, dadas pelo sistema GNSS (do inglês *Global Navigation Satellite Systems*). Porém, para garantir a precisão do georreferenciamento após o processo de geração dos mosaicos, é conveniente o uso de pontos de apoio, também conhecidos por GCPs (do inglês *Ground Control Points*) (GÓMEZ-CANDÓN et al., 2014). Para a correção geométrica das imagens coletadas por RPA, coordenadas mais precisas de objetos fixos ao nível do solo são necessárias, porém tais objetos devem ser reconhecíveis nas imagens.

Uma metodologia comum é a utilização de GCPs como apoio no processo de geração do mosaico, cujo produto final é uma imagem geograficamente ajustada. Liba e Berg-Jurgens (2015) usam esta metodologia com imagens de resolução espacial de 5cm/pixel obtidas por meio de uma RPA utilizando a ferramenta Pix4D para geração dos mosaicos. A acurácia do georreferenciamento dos mosaicos utilizando GCPs foi de 0,132m, e de 1,417m sem o uso de GCPs.

Outra metodologia é a utilização de GCPs para correção de georreferenciamento após a geração do mosaico, utilizando ferramentas que permitam o georreferenciamento de imagens, como por exemplo os sistemas de informação geográfica (SIG), como o Quantum GIS, também conhecido como QGIS. O QGIS é um software livre (*open source*) e multiplataforma que fornece várias ferramentas para visualização, edição e análise de dados georreferenciados (QGIS, 2016).

Com a finalidade de verificar a precisão de georreferenciamento dos mosaicos gerados com e sem o auxílio de GCPs, foram realizados alguns testes com base nas imagens de 3,4cm/px obtidas no dia 11/08/2016 na área experimental da Fazenda Santa Cruz.

Para que fosse possível visualizar os GCPs nas imagens, foram confeccionados marcadores (Figura 19) com fundo branco medindo 60cm x 60cm. Tais marcadores foram posicionados com auxílio de estacas a uma altura de 50cm

do solo, tendo em vista o acompanhamento no crescimento das culturas. O marcador possui um X feito com fita de marcação, para auxiliar na identificação do ponto central, onde é feita a amarração com a coordenada geográfica.



Figura 19. Marcadores utilizados como GCPs para a correção de georreferenciamento

Para a coleta da coordenada geográfica de cada um dos GCPs foi utilizado um receptor geodésico Trimble R4 (Figura 20a) com coordenadas pós-processadas com base nas estações da Rede Brasileira de Monitoramento Contínuo dos Sistemas GNSS (RBMC) de Curitiba-PR e Guarapuava-PR. Para isso, foram definidos 10 pontos (Figura 20b) de controle ao longo da área de estudo.



Figura 20. (a) GPS Trimble R4 e (b) a disposição dos GCPs na área experimental

Foram gerados três mosaicos pelo software Pix4D, o primeiro sem a utilização de GCPs (Pix0), somente com base na coordenada geográfica central das imagens originais, o segundo utilizando cinco GCPs (Pix5) e o terceiro dez GCPs (Pix10). Além disso, outros 2 mosaicos foram gerados através do QGIS, o quarto (Qgis5) e quinto (Qgis10), com base em uma correção geométrica no primeiro mosaico (Pix0) utilizando 5 e 10 GCPs, respectivamente.

Para a correção realizada através do QGIS, foi utilizado o método de transformação “polinomial 1”, pois preserva a colinearidade e permite apenas o escalonamento, translação e rotação da imagem. O método de reamostragem selecionado foi do vizinho mais próximo, uma vez que não suaviza a imagem e mantém suas estatísticas (QGIS, 2016).

A escolha da quantidade de GCPs se deu devido à recomendação do Pix4D, em que um número mínimo de 5 pontos de controle é recomendado, e de 5 a 10 pontos de controle são geralmente suficientes, mesmo para grandes projetos. Mais pontos de controle não contribuem de forma significativa para aumentar a precisão (PIX4D, 2016a).

Com base nas coordenadas dos GCPs medidas com o receptor geodésico, foi calculada a distância das coordenadas dos GCPs para cada um dos mosaicos gerados. Foi calculado também o erro médio quadrático (RMSE, do inglês – *Root Mean Square Error*) como indicativo de diferença entre as distâncias. Além disso, para saber se as distâncias medidas possuem uma diferença significativa entre os mosaicos, foi utilizado o teste de Tukey.

4.7 EXTRAÇÃO DE DADOS DAS IMAGENS

Para que se pudesse extrair os dados dos mapas de refletância, era necessário identificar cada uma das parcelas nas imagens. Para isso, nos voos realizados no dia 29/08/2016 foram utilizados marcadores brancos posicionados nas bordas das áreas, feitos de um tecido emborrachado (medindo 40 x 40 cm), para demarcar a separação das parcelas nas áreas da Fazenda Santa Cruz (Figura 21a) e Fazenda Escola (Figura 21b).



Figura 21. Pontos brancos para demarcação da divisão de parcelas da Fazenda Santa Cruz (a) e da Fazenda Escola (b)

A partir disso, utilizando o software Quantum GIS, foi possível traçar linhas de separação das parcelas nas áreas da Fazenda Santa Cruz (Figura 22) e da Fazenda Escola (Figura 23), com base nos marcadores (pontos brancos) visualizados nas imagens.



Figura 22. Linhas de separação de parcelas da Fazenda Santa Cruz

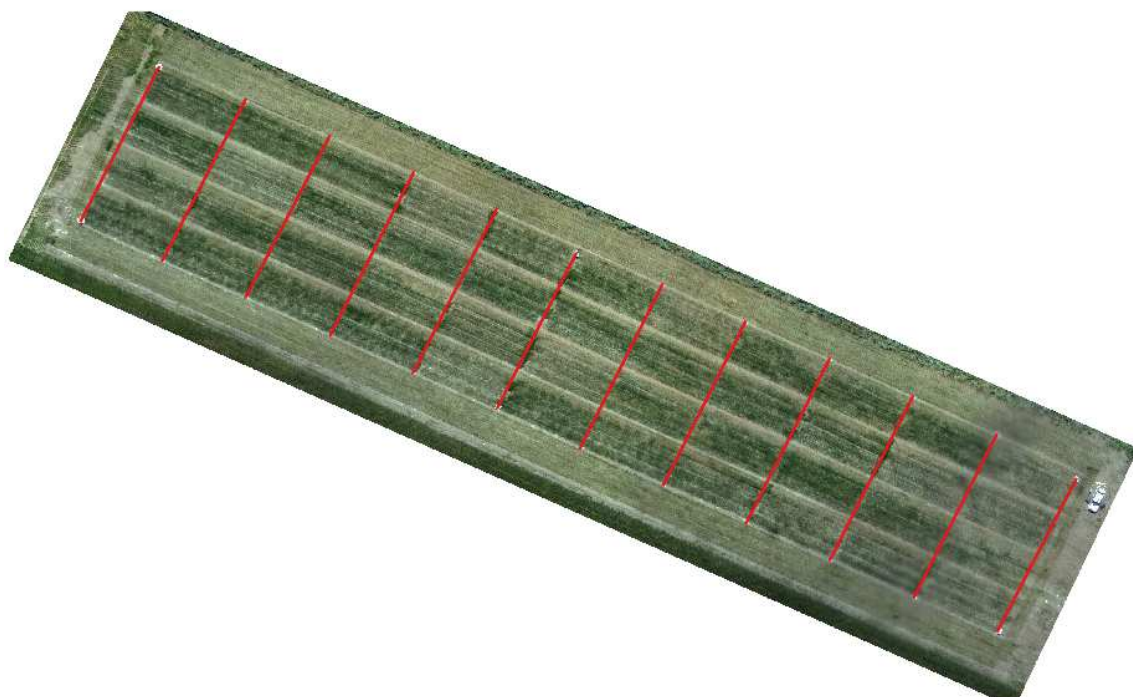


Figura 23. Linhas de separação de parcelas da Fazenda Escola

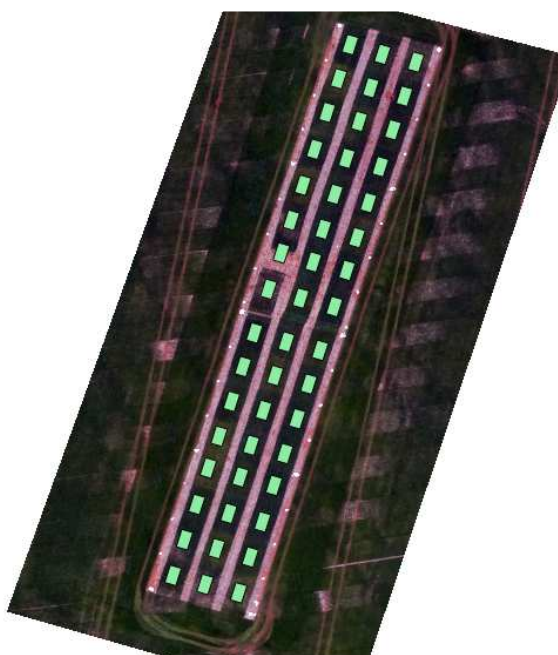


Figura 24. Polígonos representando cada parcela da Fazenda Santa Cruz

Após a realização da marcação das linhas de separação das parcelas em ambas as áreas, foram criados os polígonos para cada uma das parcelas, tanto da Fazenda Santa Cruz (Figura 24) como da Fazenda Escola (Figura 25). Os polígonos foram gerados com 4 metros de comprimento por 2,5 metros de largura (totalizando 10m^2) e posicionados no centro de cada parcela. Isso se deu pelo fato de a colheita

ser realizada somente nesse espaço dentro de cada parcela, e não em sua área total.



Figura 25. Polígonos representando cada parcela da Fazenda Escola

Depois de realizadas estas tarefas, foram criados pontos no interior de cada polígono, onde para cada ponto seriam extraídos os números digitais referentes às bandas R,G e B (da câmera RGB) e às bandas R, G e NIR (da câmera NIR). Foi feita uma estimativa da quantidade de pontos necessários para cobrir 100% dos pixels em cada parcela, calculada pela razão entre o tamanho da parcela ($10m^2$) e a área de cada pixel, levando em conta a resolução das imagens (3,4cm/px, 10cm/px e 20cm/px).

Com o objetivo de verificar a influência da quantidade de pontos na estimativa de produtividade, além da quantidade de pontos necessários para cobrir 100% de cada parcela, foram consideradas também as proporções de 70%, 40%, 20% e 10%, conforme pode ser visualizado na Tabela 4.

A quantidade de pontos resultantes por imagem pode ser observada na Tabela 5, considerando as 48 parcelas da Fazenda Santa Cruz e as 44 parcelas da Fazenda Escola, com base no percentual de cobertura e na resolução das imagens. Essa quantidade de pontos refere-se ao tamanho (número de registros) presentes nas bases de dados para posterior mineração de dados.

Tabela 4. Número de pontos de acordo com a taxa de cobertura de cada parcela de acordo com a resolução da imagem

Cobertura	3.4 cm/px	10 cm/px	20cm/px
100%	8.650	1.000	250
70%	6.055	700	175
40%	3.460	400	100
20%	1.730	200	50
10%	865	100	25

Tabela 5. Número total de pontos de acordo com a taxa de cobertura por área experimental com base na resolução das imagens

	Fazenda Santa Cruz			Fazenda Escola		
	3,4cm/px	10cm/px	20cm/px	3,4cm/px	10cm/px	20cm/px
100%	415.200	48.000	12.000	380.600	44.000	11.000
70%	290.640	33.600	8.400	266.420	30.800	7.700
40%	166.080	19.200	4.800	152.240	17.600	4.400
20%	83.040	9.600	2.400	76.120	8.800	2.200
10%	41.500	4.800	1.200	38.060	4.400	1.100

A geração desses pontos foi feita pelo software QGIS, utilizando a ferramenta “geração de pontos aleatórios no interior de polígonos”. Com tal ferramenta pode-se definir a quantidade de pontos que será gerada no interior de cada parcela, de acordo com a Tabela 5, sendo que o posicionamento de cada ponto é feito de forma aleatória. Na Figura 26, Figura 27 e Figura 28 pode-se visualizar os pontos em diferentes proporções em uma imagem de 3,4cm, 10cm e 20cm, respectivamente.

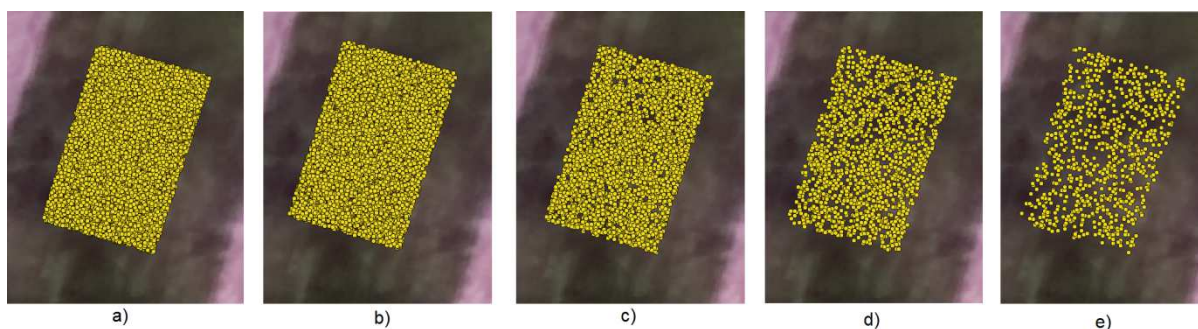


Figura 26. Visualização dos pontos de uma parcela em uma imagem de 3,4cm/px Considerando 100% (a), 70% (b), 40% (c), 20% (d) e 10% (e) de cobertura

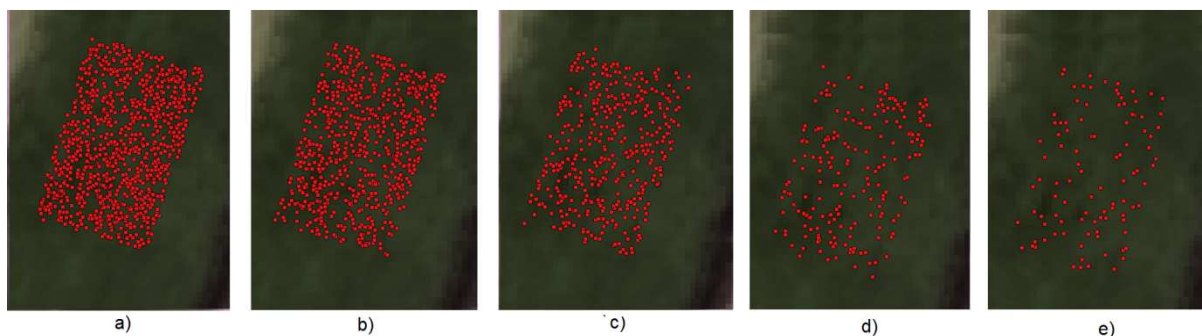


Figura 27. Visualização dos pontos de uma parcela em uma imagem de 10cm/px
Considerando 100% (a), 70% (b), 40% (c), 20% (d) e 10% (e) de cobertura

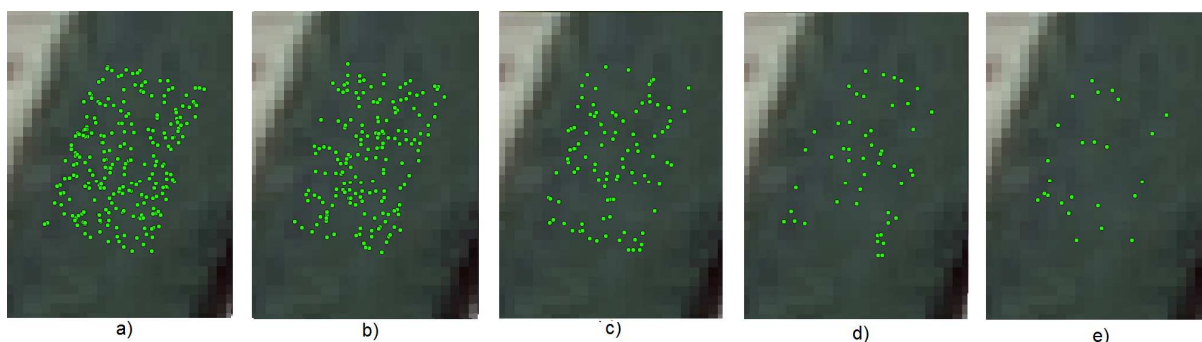


Figura 28. Visualização dos pontos de uma parcela em uma imagem de 20cm/px
Considerando 100% (a), 70% (b), 40% (c), 20% (d) e 10% (e) de cobertura

Os valores digitais foram extraídos com auxílio da ferramenta *Point Sampling Tool* (ferramenta de amostragem de pontos) presente no QGIS. Dessa forma, para cada um dos voos realizados foram extraídos os valores de refletância, das imagens obtidas pela câmera RGB, NIR, e do NDVI, calculado através da aritmética de banda das imagens da câmera NIR, conforme mostrado na Tabela 6. Em cada linha da Tabela 6 são mostrados os valores extraídos por ponto da base de dados. As colunas NIR_R, NIR_G e NIR apresentam os valores referentes às três bandas das imagens da câmera NIR e as colunas RGB_R, RGB_G e RGB_B referem-se às três bandas da câmera RGB.

Para cada uma das datas, foram geradas bases de dados com a quantidade de pontos para cobertura de 100%, 70%, 40%, 20% e 10%, ou seja, para cada data foram obtidas cinco bases de dados. Nas datas de 08/10/2016 e 17/10/2016 foram geradas uma quantidade maior de bases, uma vez que em tais datas foram feitas coletas não só na resolução de 3,4cm/px, mas também em 10cm/px e 20cm/px.

Tabela 6. Exemplo de valores extraídos a partir do *Point Sampling Tool*

Parcela	NDVI	RGB_R	RGB_G	RGB_B	NIR_R	NIR_G	NIR
Parcela01	0,47956	1616,94043	2357,36743	1496,96082	14302,70703	16576,34766	40660,73438
Parcela01	0,46304	1737,62427	2499,43994	1637,93054	15764,42383	18228,97656	42952,65625
Parcela01	0,48869	1647,87073	2422,47705	1550,94812	14643,59570	17475,74219	42635,07031
Parcela01	0,49678	1542,22058	2273,51465	1445,49988	13824,41016	16648,04883	41119,29297
Parcela01	0,48068	1795,83582	2537,70557	1708,34058	14554,08691	16951,58008	41496,88672
Parcela01	0,49016	1505,70410	2271,47778	1408,15967	13967,84082	16160,47168	40824,95313
Parcela01	0,46922	1682,14001	2424,01514	1596,58936	15392,28516	18276,70313	42605,94531
Parcela01	0,49834	1484,58777	2195,21997	1374,21570	13036,05273	15485,34082	38935,48047
Parcela01	0,48665	1136,39111	1894,78369	1030,63623	13574,87598	16162,37793	39547,83203
Parcela01	0,44357	1899,27478	2640,16675	1801,18762	16290,90723	18888,72070	42264,23438
Parcela01	0,46397	1983,04797	2728,08228	1894,70923	16087,25195	18765,96289	43936,76563
Parcela01	0,47778	1696,91589	2442,95947	1596,92969	15500,45605	18283,92773	43863,40625
Parcela01	0,48345	1687,89465	2414,33008	1583,03723	13937,72461	15833,36914	40027,45313
Parcela01	0,49690	1543,60706	2319,58325	1447,12830	14044,51953	16291,92188	41787,65234

4.8 MINERAÇÃO DOS DADOS

Para mineração das bases de dados, foi utilizado o software WEKA (versão 3.8), que é uma coleção de algoritmos de aprendizado de máquina aplicada a tarefas de mineração de dados. O software é implementado na linguagem de programação Java e permite que os algoritmos sejam aplicados diretamente às bases de dados, ou importados através de código Java. O WEKA também fornece ferramentas de pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização (WEKA, 2016).

Para geração dos modelos preditivos foi escolhido o algoritmo SMOReg. SMOReg é um algoritmo baseado em SVR, implementando um algoritmo de otimização mínima sequencial para aprender e gerar um modelo de regressão. O algoritmo normaliza os dados por padrão, ou seja, transforma os valores da base, dimensionando-os para que sejam representados entre zero e um (WITTEN et al., 2016).

A execução da análise de cada uma das bases (considerando as proporções apresentadas na seção anterior) foi repetida por três vezes, contemplando cada uma das câmeras (RGB e NIR) e o NDVI, para estimativa da produtividade do trigo. Tal separação foi feita para que o potencial de predição de cada câmera fosse avaliado, assim como do NDVI.

5 RESULTADOS E DISCUSSÃO

No presente capítulo, os resultados e discussões são apresentados em subseções. A primeira mostra os resultados das correções de georreferenciamento realizadas para os ortomosaicos gerados. A segunda apresenta o comportamento dos valores de NDVI e os valores digitais de refletância para cada banda no decorrer de cada coleta, considerando os comprimentos de onda utilizados. A terceira subseção mostra os resultados da mineração de dados, bem como os modelos gerados.

5.1 MOSAICOS GERADOS PARA EXTRAÇÃO DOS DADOS

A construção dos mosaicos, a partir dos quais foram extraídos os dados para análise e geração dos modelos preditivos de produtividade, foi realizada de acordo com a metodologia explicada na seção 4.5. Os pontos em amarelo na Figura 29 representam as coordenadas exatas dos GCPs (coletadas com o GPS em campo), e como pode ser observado, há um erro entre tal coordenada e seu respectivo marcador na imagem. Sendo assim, para todos os tratamentos foi medida a distância de erro para cada um dos GCPs (Tabela 7).



Figura 29. Demonstração do erro de georreferenciamento (tratamento Pix0)

Tabela 7. Distâncias em metros entre GCPs e a coordenada exata em solo e RMSE para cada mosaico

Ponto	PIX0	PIX5	PIX10	QGIS5	QGIS10
P01	2,039	1,027	0,67	0,275	0,277
P02	1,959	1,042	0,661	0,075	0,71
P03	2,2	1,226	0,946	0,6	0,367
P04	2,57	0,973	0,71	1,09	0,675
P05	1,539	0,395	0,43	0,253	0,25
P06	1,373	0,377	0,614	0,366	0,361
P07	1,74	0,622	0,808	0,297	0,111
P08	2,261	0,589	0,778	0,264	0,05
P09	2,41	0,542	0,744	0,26	0,107
P10	2,717	0,56	0,842	0,175	0,213
RMSE	2,122	0,789	0,732	0,456	0,378

O mosaico que obteve o menor RMSE (0,378m) foi o QGIS10. O mosaico que apresentou maior RMSE (2,122m) em relação a distância dos pontos GCP foi o mosaico PIX0.

Liba e Berg-Jurgens (2015) utilizaram imagens de RPA e 9 GCPs para a geração de um mosaico com a ferramenta Pix4D e compararam com um mosaico gerado sem o uso de GCPs, encontrando o RMSE de 0,132m e 1,417m para os mosaicos com e sem os GCPs, respectivamente. Neste estudo, os mosaicos gerados pelo Pix4D sem a utilização de GCPs demonstraram um RMSE de 2,122m, e os mosaicos gerados pelo mesmo método com 5 e 10 GCPs obtiveram um RMSE de 0,789m e 0,732m, respectivamente.

O resultado final pode sofrer influência de diversas variáveis, como a resolução espacial das imagens, equipamentos utilizados para coleta de imagens e de coordenadas dos GCPs, variáveis climáticas no momento das coletas, além do próprio software utilizado. Isso pode explicar a diferença de valores RMSE obtidos nesse trabalho em relação ao de Liba e Berg-Jurgens (2015). Porém, ainda assim a proporção das diferenças do RMSE entre os tratamentos (sem GCP e com GCP) são relativamente próximas nos dois trabalhos.

Para verificação da real diferença entre os resultados, foi aplicado o teste de Tukey, o qual foi conduzido num delineamento inteiramente casualizado, ao nível 5% de significância, com cinco tratamentos e dez repetições.

Conforme pode ser observado na Tabela 8, houve diferença significativa entre os mosaicos PIX0 em relação aos demais. Os mosaicos PIX5 e PIX10 não

apresentaram diferença significativa entre eles. Não é possível estabelecer diferença significativa do tratamento QGIS5 dos demais tratamentos, além do PIX0.

Tabela 8. Resultados do teste de Tukey na verificação da diferença entre as distâncias dos tratamentos¹

Tratamentos	Médias
PIX0	2,08 ^A
PIX5	0,74 ^B
PIX10	0,72 ^B
QGIS5	0,38 ^{BC}
QGIS10	0,35 ^C

¹Letras iguais indicam médias iguais pelo teste de Tukey (p<0.05)

Além de possuir menor erro, o processo de correção do mosaico pelo QGIS (tratamento QGIS5 e QGIS10) é menos trabalhoso quando comparado com a geração dos mosaicos corrigidos pelo Pix4D (tratamentos PIX5 e PIX10), tendo em vista que no Pix4D há a necessidade de demarcação de cada um dos GPCs em todas as imagens originais antes da geração do mosaico.

A partir desses resultados, todos os demais mosaicos foram gerados através do software Pix4D sem o auxílio de GCPs, com posterior correção através do Quantum GIS. Para correção dos mosaicos da Fazenda Santa Cruz foram utilizados 10 GCPs e para Fazenda Escola 9 GCPs.

5.2 COMPORTAMENTO DAS BASES DE DADOS

Na Fazenda Santa Cruz os maiores valores médios de NDVI (0,79) foram observados quando a cultura apresentava 67 DAS, enquanto os demais valores se mantiveram próximos (entre 0,50 e 0,36) nas coletas entre 74 DAS e 116 DAS (Figura 30a).

Na Fazenda Escola (Figura 30b) os valores de NDVI não ultrapassaram 0,317 (76 DAS), apresentando valores muito próximos em todos os períodos de coleta. Considerando os períodos entre 85 DAS e 116 DAS. No geral, os valores de NDVI para as duas áreas se comportaram de maneira diferente, o que pode ser explicado pela diferença física entre as duas áreas e a diferença na cultivar de trigo utilizado.

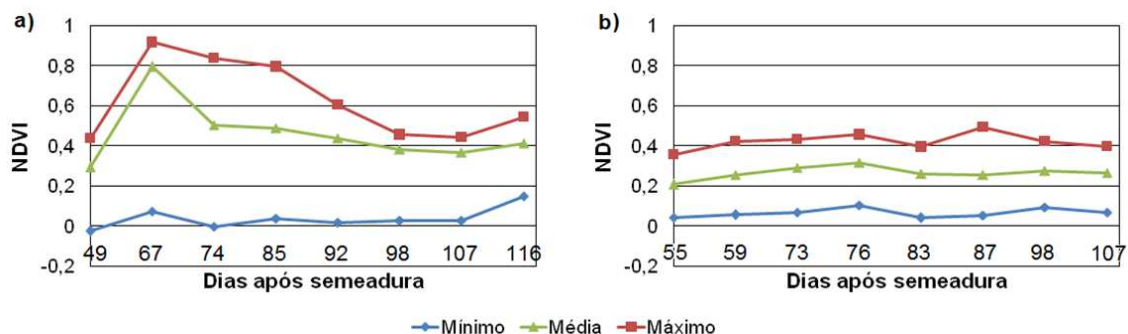


Figura 30. Comportamento dos valores de NDVI ao longo das coletas. (a) Fazenda Santa Cruz e (b) Fazenda Escola

Em relação aos valores de refletância nos canais R, G e B, obtidos com a câmera RGB, pode-se perceber que o comportamento ao longo das coletas foi similar, em que os maiores valores médios foram obtidos com 49 DAS e 92 DAS (Figura 31). Para a Fazenda Escola, os maiores valores médios foram observados com 55 e 83 DAS (Figura 32).

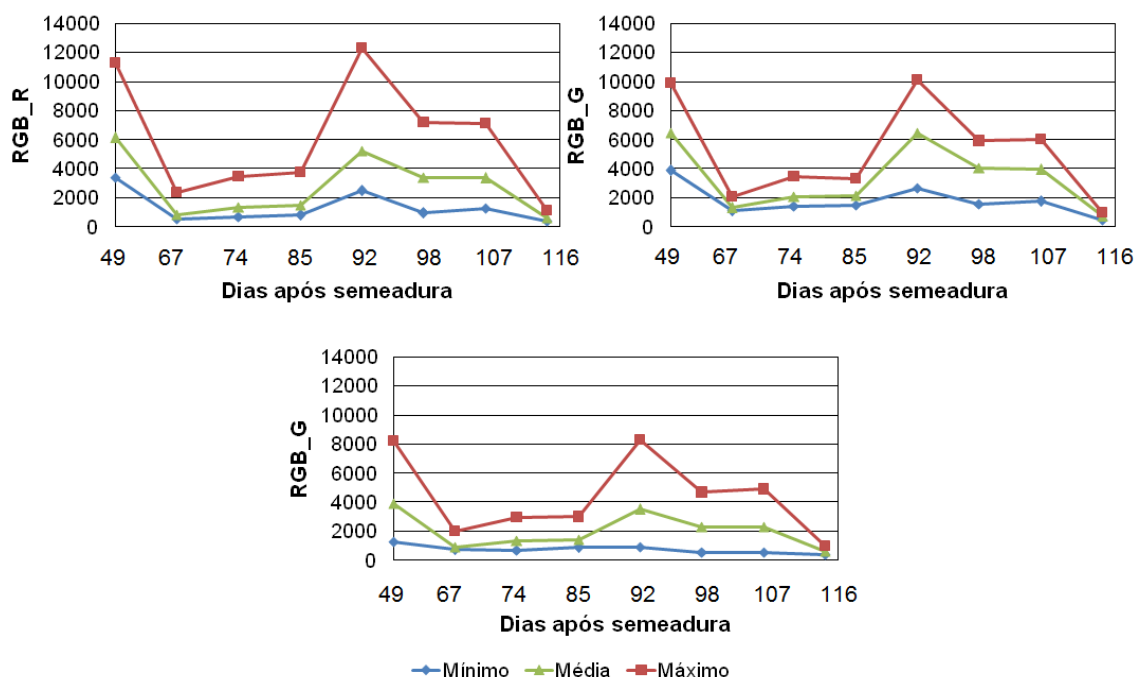


Figura 31. Comportamento dos valores digitais de R, G e B ao longo das coletas, obtidos com a câmera RGB para Fazenda Santa Cruz

Outro fato que pode ser observado ao analisar os dados dos canais RGB, é que os valores encontrados nas imagens da Fazenda Escola são no geral maiores que os valores da Fazenda Santa Cruz, considerando os mesmos períodos de coleta (ou períodos muito próximos). Tal fator pode ser explicado pela diferença nas

condições no dia da coleta, uma vez que as coletas nos mesmos períodos foram realizadas em dias diferentes, conforme mostrado na Tabela 2 (seção 4.3).

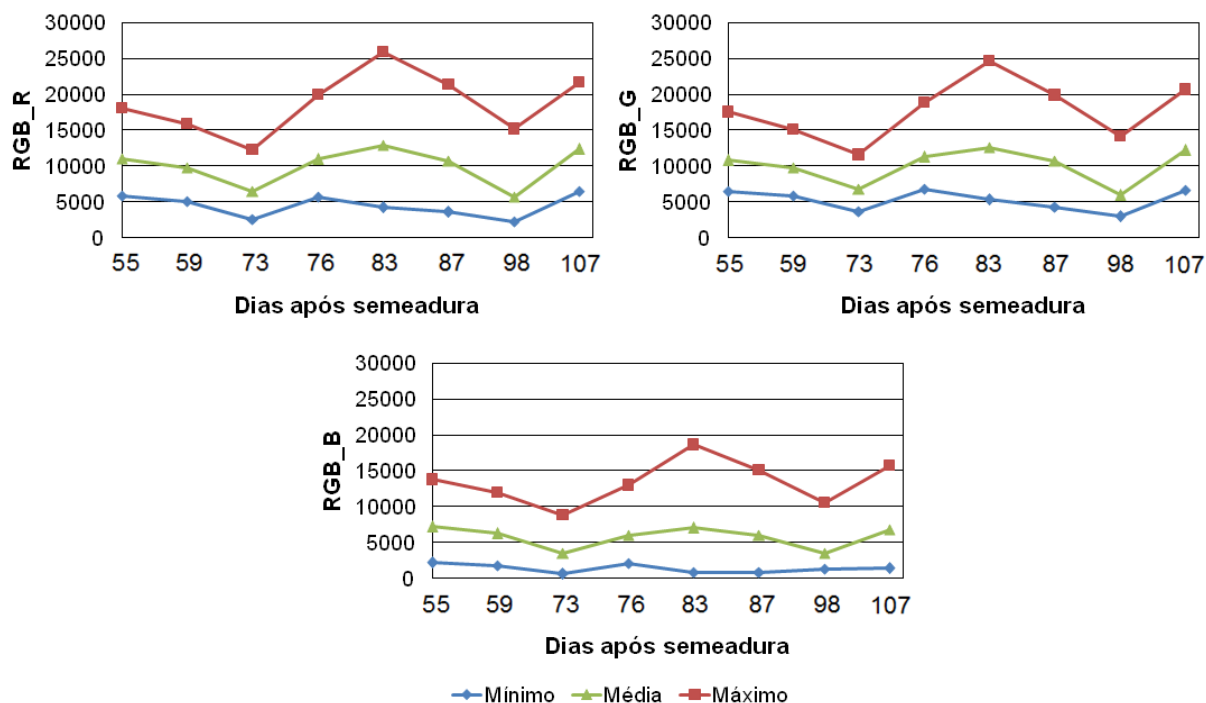


Figura 32. Comportamento dos valores digitais de R, G e B ao longo das coletas, obtidos com a câmera RGB para Fazenda Escola

Ao comparar os valores digitais para o canal R, obtidos pela câmera NIR (Figura 33a) e pela câmera RGB (Figura 33b), pode-se perceber que a distribuição dos dados é parecida, com valores altos no período de 49 DAS e 92 DAS, e valores mais baixos em 67 DAS e 116 DAS, assim como para o canal G (Figura 34).

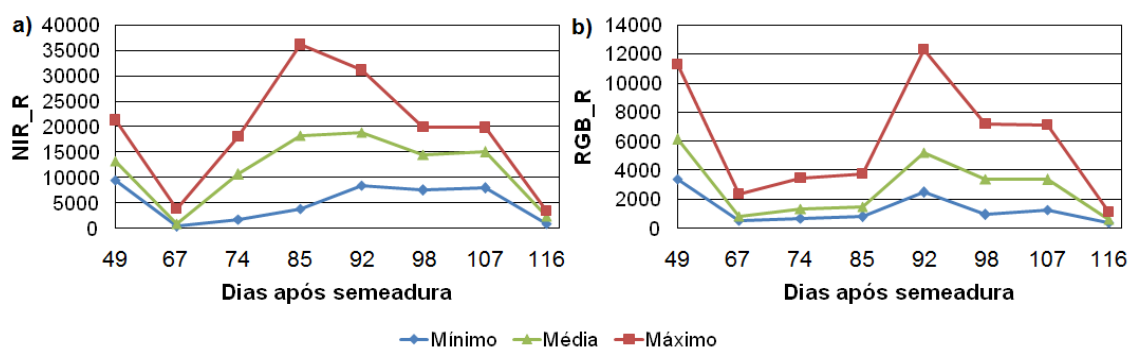


Figura 33. Comparativo de comportamento dos valores no canal R (Red) obtidos com a câmera (a) NIR e com a câmera (b) RGB, para Fazenda Santa Cruz

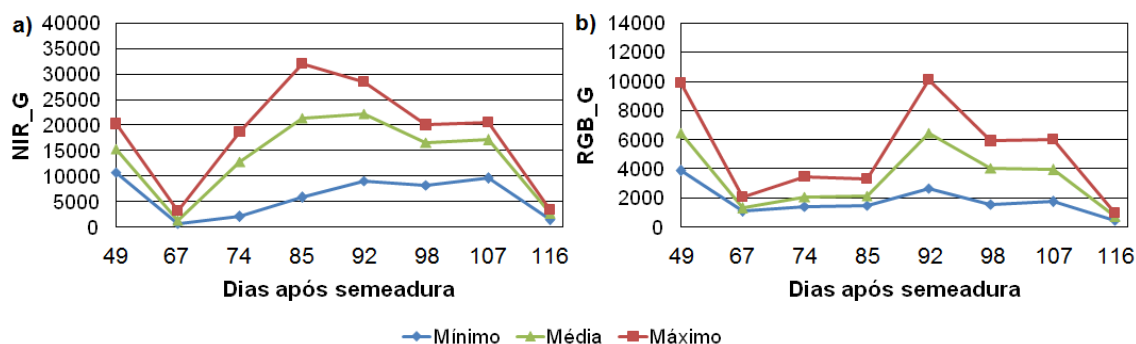


Figura 34. Comparativo de comportamento dos valores no canal G (Green) obtidos com a câmera NIR (a) e com a câmera RGB (b), para Fazenda Santa Cruz

O mesmo pode ser observado para os dados de dos canais R e G da Fazenda Escola, quando comparadas as duas câmeras (Figura 35 e Figura 36). Percebe-se uma diferença de leitura em relação ao período de 107 DAS: na câmera NIR, os valores vêm decrescendo e na câmera RGB, voltam a subir. Tal diferença pode ter sido causada pelas condições climáticas e de voo na hora de coleta, uma vez que a área experimental da Fazenda Escola é maior, levando a um tempo maior entre a coleta com a câmera RGB, a troca da câmera e o voo com a câmera NIR, possibilitando mudanças no ambiente.

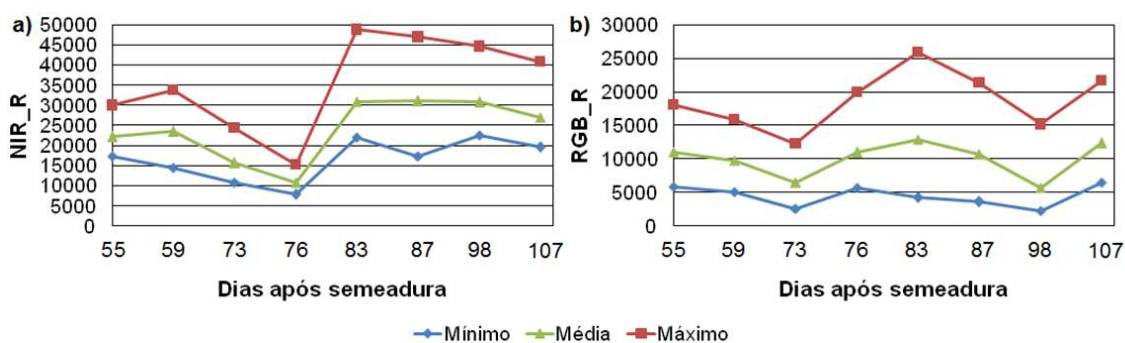


Figura 35. Comparativo de comportamento dos valores no canal R (Red) obtidos com a câmera NIR (a) e com a câmera RGB (b), para Fazenda Escola

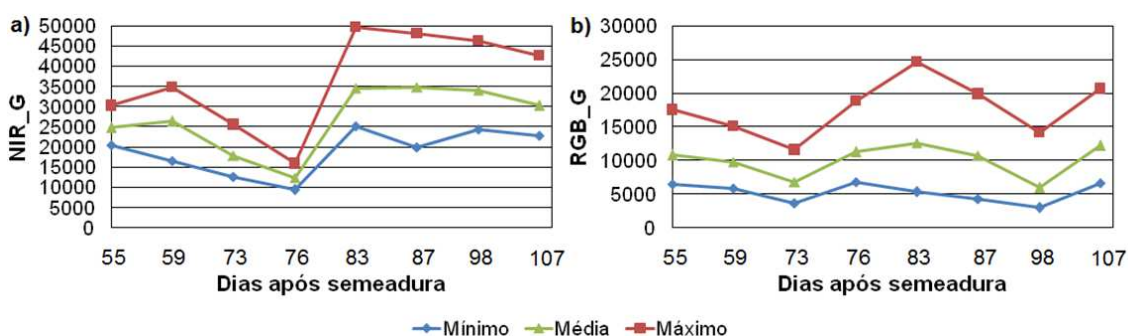


Figura 36. Comparativo de comportamento dos valores no canal G (Green) obtidos com a câmera NIR (a) e com a câmera RGB (b), para Fazenda Escola

Para o canal NIR, no geral, os valores digitais são maiores quando comparados aos outros canais, devido à alta reflexão das folhas nesse canal, como mostrado por Bernardi et al. (2014). Na Fazenda Santa Cruz, os valores médios mais altos (52861,0) foram observados em 85 DAS, enquanto na Fazenda Escola valores médios de refletância acima dos 52 mil foram observados no período entre 83 e 98 DAS, conforme mostrado na Figura 37.

Ainda, percebe-se que os valores médios de NIR nas duas áreas foram diferentes mesmo em períodos iguais ou muito próximos. Isso pode ter sido causado pela própria diferença entre as duas áreas, a diferença entre a cultivar de trigo semeada nas duas áreas, e também nas condições climáticas e de voo no momento das coletas.

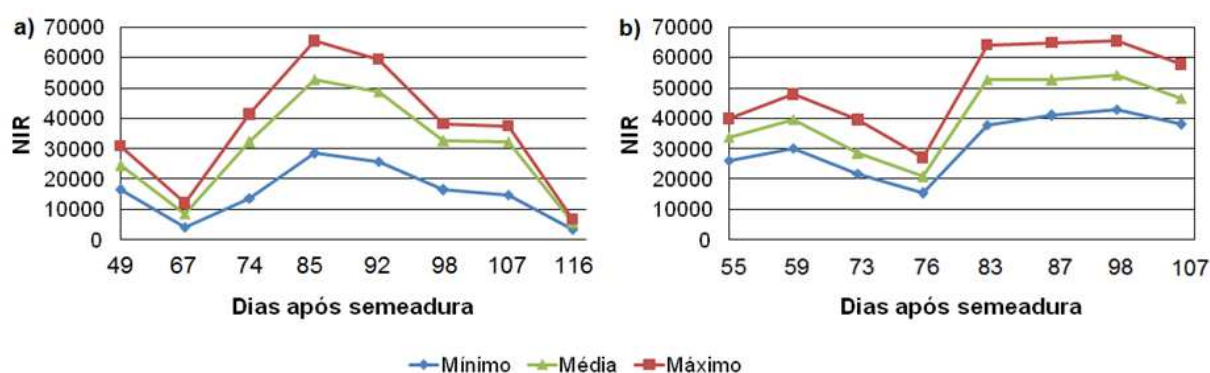


Figura 37. Comportamento dos valores de NIR ao longo das coletas. (a) Fazenda Santa Cruz e (b) Fazenda Escola

De maneira geral, os valores digitais observados na área experimental da Fazenda Santa Cruz e Fazenda Escola foram diferentes, mesmo em períodos muito próximos. Isso pode ter sido causado pela diferença nas condições de voo na hora de cada uma das coletas, uma vez que por mais que vários voos tenham sido realizados no mesmo dia, nas duas áreas, não há a garantia de que as condições no ambiente tenham sido as mesmas, além de que o período de desenvolvimento da cultura do trigo nas duas áreas era diferente, assim como a cultivar semeada.

Vale ressaltar mais uma vez que a correção radiométrica para cada uma das imagens só foi feita com base em características específicas de cada uma das câmeras, como mostrado na seção 4.5. Trabalhos como o de Calderón et al. (2013), Berni et. al. (2009) e Shi et al. (2016) realizam a correção radiométrica das imagens

utilizadas com base na radiação solar como uma garantia na precisão dos valores digitais, porém tais trabalhos não mostram os ganhos com tal correção.

Em relação ao número de instâncias (linhas) nas bases de dados, os valores são os mesmos mostrados na Tabela 5, conforme explicado na seção 4.7. A quantidade de pontos variou de acordo com resolução espacial das imagens das quais foram extraídas as bases, bem como as diferentes proporções.

5.3 RESULTADOS DA REGRESSÃO E MODELOS PREDITIVOS

A mineração das bases de dados, aplicando o algoritmo de regressão, para a geração do modelo preditivo de produtividade do trigo, foi realizada de três formas, variando os atributos de predição e mantendo sempre o mesmo atributo meta (produtividade do trigo). As três variações dos atributos de predição foram: (i) considerando somente os valores extraídos através das imagens da câmera RGB, (ii) para valores extraídos a partir da câmera NIR e (iii) para os valores de NDVI. Isso foi feito tendo em vista as diferenças em relação a cada câmera, como explicado na seção 4.2.

Essa variação de combinação de atributos foi feita para cada área experimental separadamente e para cada dia de voo, ou seja, para os diferentes períodos que foram observados. A Tabela 9, Tabela 10 e Tabela 11 apresentam parte de cada base para cada combinação, sendo que os atributos Instância e Parcela não foram usados na MD.

Tabela 9. Representação de parte da base de dados utilizada para MD com a câmera RGB para estimativa de produtividade do trigo na Fazenda Escola no período de 84DAS

Instância	Parcela	RGB_R	RGB_G	RGB_B	Produtividade
1	Parcela01	10405,43	9758,397	6660,964	3414,639
2	Parcela01	10069,46	9754,925	6598,937	3414,639
3	Parcela01	10917,18	10168,96	7029,849	3414,639
...	...	...	...	...	...
38059	Parcela44	9541,966	9315,966	6186,544	2726,598
38060	Parcela44	9114,48	9096,255	5903,063	2726,598

Tabela 10. Representação de parte da base de dados utilizada para MD com a câmera NIR para estimativa de produtividade do trigo na Fazenda Escola no período de 84DAS

Instância	Parcela	NIR_R	NIR_G	NIR	Produtividade
1	Parcela01	22460,0	24766,17	31382,34	3414,639
2	Parcela01	20844,77	23681,7	30582,76	3414,639
3	Parcela01	22435,46	24297,84	29846,9	3414,639
...	...	...	...	...	...
38059	Parcela44	23269,18	26247,86	35194,41	2726,598
38060	Parcela44	19992,42	23313,24	35290,03	2726,598

Tabela 11. Representação de parte da base de dados utilizada para MD com os valores de NDVI para estimativa de produtividade do trigo na Fazenda Escola no período de 84DAS

Instância	Parcela	NDVI	Produtividade
1	Parcela01	0,165712	3414,639
2	Parcela01	0,189354	3414,639
3	Parcela01	0,141758	3414,639
...	...	...	...
38059	Parcela44	0,203977	2726,598
38060	Parcela44	0,276717	2726,598

Os valores de correlação para a Fazenda Escola foram maiores que para a Fazenda Santa Cruz, como pode ser observado na Tabela 12 e Tabela 13. Enquanto na Fazenda Santa Cruz as correlações com base na câmera RGB não passaram de 0,3663 (116 DAS), na Fazenda Escola chegaram a 0,4819 (76 DAS). Para os valores da câmera NIR, a correlação chegou a 0,3784 (74 DAS) na área da fazenda Santa Cruz e 0,4749 (73 DAS) na Fazenda Escola.

Tabela 12. Correlações de produtividade com cada tipo diferente de dados (Câmera RGB, NIR e o NDVI) com a proporção de 10% para as bases da Fazenda Santa Cruz

Tipo de Dado	DAS							
	49	67	74	85	92	98	107	116
RGB	0,1596	0,3089	0,3350	0,3147	0,3516	0,2064	0,1129	0,3663
NIR	0,1705	0,2146	0,3784	0,3177	0,0780	0,0929	0,2475	0,1277
NDVI	0,0928	0,1574	0,0426	0,0739	0,0358	0,0727	0,0206	0,090

Tabela 13. Correlações de produtividade com cada tipo diferente de dados (Câmera RGB, NIR e o NDVI) com a proporção de 10% para as bases da Fazenda Escola

Tipo de Dado	DAS							
	55	59	73	76	83	87	98	107
RGB	0,4220	0,4461	0,4735	0,4819	0,4569	0,4770	0,4746	0,3693
NIR	0,3666	0,3806	0,4749	0,4355	0,4610	0,4376	0,4195	0,3429
NDVI	0,3569	0,3944	0,4212	0,4357	0,4384	0,3780	0,4015	0,3386

No caso do NDVI, os dados da Fazenda Santa Cruz não mostraram correlação satisfatória em nenhum período, onde o maior valor de correlação obtido foi de apenas 0,1574 em 67 DAS, período em que os valores médios de NDVI foram maiores. Já para a área da Fazenda Escola, os valores de NDVI tiveram uma correlação melhor em todos os períodos, mas ainda assim baixa, não passando de 0,4384 (83 DAS).

Os coeficientes de correlação obtidos para a câmera RGB foram, no geral, maiores que os obtidos para câmera NIR e o NDVI, o que sugere que a estimativa de produtividade do trigo utilizando dados extraídos das imagens da câmera RGB pode ser melhor.

Na Fazenda Escola, os valores de correlação para a câmera RGB foram mais altos desde as primeiras coletas, alcançando o pico ($r = 0,4819$) no período de 76 DAS e decaindo no período mais próximo a colheita (107 DAS). Já para a câmera NIR, os coeficientes de correlação foram um pouco menores no período de 55 a 59 DAS, aumentaram no período de 73 a 87 DAS e diminuíram no período mais próximo à colheita. Porém, mesmo no período de maior correlação com a câmera NIR (73 DAS), os valores de correlação para câmera RGB são muito próximos. Tais constatações sugerem que a estimativa de produtividade do trigo utilizando a câmera RGB pode ser melhor, até por que se correlacionou melhor nos períodos iniciais da cultura.

No geral, os valores de correlação obtidos na Fazenda Santa Cruz foram baixos e flutuaram bastante (Figura 38) em comparação com a Fazenda Escola, tornando difícil a interpretação e indicação de melhores ou piores períodos para estimativa.

Muitos fatores podem ter influenciado negativamente para esse resultado, os quais muitas vezes fogem ao controle, como por exemplo, a baixa variabilidade no vigor da cultura, a diferença nas condições climáticas na hora das coletas, a

diferença da cultivar de trigo, a diferença entre os tratamentos realizados nas duas áreas e até a diferença de produtividade obtida em cada um dos tratamentos pode ter ocasionado a dificuldade na geração de modelos com boa correlação.

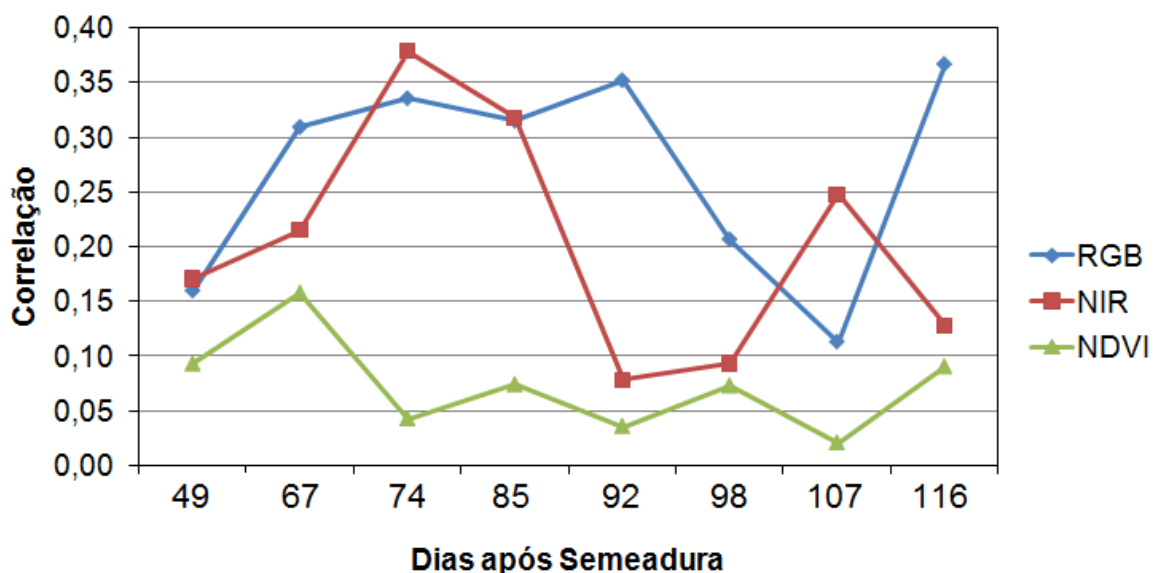


Figura 38. Índices de correlação obtidos com bases na proporção de 10% para Fazenda Santa Cruz

Não foi possível executar todas as bases de dados em todas as proporções, devido à grande quantidade de bases e a falta de hardware para tal execução, uma vez que quanto maior a proporção de cada base, maior também o tempo (Tabela 14) para construção do modelo preditivo, além de sua validação. Sendo assim, a Tabela 15 e Tabela 16 mostram as proporções que foram efetivamente executadas para Fazenda Santa Cruz, e para Fazenda Escola, respectivamente.

Tabela 14. Tempo médio para geração do modelo preditivo de acordo com a proporção da base

Proporção	Tempo Médio para Geração do Modelo
10%	48 minutos
20%	8 horas
40%	27 horas
70%	97 horas
100%	277 horas

Tabela 15. Bases de dados executadas no algoritmo de MD para Fazenda Santa Cruz, onde o X representa as bases executadas, e – (traço) as bases não executadas

Período	10%	20%	40%	70%	100%
49 DAS	X	-	-	-	-
67 DAS	X	-	X	X	-
74 DAS	X	-	-	-	-
85 DAS	X	-	X	X	-
92 DAS	X	-	X	X	-
98 DAS	X	-	X	X	-
107 DAS	X	-	X	X	-
116 DAS	X	-	-	-	-

Tabela 16. Bases de dados executadas no algoritmo de MD para Fazenda Escola, onde o X representa as bases executadas, e – (traço) as bases não executadas

Período	10%	20%	40%	70%	100%
55 DAS	X	-	X	-	-
59 DAS	X	-	X	-	-
73 DAS	X	-	X	-	-
76 DAS	X	-	X	-	-
83 DAS	X	-	X	X	-
87 DAS	X	-	X	-	-
98 DAS	X	-	X	X	-
107 DAS	X	-	X	-	-

Ao se analisar os resultados obtidos com as bases de dados em diferentes proporções, foi possível observar que mesmo com o aumento na proporção os coeficientes de correlação variaram muito pouco, não chegando sequer a um décimo de aumento na correlação. Isso pôde ser observado nos resultados em diferentes proporções nas bases de dados da Fazenda Santa Cruz (Tabela 17) e da Fazenda Escola (Tabela 18).

Tabela 17. Correlação em diferentes proporções para Fazenda Santa Cruz

DAS	RGB	NIR	NDVI
10%	0,1129	0,2475	0,0206
107 40%	0,1091	0,2506	0,0223
70%	0,1096	0,2500	-
10%	0,2064	0,0929	0,0727
98 40%	0,2088	0,0901	0,0716
70%	0,2088	0,0901	0,0714

92	10%	0,3516	0,0780	0,0358
	40%	0,3501	0,0725	0,0340
	70%	0,3498	0,0734	-
85	10%	0,3147	0,3177	0,0739
	40%	0,3174	0,3164	0,0771
	70%	0,3192	0,3170	0,0780
67	10%	0,3089	0,2146	0,1574
	40%	0,3083	0,2161	-
	70%	-	0,2142	0,1606

Tabela 18. Correlação em diferentes proporções para Fazenda Escola

	DAS	RGB	NIR	NDVI
107	10%	0,3693	0,3429	0,3386
	40%	0,3738	0,3470	0,3421
	70%	-	-	-
87	10%	0,4770	0,4376	0,3780
	40%	0,4817	0,4393	0,3842
	70%	-	-	-
83	10%	0,4569	0,4610	0,4384
	40%	0,4607	0,4604	0,4396
	70%	-	-	-
73	10%	0,4735	0,4749	0,4212
	40%	0,4771	-	0,4254
	70%	-	-	-
55	10%	0,4220	0,3666	0,3569
	40%	0,4231	0,3664	0,3552
	70%	-	-	-

Ao observar os valores médios de refletância para cada um dos canais utilizados nas bases de dados, pode-se perceber que praticamente não há diferença entre eles, como pode ser visualizado nos valores do canal R da câmera RGB para área da Fazenda Santa Cruz (Figura 39) e Fazenda Escola (Figura 40).

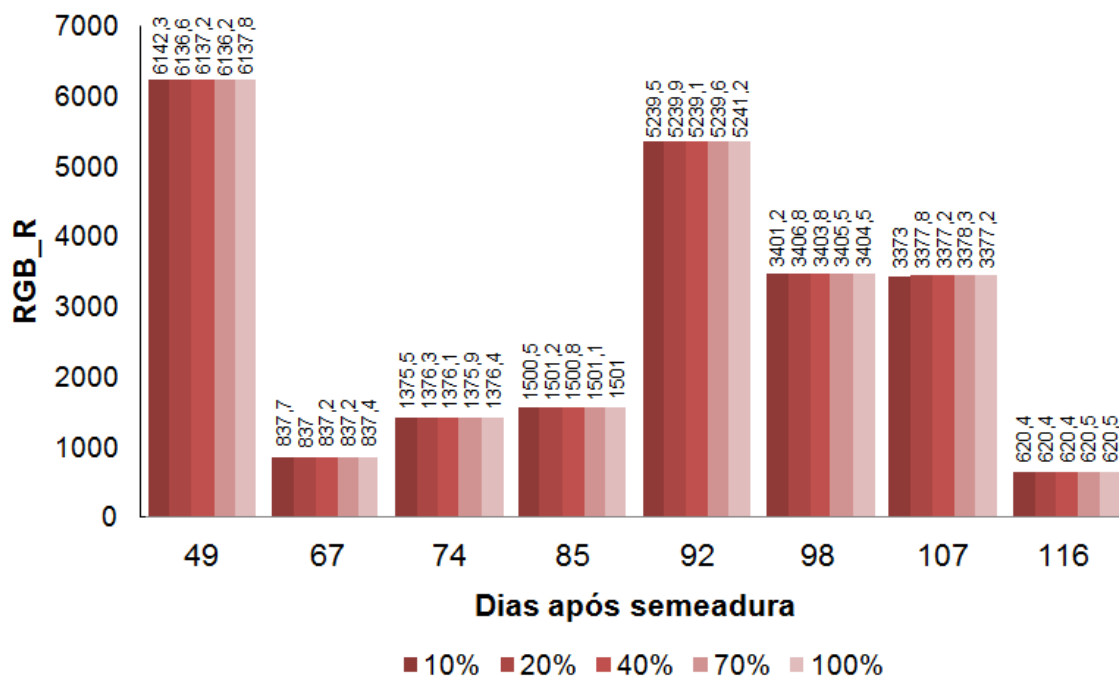


Figura 39. Valores médios de refletância para o canal RGB_R nos diferentes períodos de coleta na Fazenda Santa Cruz, considerando as diferentes proporções das bases de dados

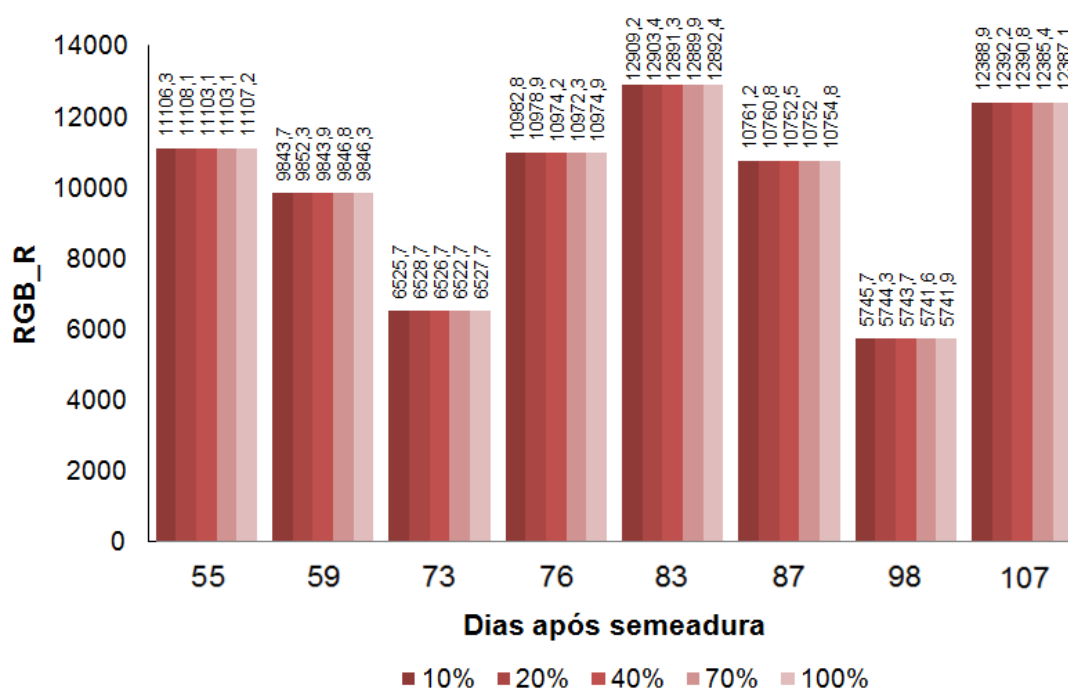


Figura 40. Valores médios de refletância para o canal RGB_R nos diferentes períodos de coleta na Fazenda Escola, considerando as diferentes proporções das bases de dados

Para que fosse possível comprovar estatisticamente a igualdade entre as bases de dados em diferentes proporções, foram criadas tabelas para realização de

um teste de Tukey, com base na média e desvio padrão dos valores digitais de cada um dos canais para todos os períodos de coleta.

No resultado do teste de Tukey para os valores digitais do canal RGB_R (Tabela 19 e Tabela 20), é possível observar que independentemente da proporção da base não há diferença significativa entre os valores digitais em cada período de coleta. O mesmo teste foi aplicado em todos os canais (RGB_G, RGB_B, NIR_R, NIR_G, NIR e NDVI) nas bases de dados das duas áreas de estudo, e mostrou o mesmo resultado em todas as aplicações.

O fato de não haver diferença nos índices de correlação independente da proporção da base acontece pois, uma vez que os valores digitais nas bases de dados são os mesmos, consequentemente os modelos preditivos gerados serão estatisticamente os mesmos, ou seja, os pesos atribuídos a cada variável do modelo serão muito parecidos, fazendo com que a correlação permaneça igual entre as proporções de base.

Tabela 19. Valores de RGB_R nos diferentes períodos de coleta da Fazenda Santa Cruz em diferentes proporções¹

	49 DAS	67 DAS	74 DAS	85 DAS
10%	6142,3 ± 921,7 ^A	837,7 ± 205,1 ^B	1375,5 ± 276,3 ^C	1500,5 ± 232,1 ^D
20%	6136,6 ± 920,2 ^A	837,0 ± 205,0 ^B	1376,3 ± 277,8 ^C	1501,2 ± 231,3 ^D
40%	6137,2 ± 921,4 ^A	837,2 ± 204,8 ^B	1376,1 ± 276,8 ^C	1500,8 ± 231,9 ^D
70%	6136,2 ± 920,6 ^A	837,2 ± 204,8 ^B	1375,9 ± 276,5 ^C	1501,1 ± 230,5 ^D
100%	6137,8 ± 921,6 ^A	837,4 ± 205,4 ^B	1376,4 ± 277,9 ^C	1501,0 ± 232,1 ^D
	92 DAS	98 DAS	107 DAS	116 DAS
10%	5239,5 ± 857,6 ^E	3401,2 ± 547,9 ^F	3373,0 ± 454,2 ^G	620,4 ± 67,7 ^H
20%	5239,9 ± 851,3 ^E	3406,8 ± 548,9 ^F	3377,8 ± 451,3 ^G	620,4 ± 67,4 ^H
40%	5239,1 ± 853,5 ^E	3403,8 ± 549,8 ^F	3377,2 ± 452,5 ^G	620,4 ± 67,5 ^H
70%	5239,6 ± 852,4 ^E	3405,5 ± 548,3 ^F	3378,3 ± 450,8 ^G	620,5 ± 67,3 ^H
100%	5241,2 ± 855,4 ^E	3404,5 ± 548,9 ^F	3377,2 ± 452,3 ^G	620,5 ± 67,7 ^H

¹ Letras iguais indicam médias iguais pelo teste de Tukey (p<0.05).

Tabela 20. Valores de RGB_R nos diferentes períodos de coleta da Fazenda Escola em diferentes proporções¹

	55 DAS	59 DAS	73 DAS	76 DAS
10%	11106,3 ± 1621,6 ^A	9843,7 ± 1557,6 ^B	6525,7 ± 1252,4 ^C	10982,8 ± 2244,1 ^D
20%	11108,1 ± 1632,4 ^A	9852,3 ± 1567,1 ^B	6528,7 ± 1255,4 ^C	10978,9 ± 2240,3 ^D
40%	11103,1 ± 1632,7 ^A	9843,9 ± 1570,9 ^B	6526,7 ± 1257,5 ^C	10974,2 ± 2242,8 ^D
70%	11103,1 ± 1632,3 ^A	9846,8 ± 1569,6 ^B	6522,7 ± 1259,5 ^C	10972,3 ± 2243,9 ^D
100%	11107,2 ± 1631,2 ^A	9846,3 ± 1568,0 ^B	6527,7 ± 1256,8 ^C	10974,9 ± 2241,4 ^D
	83 DAS	87 DAS	98 DAS	107 DAS
10%	12909,2 ± 2960,7 ^E	10761,2 ± 2258,4 ^F	5745,7 ± 2028,9 ^G	12388,9 ± 1926,3 ^H
20%	12903,4 ± 2959,8 ^E	10760,8 ± 2270,4 ^F	5744,3 ± 2021,2 ^G	12392,2 ± 1927,7 ^H
40%	12891,3 ± 2956,4 ^E	10752,5 ± 2269,8 ^F	5743,7 ± 2025,1 ^G	12390,8 ± 1937,4 ^H
70%	12889,9 ± 2959,7 ^E	10752,0 ± 2270,1 ^F	5741,6 ± 2025,7 ^G	12385,4 ± 1938,1 ^H
100%	12892,4 ± 2958,9 ^E	10754,8 ± 2270,8 ^F	5741,9 ± 2025,2 ^G	12387,1 ± 1934,5 ^H

¹ Letras iguais indicam médias iguais pelo teste de Tukey (p<0.05).

Isso pode ser observado nos modelos preditivos gerados para bases em diferentes proporções da Fazenda Santa Cruz (Tabela 21) e da Fazenda Escola (Tabela 22).

Tabela 21. Modelos de estimativa de produtividade do trigo gerados em diferentes proporções para a Fazenda Santa Cruz¹

	10%	40%	70%
98 DAS			
RGB	- 0,1046 * RGB_R	- 0,1146 * RGB_R	- 0,0980 * RGB_R
	+ 0,7751 * RGB_G	+ 0,7779 * RGB_G	+ 0,7620 * RGB_G
	- 0,3793 * RGB_B	- 0,3729 * RGB_B	- 0,3986 * RGB_B
	+ 0,2922 (r = 0,2064)	+ 0,2944 (r = 0,2028)	+ 0,2965 (r = 0,2088)
NIR	+ 0,1433 * NIR_R	+ 0,1424 * NIR_R	+ 0,1418 * NIR_R
	- 0,2254 * NIR_G	- 0,2257 * NIR_G	- 0,2260 * NIR_G
	+ 0,1178 * NIR	+ 0,1167 * NIR	+ 0,1146 * NIR
	+ 0,4898 (r = 0,0929)	+ 0,4891 (r = 0,0896)	+ 0,4864 (r = 0,0901)
NDVI	+ 0,0253 * NDVI	+ 0,0266 * NDVI	+ 0,0272 * NDVI
	+ 0,4756 (r = 0,0727)	+ 0,4751 (r = 0,0716)	+ 0,4746 (r = 0,0714)
85 DAS			
RGB	- 1,0811 * RGB_R	- 1,0696 * RGB_R	- 1,0929 * RGB_R
	+ 1,4162 * RGB_G	+ 1,4599 * RGB_G	+ 1,5044 * RGB_G
	+ 0,0320 * RGB_B	+ 0,0213 * RGB_B	+ 0,0312 * RGB_B

	+ 0,2337 (r = 0,3147)	+ 0,2115 (r = 0,3174)	+ 0,2106 (r = 0,3192)
	- 2,6983 * NIR_R	- 2,6657 * NIR_R	- 2,7477 * NIR_R
NIR	+ 3,0969 * NIR_G	+ 2,9579 * NIR_G	+ 3,1878 * NIR_G
	- 1,2376 * NIR	- 1,2771 * NIR	- 1,2871 * NIR
	+ 0,6149 (r = 0,3177)	+ 0,7749 (r = 0,3164)	+ 0,6406 (r = 0,3170)
NDVI	- 0,0738 * NDVI	- 0,0782 * NDVI	- 0,0764 * NDVI
	+ 0,5376 (r = 0,0739)	+ 0,5375 (r = 0,0771)	+ 0,537 (r = 0,0780)

¹ Valores normalizados de RGB_R, RGB_G, RGB_B, NIR_R, NIR_G, NIR e NDVI

Tabela 22. Modelos de estimativa de produtividade do trigo gerados em diferentes proporções para a Fazenda Escola¹

	10%	40%	70%
98 DAS			
RGB	- 3,7997 * RGB_R + 3,3436 * RGB_G + 0,2843 * RGB_B + 0,4605 (r = 0,4746)	- 3,8933 * RGB_R + 3,5330 * RGB_G + 0,2091 * RGB_B + 0,4322 (r = 0,4774)	-
NIR	+ 0,7590 * NIR_R - 1,4930 * NIR_G + 0,4258 * NIR + 0,5301 (r = 0,4195)	+ 0,7541 * NIR_R - 1,5465 * NIR_G + 0,4415 * NIR + 0,4744 (r = 0,4236)	+ 0,7876 * NIR_R - 1,6267 * NIR_G + 0,4405 * NIR + 0,515 (r = 0,4256)
NDVI	+ 0,5598 * NDVI + 0,0736 (r = 0,4015)	+ 0,5954 * NDVI + 0,0477 (r = 0,4055)	+ 0,5996 * NDVI + 0,0484 (r = 0,4073)
	10%	40%	70%
83 DAS			
RGB	- 2,6158 * RGB_R + 3,2657 * RGB_G - 0,7985 * RGB_B + 0,4657 (r = 0,4569)	- 2,8135 * RGB_R + 3,5293 * RGB_G - 0,8637 * RGB_B + 0,4684 (r = 0,4607)	- 2,8128 * RGB_R + 3,4692 * RGB_G - 0,8117 * RGB_B + 0,4184 (r = 0,4637)
NIR	+ 0,555 * NIR_R - 1,1096 * NIR_G + 0,7415 * NIR + 0,1961 (r = 0,4610)	+ 0,6074 * NIR_R - 1,2147 * NIR_G + 0,7527 * NIR + 0,209 (r = 0,4604)	+ 0,6027 * NIR_R - 1,2229 * NIR_G + 0,753 * NIR + 0,2234 (r = 0,4631)
NDVI	+ 0,609 * NDVI + 0,0059 (r = 0,4384)	+ 0,6068 * NDVI + 0,0078 (r = 0,4396)	-

¹ Valores normalizados de RGB_R, RGB_G, RGB_B, NIR_R, NIR_G, NIR e NDVI

Os mesmos testes foram aplicados nas bases de dados extraídas de imagens nas resoluções de 10cm/px e 20cm/px, e mostraram o mesmo comportamento. Dessa forma, as bases em diferentes proporções na mesma resolução continuam iguais, mantendo os coeficientes de correlação quando aplicadas no algoritmo de regressão. Porém, ao comparar bases entre resoluções diferentes, podem-se observar diferenças, como mostrado na Tabela 23 e Tabela 24.

Tabela 23. Coeficientes de correlação em diferentes resoluções. Fazenda Escola a 87 DAS

	3,4 cm/px	10 cm/px	20 cm/px
RGB	0,4770	0,5871	0,6060
NIR	0,4376	0,5377	0,5423
NDVI	0,3780	0,5094	0,5424

Tabela 24. Coeficientes de correlação em diferentes resoluções. Fazenda Escola a 98 DAS

	3.4 cm/px	10 cm/px	20 cm/px
RGB	0,4746	0,5608	0,6168
NIR	0,4195	0,5006	0,5411
NDVI	0,4015	0,4741	0,5113

Como pode-se perceber, há uma diferença no coeficiente de correlação das três bases com resolução de 3,4cm/px em comparação com as resoluções de 10cm/px e 20cm/px. Ao se diminuir a resolução espacial das imagens, características da área de estudo são perdidas, uma vez que uma área maior será representada por um único pixel.

Nesse caso, através dos coeficientes de correlação alcançados, é possível observar que os maiores coeficientes de correlação foram atingidos em bases da Fazenda Escola, extraídas a partir de imagens na resolução de 20cm/px, ou seja, imagens que possuem menos características (Figura 41).

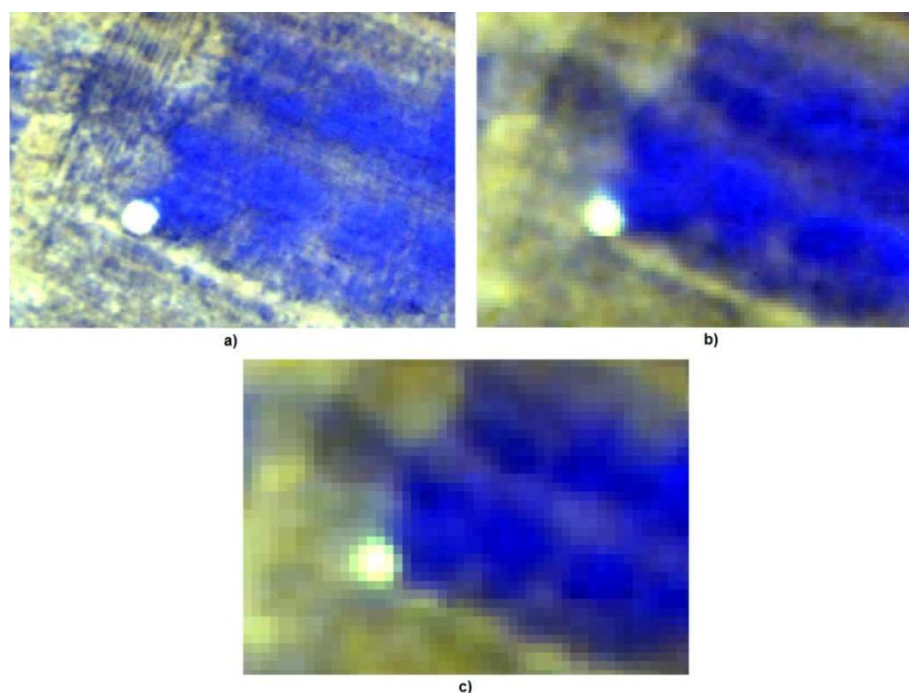


Figura 41. Diferença nas características de uma imagem nas resoluções de 3,4cm/px (a), 10cm/px (b) e 20cm/px (c)

Tais resultados mostram que a utilização de imagens em alta resolução podem não ser adequadas para a estimativa da produtividade do trigo. Os respectivos modelos preditivos para cada uma das bases que obtiveram maior correlação são apresentados na Tabela 25.

Tabela 25. Modelos gerados para os maiores coeficientes de correlação¹

RGB	NIR	NDVI
87 DAS		
- 2,3146 * RGB_R + 1,3092 * RGB_G + 0,6121 * RGB_B + 0,6263 (r = 0,6060)	+ 0,1534 * NIR_R - 0,5326 * NIR_G + 0,6958 * NIR + 0,2051 (r = 0,5423)	+ 0,6674 * NDVI + 0,0133 (r = 0,5424)
98 DAS		
- 1,7176 * RGB_R - 1,5705 * RGB_G + 2,9794 * RGB_B + 0,4395 (r = 0,6168)	+ 0,4867 * NIR_R - 0,8905 * NIR_G + 0,6062 * NIR + 0,2875 (r = 0,5411)	+ 0,5897 * NDVI + 0,0383 (r = 0,5153)

¹ Valores normalizados de RGB_R, RGB_G, RGB_B, NIR_R, NIR_G, NIR e NDVI

No trabalho de Bose et al. (2016), também foi aplicado um algoritmo de SVR para a estimativa de produtividade do trigo com base no NDVI, a partir de imagens de satélite, com resolução de 250 metros, obtendo coeficiente de correlação de 0,67. Além do SVR, os autores também utilizaram regressão linear, algoritmo do K-vizinho mais próximo e redes neurais, obtendo as maiores correlações (r = 0,81) através das redes neurais. A que se mencionar aqui que a SVM tem a vantagem sobre a rede neural de gerar um modelo numérico explícito, o que permite compreender como cada atributo de predição interfere no modelo resultante.

No trabalho de Araújo et al. (2005), os autores realizaram a estimativa de produtividade para duas áreas cultivadas com trigo, com base em imagens de videografia aérea, utilizando dados do comprimento de onda do vermelho, NIR e NDVI. Para a primeira área, foi obtida correlação de 0,819 para o NDVI, 0,594 para o NIR e 0,578 para o vermelho. Na segunda área, os índices de correlação foram muito baixos (0,158, 0,103 e 0,095 para NDVI, NIR e vermelho, respectivamente),

onde os autores associaram a problemas na câmera durante as coletas de dados. O fato dos resultados terem sido contrários aos encontrados nessa pesquisa, onde índices de correlação mais altos foram encontrados para câmera RGB e NIR, e mais baixos para o NDVI, pode ser explicado pela diferença nos equipamentos utilizados (câmeras e plataforma de voo).

Tattaris et al. (2016) utilizaram uma RPA para coleta de imagens com uma câmera NIR, para posterior cálculo de NDVI e estimativa de produtividade do trigo irrigado. Nesse trabalho, os autores utilizaram imagens com resolução de 46 cm/px, sendo que a correlação mais alta ($r = 0,73$) foi alcançada somente em uma data de coleta. Na maioria dos casos, as correlações não ultrapassaram 0,6, com valores muito baixos em alguns casos (0,16, 0,21 e 0,36). Isso pode indicar que o NDVI obtido a partir de imagens de RPA pode não ser a melhor variável de estimativa de produtividade do trigo, tendo em vista que os resultados alcançados nessa pesquisa também não mostraram correlações altas para o NDVI.

6 CONCLUSÃO E TRABALHOS FUTUROS

A mineração de dados em imagens é um método de análise adequado para o uso durante o desenvolvimento da cultura, permitindo que se tenha um indicativo do desenvolvimento do trigo antes mesmo da colheita. Foi possível gerar vários modelos predição, dentre eles, os gerados a partir de dados extraídos das imagens da câmera RGB se mostraram mais apropriados, por apresentar modelos de estimativa da produtividade de trigo com as maiores correlações. O uso do NDVI não se mostrou adequado para tal estimativa.

O software Quantum GIS apresentou melhores resultados para a correção do georreferenciamento dos mosaicos se comparado ao Pix4D. Portanto, é mais vantajoso e menos trabalhoso gerar os mosaicos pelo Pix4D sem qualquer auxílio de GCPs (pontos de controle em solo) e depois corrigi-los através do Quantum GIS.

Entre as resoluções de imagens utilizadas, a que se mostrou melhor para estimativa de produtividade foi a de 20cm/px, apresentando os maiores índices de correlação. A resolução de 3,4cm/px, que permite a visualização de mais características nas imagens, apresentou coeficientes de correlação baixos. Portanto, pode-se concluir que imagens com menor resolução espacial são melhores para estimativa de produtividade do trigo.

Os valores digitais de refletância obtidos nos diferentes comprimentos de onda são estatisticamente iguais nos diferentes tamanhos de bases avaliadas, nas proporções de 10%, 20%, 40%, 70% e 100%. Dessa forma, a utilização de bases de dados na proporção de 10% é mais vantajosa por exigir menor tempo de execução do algoritmo.

Como trabalhos futuros sugere-se:

- Aplicar os modelos gerados nos dados para verificar os erros na estimativa de produtividade do trigo, com intuito de analisar se os modelos gerados, mesmo não apresentando altos índices de correlação, podem se adequados para utilização em campo.
- Realizar a correção radiométrica com base na irradiação solar, pois dessa forma poderia se garantir a precisão da refletância em cada um dos comprimentos de onda.

- Verificar a real diferença das imagens com e sem correção radiométrica, e como cada uma afeta os resultados do processo de mineração de dados em imagens de RPA.
- Aplicar as bases de dados em diferentes algoritmos para mineração de dados, com o intuito de identificar o melhor algoritmo para estimativa de produtividade do trigo.

REFERÊNCIAS

- ARAÚJO, J. C.; VETTORAZZI, C. A.; MOLIN, J. P. Estimativa da produtividade e determinação de zonas de manejo, em culturas de grãos, por meio de videografia aérea multiespectral. **Acta Scientiarum Agronomy**, v. 27, n. 3, p. 437-447, Maringá, 2005.
- BENDIG, J.; YU, K.; AASEN, H.; BOLTEN, A.; BENNERTZ, S.; BROCHAIT, J.; GNYP, M. L.; BARETH, G.; Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. **International Journal of Applied Earth Observation and Geoinformation**, v. 39, p. 79–87, 2015.
- BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. Agricultura de Precisão: resultados de um novo olhar. **Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA**. Editora Cubo, 2014.
- BERNI, J. A. J.; ZARCO-TEJADA, P. J.; SUARÉZ, L.; FERERES, E.; Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring From An Unmanned Aerial Vehicle. **IEEE transactions on Geoscience and Remote Sensing**, v. 47, n. 3, p. 722-738, 2009.
- BRAGAGNOLO, J.; AMADO, T. J. C.; NICOLOSO, R. S.; SANTI, A. L.; FIORIN, J. E.; TABALDI, F. Sensor óptico na fertilização nitrogenada à dose variável no milho: II - índices de eficiência da fertilização e produtividade de grãos de milho. **Revista Brasileira de Ciência e Solo**, v. 37, n. 5, p. 1299-1309, 2013.
- BORÉM, A.; SCHEEREN, P. L.; **Trigo, do plantio à colheita**. Editora UFV, 2015.
- BOSE, P.; KASABOV, N. K.; BRUZZONE, L.; HARTONO, R. N. Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. **IEEE Transactions on Geoscience and Remote Sensing**. V. 54, n. 11, p. 6563-6573, 2016.
- CALDERÓN, R.; NAVAS-CORTÉS, J. A.; LUCENA, C.; ZARCO-TEJADA, P. J.; High-resolution airborne hyperspectral and thermal imagery for early detection of verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. **Remote Sensing of Environment**, v. 139, p. 231-245, 2013.
- CAMPONOGARA, A.; GALLIO, E.; BORBA, W. F.; GEORGIN, J. O atual contexto da produção de trigo no Rio Grande do Sul. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**, v. 19, n. 2, p. 246-257, 2015.
- CARVALHO, D. R.; DALLAGASSA, M. R. Mineração de dados: aplicações, ferramentas, tipos de aprendizado e outros subtemas. **AtoZ: novas práticas em informação e conhecimento**, v. 3, p. 82, 2014.

COLOMINA, I.; MOLINA, P.; Unmanned aerial systems for photogrammetry and remote sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 92, p. 79-97, 2014.

CONAB, **Acompanhamento da Safra Brasileira Grãos**, V. 4 Safra 2016/2017 – Segundo Levantamento, Brasília, p1-156, 2017.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; MERCANTE, E.; LAMPARELLI, R. A. C. Spatial autocorrelation of ndvi and gvi indices derived from landsat/tm images for soybean crops in the western of the state of parana in 2004/2005 crop season. **Engenharia Agrícola**, v. 33, n. 3, p. 525-537, 2013.

DECEA, Sistemas de aeronaves remotamente pilotadas e o acesso ao espaço aéreo brasileiro. **Departamento de Controle do Espaço Aéreo**. 2015.

EISENBEISS, H.; UAV Photogrammetry. Tese de Pós Doutorado. **Institut für Geodesie und Photogrammetrie**, Zurich, Suíça, 2009.

FIGUEIREDO, D. **Conceitos básicos de sensoriamento remoto**. Companhia Nacional de Abastecimento – CONAB. 2007. Disponível em: <http://www.conab.gov.br/conabweb/download/SIGABRASIL/manuais/conceitos_sm.pdf>. Acesso em: 15/11/2016.

GILL, M. K.; ASEFA, T.; KEMBLOWSKI, M. W.; MCKEE, M., Soil moisture prediction using support vector machines. **Journal of The American Water Resources Association**, v. 42, n. 4, p.1033-1046, 2006.

GÓMES-CANDÓN, D.; CASTRO A. I. D.; LÓPEZ-GRANADOS, F.; Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. **Precision Agriculture**, v. 15, p. 44-56, 2014.

GONZALEZ, R. C.; WOODS, R. E.; **Digital Imaging Processing**. 3 ed. Upper Saddle River: Pearson Prentice Hall, 2008.

GROHS, D. S.; BREDEMEIER, C.; MUNDSTOCK, C. M.; POLETTTO, N.; Modelo Para estimativa do potencial produtivo em trigo e cevada por meio do sensor *greenseeker*. **Engenharia Agrícola**, v. 29, n. 1, p.101 – 112. 2009.

HAN, J.; KAMBER, M.; **Data Mining: Concepts and Techniques**. Elsevier, 3 ed., 2012.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295-309, 1988.

IAPAR, **Instituto Agronômico do Paraná**, 2015. Disponível em: <<http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=863>> Acesso em: 19/01/2017.

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 ujm talhão com diferentes potenciais produtivos. **Revista Brasileira de Ciência e Solo**, v. 38, n. 04, p. 1190-1199, 2014.

LIBA, N.; BERG-JURGENS, J. Accuracy of Orthomosaic Generated by Different Methods in Example of UAV Platform MUST Q. **Materials Science and Engineering**, n. 96, v. 1, p. 1-8, 2015.

LÓPEZ-GRANADOS, F.; TORRES-SÁNCHEZ, J.; SERRANO, N.; ARQUERO, O.; PEÑA, J. M.; High-throughput 3-D monitoring of agricultural-tree plantations with Unamanned Aerial Vehicle (UAV) technology. **Plos One**, v. 10, n. 06, p. 1-20, 2015.

MAPA, Agricultura de Precisão, **Ministério da Agricultura, Pecuária e Abastecimento**, 3rd ed., p. 1-40, 2013. Disponível em: <http://www.agricultura.gov.br/arq_editor/file/Desenvolvimento_Sustentavel/Agricultura-Precisao/Boletim%20T%C3%A9cnico%20-%20Agricultura%20de%20Precis%C3%A3o%202013.pdf>. Acesso em: 15/11/2016.

MENESES, P. R., ALMEIDA, T. **Introdução ao processamento de imagens de sensoriamento remoto**. UNB, E-Book, Brasília, 2012. Disponível em: <<http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8>>. Acesso em 15/11/2016.

MEYER, P.; ITTEN, K. I.; KELLENBERGER, T.; STEFAN, S.; SANDMEIER, R.; Radiometris corrections of topographically induced effects on landsat tm data in alpane terrain. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 48, n. 4, p. 17-28, 1993.

NAVIDI, W.,. **Probabilidade e estatística para ciências exatas**. AMGH Editora. 1 Ed, 2012.

PIX4D, Pix4Dmapper 3.0 user manual, **Pix4D SA**, Lausanne, Suíça, 2016a. Disponível em: <<https://support.pix4d.com/hc/en-us/articles/204272989-Offline-Getting-Started-and-Manual-pdf-#gsc.tab=0>>. Acesso em: 10/12/2016.

PIX4D, Difference between the reflectance map and the orthomosaic, **Pix4D Support Site**, 2016b. Disponível em: < <https://support.pix4d.com/hc/en-us/articles/202739409-Difference-between-the-Reflectance-Map-and-the-Orthomosaic#gsc.tab=0>>. Acesso em: 10/12/2016.

PIX4D, Camera requirements for precision agriculture. **Pix4D AS**, Lausanne, Suíça, 2016c. Disponível em: <<https://support.pix4d.com/hc/en-us/articles/204894705-Camera-Requirements-for-Precision-Agriculture#gsc.tab=0>>. Acesso em 12/12/2016.

QGIS, Quantum GIS Documentation – Georeferencer Plugin, 2016. Disponível em: <http://docs.qgis.org/2.0/en/docs/user_manual/plugins/plugins_georeferencer.html>. Acesso em 10/12/2016.

RABELO, P. M. Conjuntura Semanal – Trigo. **CONAB – Companhia Nacional do Abastecimento**. 2016. Disponível em: <

http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_09_13_15_20_18_trigo_-_semana_-_05_a_09.pdf>. Acesso em 17/11/2016.

SANTIAGO-CINTRA, eBee – Camera S110, **Santiago e Cintra Geotecnologias**, São Paulo, 2016a.

SANTIAGO-CINTRA, eBee – Camera WX, **Santiago e Cintra Geotecnologias**, São Paulo, 2016b.

SANTOS, J.; Índices de vegetação e monitoramento do uso do solo e cobertura vegetal na bacia do rio Araguari – MG – utilizando dados do sensor MODIS. Dissertação (Mestrado em Geografia). **Universidade Federal de Uberlândia**. Uberlândia, Mg, 2005.

SIC, WorldView-4 Satellite Sensor (0.31m). **Satellite Imaging Corporation**. 2016. Disponível em: < <http://www.satimagingcorp.com/satellite-sensors/geoeye-2/>>. Acesso em 17/01/2017.

SATIR, O.; BERBEROGLU, S. Crop yield prediction under soil salinity using satellite derived vegetation indices. **Field Crops Research**, v. 192, p.134-143, 2016.

SCHADECK, F. A.; CARDOSO, C. D. V.; Fertilidade de solo e viabilidade técnica – econômica da agricultura de precisão na região das Missões RS. **Revista Brasileira de Geomática**, v. 04, n. 03, p. 139-150, 2016.

SHI, Y.; THOMASSON, J. A.; MURRAY, S. C.; PUGH, N. A.; ROONEY, W. L.; SHAFIAN, S.; RAJAN, N.; ROUZE, G.; MORGAN, C. L. S.; NEELY, H. L.; RANA, A.; BAGAVATHIANNAN, M. V.; HENRICKSON, J.; BOWDEN, E.; VALASEK, J.; OLSENHOLLER, J.; BISHOP, M. P.; SHERIDAN, R.; PUTMAN, E. B.; POPESCU, S.; BURKS, T.; COPE, D.; IBRAHIM, A.; MCCUTCHEN, B. F.; BALTENSBERGER D. D.; AVANT, R. V.; VIDRINE, M.; YANG, C., Unmanned aerial vehicles for high-throughput phenotyping and agronomic reseracjh. **PLoS ONE**, v. 11, n.7, p. 1-26, 2016.

SILVA, F. C. M.; SILVA, N. M.; CÂNDIDO, A. K. A. A; Seleção de técnicas de classificação de fotografias aéreas derivadas de VANT na análise ambiental de área de cerrado. **Revista Eletrônica do PRODEMA**, v.10, n. 01, p. 74-84, 2016.

SOUZA, E. M. F. R.; VICENS, R. S.; CRUZ, C. B. M.; Processamento digital de imagem hyperspectral do sensor Hyperion. **Revista Brasileira de Cartografia**, v. 65, n. 5, p. 895-906, 2013.

TAN, P. N.; STEINBACH, M.; KUMAR, V.; **Introdução ao Data Mining – Mineração de Dados**. Ciência Moderna, 1 Ed, 2009.

TARTACHNYK, I. I.; RADEMACHER, I.; KUHBAUCH, W.; Distinguishing nitrogen deficiency and fungal infection of winter wheat by laser-induced fluorescence. **Precision Agriculture**, v. 07, p. 281-193, 2006.

TATTARIS, M.; REYNOLDS, M. P.; CHAPMAN, S. C.; A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. **Frontiers in Plant Science**, v. 7, artigo 1131, 2016.

TEILLET, P. M.; Image correction for radiometric effects in remote sensing. **International Journal of Remote Sensing**, v.7, n.12, p. 1637-1651, 1986.

TORRES-SÁNCHEZ, J.; PEÑA, J. M.; DE CASTRO, A. I.; LÓPEZ-GRANADOS, F. Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. **Computers and Electronics in Agriculture**, v. 103, p. 104–113, 2014

USGS, Landsat 8. **United States Geological Survey**, v. 10, v. 04, 2016. Disponível em: < <https://landsat.usgs.gov/landsat-8>>. Acesso em 17/01/2017.

VERGER, A.; VIGNEAU, N.; CHÉRON, C.; GILLIOT, J. M.; COMAR, A.; BARET, F. Green area index from an unmanned aerial system over wheat and rapeseed crops. **Remote Sensing of Environment**, v. 152, p. 654–664, 2014.

WANG, L.; TIAN, Y.; YAO, X.; ZHU, Y.; CAO, W. Predicting grain yield and protein content in wheat by fusing multisensor and multi-temporal remote-sensing images. **Field Crops Research**, v. 164, p. 178-188, 2014.

WEKA, Weka 3: Data mining software in java. **The University of Waikato**, 2016. Disponível em: < <http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em 13/12/2016.

WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. **Data mining: practical machine learning tools and techniques**. Morgan Kauffmann, Ed. 4, 2016.

XIAO, J.; MOODY, A. A comparison of methods for estimating fractional green vegetation cover within a desert-to-upland transition zone in central New Mexico, USA. **Remote Sensing of Environment**, v. 98, n. 2, p. 237 – 250, 2005.

YANG, C. A high-resolution airborne four-camera imaging system for agricultural remote sensing. **Computers and Electronics in Agriculture**, v. 88, p. 13-24, 2012.

YU, Z.; CAO, Z.; WU, X.; BAI, X.; QIN, Y.; ZHUO, W.; XIAO, Y.; ZHANG, X.; XUE, H. Automatic image-based detection technology for two critical growth stages of maize: Emergence and three-leaf stage. **Agricultural and Forest Meteorology**, v. 175, p. 65-84, 2013.