

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**PREVISÃO E ESTIMATIVA DE VARIÁVEIS DE CULTIVO UTILIZANDO NÍVEIS E
FORMAS DE SENSORIAMENTO REMOTO E TÉCNICAS DE APRENDIZADO DE
MÁQUINA**

**Mailson Freire de Oliveira
Engenheiro Agrônomo**

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**PREVISÃO E ESTIMATIVA DE VARIÁVEIS DE CULTIVO UTILIZANDO NÍVEIS E
FORMAS DE SENSORIAMENTO REMOTO E TÉCNICAS DE APRENDIZADO DE
MÁQUINA**

Mailson Freire de Oliveira

**Orientador: Prof. Dr. Rouverson Pereira da
Silva**

**Coorientadores: Prof. Dra. Brenda Valeska
Ortiz**

Prof. Dr. Carlos Eduardo Angeli Furlani

Prof. Dr. Glauco de Souza Rolim

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – UNESP, Câmpus de Jaboticabal, como parte das exigências para obtenção do título de Doutor em Agronomia (Produção Vegetal).

2021

O48p Oliveira, Mailson Freire de

Previsão e estimativa de variáveis de cultivo utilizando níveis e formas de sensoriamento remoto e técnicas de aprendizado de máquina / Mailson Freire de Oliveira. -- Jaboticabal, 2021

109 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Rouverson Pereira da Silva

Coorientadora: Brenda Valeska Ortiz

1. Agricultura de precisão. 2. Processamento de dados. 3. Aprendizado de Máquinas. 4. Previsão de Variáveis Agronômicas. 5. Cultivos Agrícolas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: PREVISÃO E ESTIMATIVA DE VARIÁVEIS DE CULTIVO UTILIZANDO NÍVEIS E FORMAS DE SENSORIAMENTO REMOTO E TÉCNICAS DE APRENDIZADO DE MÁQUINA

AUTOR: MAILSON FREIRE DE OLIVEIRA

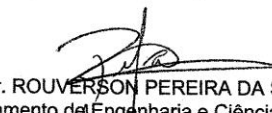
ORIENTADOR: ROVERSON PEREIRA DA SILVA

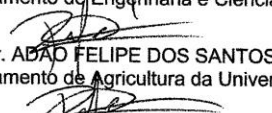
COORIENTADOR: GLAUCO DE SOUZA ROLIM

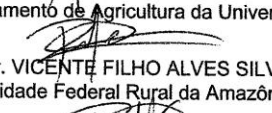
COORIENTADORA: BRENDA VALESKA ORTIZ

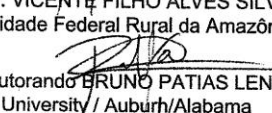
COORIENTADOR: CARLOS EDUARDO ANGELI FURLANI

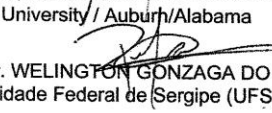
Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (PRODUÇÃO VEGETAL), pela Comissão Examinadora:


Prof. Dr. ROVERSON PEREIRA DA SILVA (Participação Virtual)
Departamento de Engenharia e Ciências Exatas (DECEX) / FCAV / UNESP - Jaboticabal


Prof. Dr. ADAO FELIPE DOS SANTOS (Participação Virtual)
Departamento de Agricultura da Universidade Federal de Lavras (UFLA) / Lavras/MG


Prof. Dr. VICENTE FILHO ALVES SILVA (Participação Virtual)
Universidade Federal Rural da Amazônia-Campus Parauapebas / Parauapebas/PA


Pós-Doutorando BRUNO PATIAS LENA (Participação Virtual)
Auburn University / Auburn/Alabama


Prof. Dr. WELLINGTON GONZAGA DO VALE (Participação Virtual)
Universidade Federal de Sergipe (UFS) / São Cristóvão/SE

Jaboticabal, 04 de maio de 2021

DADOS CURRICULARES DOS AUTORES

ME. MAILSON FREIRE DE OLIVEIRA – nascido em Parauapebas, Pará, Brasil, no dia 28 de junho de 1993, filho de Marleide Freire de Oliveira e Antônio Wilson Rodrigues de Oliveira. cursou o ensino fundamental nas escolas Autonomia e Escola Municipal de Ensino Fundamental e Médio Carlos Henrique de Andrade, deste município e ensino médio na Escola de Ensino Fundamental e Médio Eduardo Angelim em Parauapebas, tendo finalizado o ensino básico e médio no ano de 2011. Ingressou no ensino superior no ano de 2011 no curso de Engenharia Agrônoma pela Universidade Federal Rural da Amazônia (UFRA), Câmpus de Parauapebas, obtendo o título de Engenheiro Agrônomo em março de 2016. Durante a graduação foi bolsista de Iniciação Científica durante três anos, sendo um ano pela Pró-reitoria de Ensino e dois anos pelo convênio UFRA/VALE, nas áreas de Fertilidade do Solo e Manejo Integrado de Plantas Daninhas, sob orientação dos professores Dr. Eduardo do Valle Lima e Rafael Gomes Viana nas respectivas áreas. Participou do projeto Controle de Gramíneas Exóticas Invasoras na Flona de Carajás, como integrante do grupo de pesquisa Manejo Integrado de Plantas Daninhas na Amazônia. Também colaborou na execução e apresentação de trabalhos publicados em congressos e revistas científicas. Em 2016 no mês de março, iniciou no curso de Mestrado Acadêmico em Agronomia (Produção Vegetal), com concentração nas áreas de Máquinas Agrícolas e Agricultura de Precisão, pela Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, São Paulo, no Laboratório de Máquinas e Mecanização Agrícola, localizado no Departamento de Engenharia Rural da UNESP. No decorrer do curso realizou pesquisas com ênfase em Agricultura de Precisão, principalmente sensoriamento remoto proximal em diferentes culturas: milho, amendoim, soja e café. Em julho de 2017 obteve o título de Mestre em Agronomia e, em agosto de 2017 iniciou o curso de doutorado no mesmo Programa. Em maio de 2021 submeteu-se à banca examinadora para obtenção do título de Doutor em Agronomia (Produção Vegetal).

Ao olhar o mundo, capture cumprimentos de
onda além do visível, procure por radiações
desconhecidas, e enxergará o que está ao
seu alcance de forma fenomenal.

Mailson Oliveira

Aos meus pais, Marleide Freire de Oliveira e Antônio Wilson Rodrigues de Oliveira, que me ensinaram os maiores valores da vida, que me norteiam nos caminhos da vida.

DEDICO

Ao meu irmão Wildiney Freire de Oliveira,
meu afilhado Kayo Krysóstomo Maciel Viana
Freire, minha avó Isabel de Sousa Barros
Silva (*In memorian*) e meu avô João Soares
(*In memorian*).

OFEREÇO

AGRADECIMENTOS

Agradeço primeiramente a Deus, por toda a compaixão e bondade para conosco e que em um momento singular criou o que hoje chamamos de tudo, e a partir disso nós evoluímos ao ponto de poder desfrutar das belezas de sua criação.

Aos meus pais, Marleide Freire de Oliveira e Antônio Wilson Rodrigues de Oliveira, que nunca mediram esforços, muitas vezes se sacrificando para me dar todas as condições para que eu pudesse estar aqui hoje realizando uma grande meta da minha vida e por todo carinho, educação, amor e exemplo que sempre me deram para que eu pudesse me tornar a pessoa que sou hoje.

Ao meu irmão Wildiney Freire de Oliveira, pelo grande exemplo de humildade e genialidade que sempre me passou na vida pessoal e acadêmica.

Ao meu afilhado Kayo Krysóstomo Maciel Viana Freire pelos momentos vividos e as batalhas vividas em Summoner's Rift, que por vezes foram travadas para desestressar, mas falhavam miseravelmente, gerando ainda mais estresse.

À Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal-SP, Brasil e de modo especial ao Programa de Pós-Graduação em Agronomia (Produção Vegetal). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo 169775/2017-8.

Ao meu orientador Prof. Dr. Rouverson Pereira da Silva por ter me acolhido em um momento de total desilusão, me guiou de forma magistral ao caminho da ciência de ponta, confiando no meu trabalho e sempre me incentivando a continuar.

Ao meu coorientador Prof. Dr. Glauco Rolim que me apresentou o caminho da modelagem e hoje sigo firme como um jovem cientista em uma área tão encantadora, por vezes complexa (de perder cabelos) mas de grande importância para a compreensão dos fenômenos agrícolas.

À minha coorientadora Dr. Brenda Ortiz, por me receber em seu grupo de pesquisa durante meu doutorado sanduiche, e me ensinar muitos princípios de ciência e compromisso com a ciência de qualidade.

Ao meu coorientador Prof. Dr. Eduardo Furlani, pelos ensinamentos e conselhos para o planejamento da minha carreira profissional.

Aos colegas e amigos do Laboratório de Máquinas e Mecanização Agrícola (LAMMA) – Aline Alcântara, Danilo Tedesco, Edcarla Nicolau, Leonardo Bernache,

Lucas Girio, Nivanda Lima, Rafael Corrêa, Renata Queiroz, Tiago Tavares, Francieli Moreli, Letícia Santos, Jean Lucas, Marcelo Barbosa Junior e Bruno Rocca. Obrigado pelos momentos vividos!

Ao Departamento de Engenharia Rural da FCAV, em especial aos funcionários Valdecir Aparício (Maranhão) e Sebastião Filho (Tiãozinho) pelos conhecimentos práticos ensinados e a todos os funcionários e professores do departamento.

Aos amigos que fiz nos Estados Unidos Bruno, Thays Silva, Guilherme Morata, Gabi Zaghi (Pomba Neném), Helen Oliveira (Marreco Carioca) e aos companheiros de AP Rafael Dias (Índio Donizete) e Brodrick Brown. Certamente não teria aguentado a barra dos Estados Unidos sem vocês.

A minha segunda família (República Toca Fogo), aos seguintes irmãos; Luan Oliveira, Karol Tomaz, Edgard Henrique, Jarlyson Sousa, Samira Ratum, Armando Brito, Franciele Morlin, Seu Paulo, João Godinho, Maria Albertina, Helena Carvalho, Ancelmo Cazuza. Meus caros, obrigado de coração pelo afeto familiar.

Ao meu irmão de república e padrinho de fogueira Adão, um cara que sempre acreditou no meu potencial e me guiou dando exemplo de excelência e superação pessoal e profissional. Agradeço de coração os conselhos e momentos vividos.

A todos que de alguma forma contribuíram para que este momento memorável se tonasse possível ficam aqui meus agradecimentos.

PREVISÃO E ESTIMATIVA DE VARIÁVEIS DE CULTIVO UTILIZANDO NÍVEIS E FORMAS DE SENSORIAMENTO REMOTO E TÉCNICAS DE APRENDIZADO DE MÁQUINA

Resumo

A melhoria das previsões e estimativas de variáveis agronômicas por meio de técnicas não destrutivas é motivo de pesquisas para o avanço do conhecimento dessas relações funcionais. Buscando a melhora da estimativa e previsão de variáveis agronômicas de forma remota, foram desenvolvidos experimentos para as culturas do milho, café e amendoim. No primeiro capítulo apresenta-se à revisão de literatura. No segundo capítulo é apresentado o experimento para cultura do milho no qual objetivou-se prever a produtividade, ao nível de zona de manejo, usando algoritmos de aprendizado de máquina, associados a dados de sensoriamento remoto orbital e de topografia, em que se chegou à conclusão de que índices topográficos aumentam a acurácia da previsão de produtividade de milho, quando associados a bandas espectrais como entrada em modelos de aprendizado de máquina. No capítulo três, foi realizado um experimento na cultura do café, no qual se investigou a possibilidade de determinar o volume de aplicação de defensivos usando o princípio *Tree Row Volume*, baseado no acesso de características físicas de plantas de café, concluindo-se que usando sensoriamento remoto proximal e redes neurais artificiais, é possível estimar o volume de dossel de plantas de café com razoável acurácia, e que isso pode ser feito por meio de um modelo de rede neural multicamadas, utilizando índice de vegetação de diferentes partes da planta como entrada do modelo. No quarto capítulo, no experimento da cultura do amendoim, avaliou-se se redes neurais artificiais e múltiplos índices de vegetação modificados e não-modificados são capazes de estimar de forma não destrutiva a maturação do amendoim, sendo que as várias bandas espectrais derivadas de imagens de drone, juntamente com os parâmetros da rede resultante, permitiram a estimação da maturação de amendoim determinada em campos comerciais com elevada acurácia e de forma não destrutiva. Por fim, no capítulo cinco, tem-se as considerações finais com as implicações das pesquisas realizadas assim como os próximos passos para pesquisas futuras.

Palavras-chave: Agricultura digital, topografia, agricultura de precisão, sensoriamento de cultivos, modelagem de cultivo

FORECAST AND ESTIMATION OF CULTIVATION VARIABLES USING REMOTE SENSING LEVELS AND FORMS AND MACHINE LEARNING TECHNIQUES

Abstract

The improvement of forecasts and estimates of agronomic variables through non-destructive techniques is the reason for research to advance the knowledge of these functional relationships. Seeking to improve the estimation and prediction of agronomic variables remotely, experiments were developed for the crops of corn, coffee and peanuts. The first chapter of this thesis presents the literature review. The second chapter presents the experiment for maize culture in which the objective was to predict productivity, at the level of the management zone, using machine processing algorithms, associated with data from orbital remote sensing and topography, in which the completion of topographic indexes increases the accuracy of the corn yield forecast, when associated with spectral bands as input into machine learning models. In third chapter, an experiment was carried out on coffee culture, in which the possibility of determining the volume of application of pesticides was investigated using the Tree Row Volume principle, based on the access of physical characteristics of coffee plants, concluding that using proximal remote sensing and artificial neural networks, it is possible to estimate the canopy volume of coffee plants with reasonable accuracy, and that this can be done using a multilayer neural network model, using vegetation index of different parts of the plant as input of model. In the fourth chapter, in the peanut culture experiment, it was evaluated whether artificial neural networks and multiple modified and unmodified vegetation indices are capable of estimating non-destructively the peanut maturation, with the various spectral bands derived from images drone, together with the parameters of the resulting network, allowed the estimation of peanut ripeness determined in commercial fields with high accuracy and in a non-destructive way. Por fim, no capítulo cinco, tem-se as considerações finais com as implicações das pesquisas realizadas assim como os próximos passos para pesquisas futuras.

Keywords: Digital agriculture, topography, precision agriculture, crop sensing, crop modeling

SUMÁRIO

RESUMO.....	xi
ABSTRACT.....	xii
CAPÍTULO 1 – Considerações gerais.....	1
Introdução.....	1
Hipóteses	2
Objetivos.....	3
Revisão bibliográfica	3
Cultura do milho.....	3
Cultura do Café.....	4
Cultura do amendoim	6
Sensoriamento Remoto	7
Previsão de variáveis agronômicas por sensoriamento remoto	9
Aprendizado de máquinas em agricultura de precisão	11
Referências bibliográficas.....	16
CAPÍTULO 2 - Forecasting corn yield at the management zone level	22
Abstract.....	22
Introduction	22
Materials and Methods.....	24
Study area.....	24
Satellite imagery.....	27
Data processing and building the dataset	27
Yield data	28
Topographical data	29
Dataset	29
Auto-ML.....	30
Model performance analysis	30
Theoretical framework	31
Results	32
Descriptive statistics of corn yield, spectral bands, and TWI	33
Auto-ML for predicting corn yield using TWI and spectral bands	34
Comparison of models using different features in three scenarios in terms of accuracy, relative error, and tendency	41

Discussion	51
Conclusions.....	53
References	54
CAPÍTULO 3 - Determination of application volume for coffee plantations using artificial neural networks and remote sensing	58
Introduction	58
Materials and Methods.....	60
Description of the experimental site.....	60
Proximal remote sensing variables	60
Plant physical variables.....	61
Neural Network Model Description	63
Multi-layer perceptron (MLP) neural network	63
Radial basis function (RBF).....	64
Topology optimization for estimation of plant height and diameter.....	64
Models performance analysis	65
Tree row volume (TRV) and variable rate approach.....	66
Results	67
Discussion	73
Conclusions.....	75
References	76
CAPÍTULO 4 - Mapeando a maturação abaixo do solo usando UAV: estudo de caso em amendoim	79
Introdução.....	79
Metodologia	81
Campo experimental e mensuração da refletância da cultura.....	81
Índices de vegetação	82
Análise estatística	83
Modelagem da maturação da cultura por índices de vegetação.....	83
Rede MLP	85
RBF.....	85
Banco de dados para treinamento e validação	86
Avaliação do desempenho dos modelos	87
Resultados	87
Conclusões	97

Referências	97
Material suplementar	101
CAPÍTULO 5 - CONSIDERAÇÕES FINAIS.....	108

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

Introdução

Esta tese está estruturada em cinco capítulos, que abordam técnicas de sensoriamento remoto para prever variáveis agronômicas de culturas agrícolas. Este estudo é fundamental para o avanço da aplicação de técnicas de aprendizado de máquina na agricultura, pois traz novas abordagens metodológicas para previsão de produtividade de milho, determinação de volume de aplicação de defensivos para cultura do café e estimativa da maturação de amendoim.

A melhoria das previsões e estimativas de variáveis agronômicas por meio de técnicas não destrutivas é um aspecto essencial quando o objetivo é aprimorar o conhecimento dessas relações funcionais. Dentre as técnicas não destrutivas de estimativa e previsão de variáveis agronômicas, o sensoriamento se destaca por apresentar diversidade de sensores, níveis de aquisição e qualidade na captura de informações das plantas. Esforços científicos têm se concentrados no descobrimento de relações de informações espectrais com variáveis de plantas, em que técnicas de aprendizado de máquina são utilizadas para o processamento não somente de grandes volumes de dados, mas também da diversidade dos dados de sensores remotos. Desta forma, visando melhorar a previsão de variáveis agronômicas, acredita-se que o sensoriamento remoto e técnicas de aprendizado de máquinas possam ser ferramentas capazes de melhorar a previsão e estimativa de variáveis agronômicas.

No primeiro capítulo, é apresentado o referencial teórico que suporta as hipóteses de que o sensoriamento remoto pode ser aplicado na estimativa de variáveis agronômicas, especificamente na previsão da produtividade, determinação do volume de aplicação de defensivos, e estimativa da maturação de amendoim.

Baseado em sensoriamento remoto orbital, foi desenvolvido, no segundo capítulo, um método para estimativa da produtividade de milho, ao nível de zonas de manejo, utilizando algoritmos de aprendizado de máquina, bandas espectrais e índices topográficos. Para este capítulo, foram utilizados dados de um campo comercial de produção de milho nos Estados Unidos.

No terceiro capítulo são apresentados os resultados da utilização de sensoriamento remoto ativo, associado a redes neurais artificiais, para a determinação

do volume de aplicação de defensivos na cultura do café. Nesse capítulo, foi utilizado o princípio do Tree Row Volume (TRV) para desenvolver um método não destrutivo de determinação de volume de aplicação variado para a cultura do café.

Dados provenientes de lavouras comerciais de amendoim localizadas no estado da Georgia, Estados Unidos, de áreas irrigada e não irrigada, foram utilizados no Capítulo 4 visando à determinação da maturação de amendoim por meio de imagens multiespectrais de plataforma aérea. Por fim, no quinto capítulo são apresentados os melhores resultados dos experimentos desenvolvidos na tese, assim como as limitações encontradas e as diretrizes para pesquisas futuras.

Hipóteses

O desenvolvimento desta tese baseou-se em quatro hipóteses, descritas a seguir:

1. Redes neurais e algoritmos de aprendizado de máquina associados às técnicas de sensoriamento remoto podem ser empregados na determinação e estimativa de variáveis agronômicas.
2. Algoritmos de aprendizado de máquinas e características topográficas, utilizando dados de zonas de manejo, podem aumentar a acurácia da previsão de produtividade de milho quando comparados com algoritmos baseados em dados da área inteira.
3. Redes neurais artificiais, associadas ao sensoriamento remoto proximal utilizando o TRV, podem determinar o volume de aplicação de defensivos para a cultura do café.
4. Devido à capacidade de aprendizado de redes neurais, é possível estimar a maturação da cultura do amendoim por meio de índices de vegetação obtidos por sensoriamento remoto aéreo.

Objetivos

Objetivou-se neste trabalho empregar e avaliar técnicas de inteligência artificial e sensoriamento remoto na determinação e estimativa de variáveis agronômicas.

Prever a produtividade ao nível de zona de manejo usando algoritmos de aprendizado de máquinas, associados a dados de sensoriamento remoto orbital e de topografia.

Investigar a possibilidade de determinar o volume de aplicação de defensivos usando o princípio *Tree Row Volume*, baseado no acesso de características físicas de plantas de café.

Avaliar se redes neurais artificiais e múltiplos índices de vegetação modificados e não-modificados são capazes de estimar de forma não destrutiva a maturação do amendoim.

REVISÃO BIBLIOGRÁFICA

Cultura do milho

O milho pertence à ordem *Gramineae*, subfamília *Panicoideae*, gênero *Zea*, espécie *Zea mays* L., e apresenta grande aplicabilidade na alimentação humana e animal (Souza et al., 2012; Torres et al., 2014). O milho tem sua origem nas Américas e é cultivado nos mais diversos ambientes e climas (Werle et al., 2011). Devido à sua grande adaptabilidade, o milho se tornou uma das culturas mais importantes do mundo, podendo ser encontrado em diversos tipos de formatos e cores (Schmitt, 2014). Descreve-se que a origem do milho ocorreu a partir da domesticação do teosinto a cerca de 10 mil anos atrás, na América central (Schmitt, 2014).

O milho possui grande importância científica e econômica, devido principalmente às propriedades nutritivas encontradas em seus grãos, servindo para o emprego na alimentação animal e humana, além de fornecer matéria para indústria (Souza et al., 2012; Torres et al., 2014). Esse cereal é principalmente destinado à alimentação de animais não ruminantes como aves e suínos com certa de 70 – 80% de sua produção mundial, compondo em sua maioria rações (Alves et al., 2015).

O potencial produtivo é uma das variáveis agrônomicas de maior interesse por parte dos produtores. No entanto, este potencial é resultado de diferentes processos do sistema produtivo como densidade de semeadura, população final de plantas, número de espigas por planta, comprimento de espigas, número de fileiras de grãos por espiga, número de grãos por fileira, massa média do grão e pelo genótipo utilizado (Balbinot, et al., 2005). Além de parâmetros relacionados aos aspectos da cultura

semeada, sua interação com o meio, seja ele o solo e ou o clima, são responsáveis pela determinação de 60% do potencial produtivo (Resende, 2003). A variabilidade espacial da produtividade de grãos de milho possui relação de fatores que interagem de forma complexa e condicionam a expressão da cultura (Vian et al., 2016).

Cultura do Café

O café pertence à família das Rubiaceas e gênero *Coffea*, sendo que é possível encontrar cerca de 120 espécies deste gênero, em termos de representação comercial duas espécies se destacam a *Coffea arábica* L. (café arábica) e *Coffea canephora* Pierre (café conilon) (Davis et al., 2011). São produzidas no mundo 175,5 milhões de toneladas de café (USDA, 2017). O Brasil é o maior produtor mundial de grão de café (FAOSTAT, 2018). Esta cultura da produção de café para grãos representa parte fundamental da balança comercial brasileira (Coelho et al., 2009).

No ano de 2019, a produção estimada de café foi de 45 milhões de sacas de 60 kg beneficiadas (CONAB, 2020). A área cultivada com espécies cafeeira é em torno de 2,2 milhões de hectares (CONAB, 2020). Os principais estados produtores de café são: Minas Gerais, São Paulo, Espírito Santo, Bahia e Paraná. Em São Paulo foi estimada produção de seis milhões de sacas de café beneficiado (CONAB, 2020).

A cultura do café demanda grande quantidade de manejos para altas produtividade e qualidade de frutos. Dentre as operações de manejo a aplicação de produtos fitossanitários é de suma importância para a certificação (Rede de Agricultura Sustentável, 2014). Para racionalização da aplicação de produtos fitossanitários, é necessário o emprego correto da tecnologia de aplicação para que o ingrediente ativo seja depositado no alvo no momento adequado, na quantidade certa, com mínimo de desperdício e contaminação ambiental (Matthews, 2008).

A forma convencional de aplicação de produtos fitossanitários em lavouras de café considera-se a quantidade de volume aplicado por área de lavoura, esta forma independe do volume vegetativo (Sousa Junior et al., 2017). Esta forma de aplicação pode proporcionar o excesso de calda aplicada em lavoura mais novas, devido à pouca área foliar, o que pode ocasionar escoamento superficial e contaminação do solo (Matuo, 1990). Formas alternativas de aplicação de produtos fitossanitários vem sendo propostas como a variação em função do volume vegetativo (Gil et al., 2007). O volume vegetativo é influenciado por diversos fatores, como idade das plantas,

desfolha decorrente da colheita (Santinato et al., 2014), ataque de pragas e doenças (Matielo et al., 2010), variações sazonais em decorrência da disponibilidade hídrica (Damatta et al., 2007). As variações do volume vegetativo têm influência direta nas operações de aplicação de produtos fitossanitários, principalmente na penetração, distribuição e deposição de calda (Silva et al., 2014)

Cultura do amendoim

O amendoim pertence à família Fabaceae e ao gênero *Arachis*. O amendoim tem sua origem na América do sul, ocorre no Brasil, Paraguai, Bolívia, Argentina e Uruguai (Krapovickas e Gredory, 1994). O amendoim atualmente está presente em todos os continentes do mundo e das 85 espécies existentes o Brasil possui 48 delas em sua flora nativa. Dentre as espécies a *Arachis hypogea* L. é a de maior interesse econômico (Veiga et al., 2001).

A produção mundial de amendoim em 2019 foi por volta de 41,77 milhões de toneladas. Destacam-se como produtores a China, Índia, Nigéria e Estados Unidos, estes países produziram de forma conjunta cerca de 27 milhões de toneladas (USDA, 2019).

Na América do Sul os principais países produtores são a Argentina e o Brasil, no entanto somam apenas 1,47 milhão de toneladas. O Brasil em termos de produção mundial não se coloca de forma expressiva possuindo apenas 1% do total mundial (USDA, 2019). Dentre os estados brasileiros, São Paulo é o estado com maior produção, produzindo entre 70 e 80 % de todo o amendoim nacional (Suassana, et al., 2014). O amendoim é produzido no Brasil principalmente em sistema de preparo convencional no solo, principalmente por não ser a cultura principal das áreas de produção, pois é utilizado em áreas de reforma de canaviais (Ormond et al., 2018).

A maturação das vagens de amendoim é o fator primordial para início da colheita mecanizada da cultura, visando garantir máxima produtividade e redução de perdas. Por se tratar de uma planta com hábito de crescimento indeterminado, o amendoim apresenta vagens em diferentes estádios de maturação (Carley et al., 2008; Dorner, 2008), o que dificulta a identificação da maturação, podendo acarretar prejuízos econômicos para os produtores (Santos et al., 2013).

Uma forma de determinar a maturação de vagens de amendoim é a utilização do Quadro de Maturação (Williams e Drexler, 1981) em conjunto com o índice de

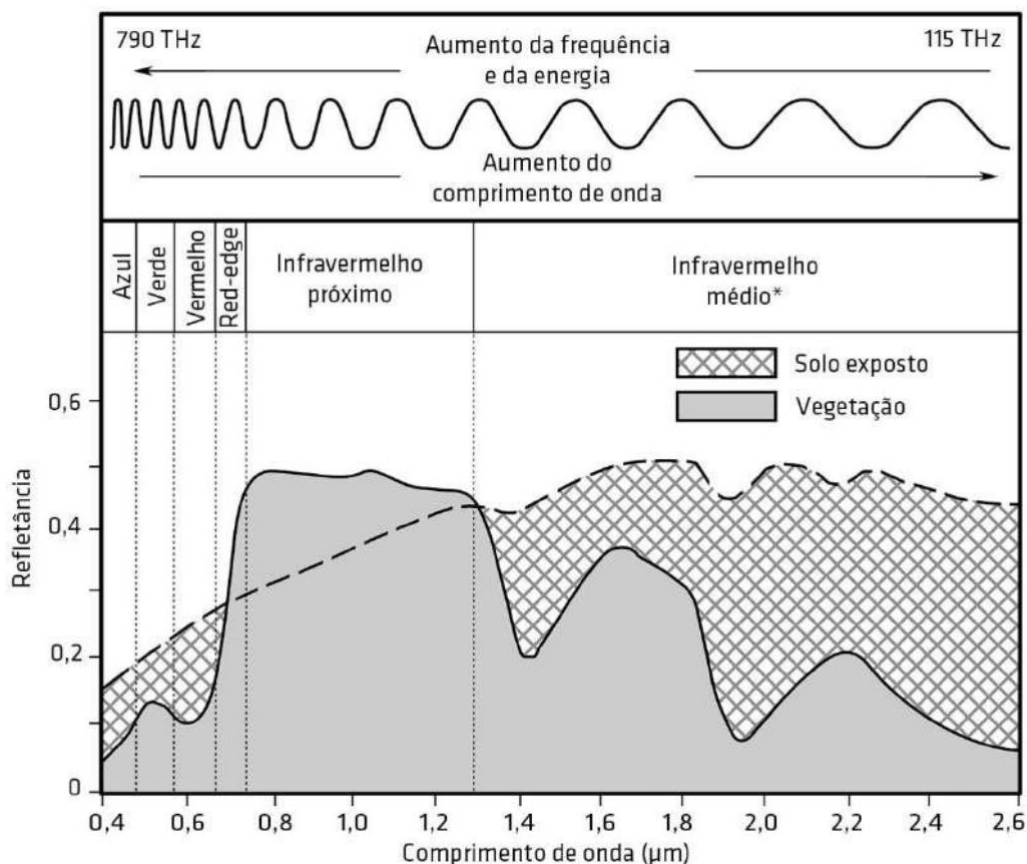
maturação do amendoim (Rowland et al., 2006) em que a amostra atinge maturação ótima quando o índice é de 0,7. No geral, este método é trabalhoso, subjetivo e sujeito a micro variabilidade de campo, que pode afetar na maturidade das plantas individuais. Diante disso, são necessárias alternativas que possam estimar a maturação em campos de produção, ao invés da maturação de plantas individuais, por meio de amostras não-destrutivas.

Rowland et al. (2008) demonstraram que existe correlação das bandas espectrais no infravermelho com maturação de vagens de amendoim utilizando o sensoriamento remoto proximal. Em um estudo em que escala de parcelas Vellidis e Beasley (2013) demonstraram que o Índice de Vegetação Não-Linear (NLI) pode prever a maturação da cultura do amendoim. No entanto, uma abordagem em escala comercial ainda não foi desenvolvida, principalmente modificando índices de vegetação e descobrindo padrões entre índices de vegetação modificados e a maturação das vagens de amendoim.

Sensoriamento remoto

O sensoriamento remoto pode ser definido como uma tecnologia na qual se adquire informações sobre objetos sem que haja contato físico com os mesmos, através do uso de sensores e equipamentos para o processamento de dados (Lillesand e Kiefer, 1994). O sensoriamento remoto presume que cada alvo possui propriedades específicas de interação com a radiação eletromagnética (Molin et al., 2015). Estas propriedades específicas de interação podem ser demonstradas na forma de assinatura espectral (Figura 1).

Figura 1. Comportamento espectral de vegetação e solo exposto em comprimentos de ondas variando do azul ao infravermelho médio



Fonte: (Molin et al., 2015).

Os sensores remotos podem ser classificados quanto à plataforma em que estão acoplados, sendo: orbital, aéreo e terrestre (Moreira, 2011). Sensores orbitais estão geralmente embarcados em satélites, os quais podem ser definidos como sensores imageadores. Já os sensores aéreos destacam-se por apresentarem fotografia aérea, enquanto no sensoriamento terrestre é comum o uso de radiômetro portáteis e espectralradiômetros, em que há a possibilidade de embarcá-los em máquina agrícolas (Moreira, 2011).

Outra classificação importante é quanto a fonte de radiação do sensor, que podem ser sensores ativos e passivos. Os sensores ativos possuem sua própria fonte de radiação, enquanto os sensores passivos não possuem fonte de radiação própria (Moreira, 2011). Dentre os sensores, podemos ter satélites, aeronaves remotamente pilotada e sensores ativos de dossel.

Os satélites orbitam a terra e podem fornecer dados para diferentes tipos de aplicações, dependendo da resolução espacial, espectral, temporal e custos

requeridos (Moreira, 2011). O sensoriamento remoto por satélites oferece grande variedade de parâmetros de cultivos e vegetação, como índice de área foliar (Bouman, 1995; Yao et al., 2015), nitrogênio foliar (Huang et al., 2013), biomassa de cultivo (Bouman, 1995; Mkhabela et al., 2011; Gao et al., 2018), radiação fotossinteticamente ativa (Morel et al., 2014).

O uso de aeronaves remotamente pilotadas vem aumentando devido às suas diversas aplicações (Tsouros et al., 2019; Yang et al., 2017). Essas aeronaves podem ser embarcadas com sensores que capturam a radiação do visível, até sensores hiperespectrais (Amaral et al., 2020). Esta plataforma pode ser usada para o monitoramento do estado de água na planta, por meio de bandas do infravermelho (Kandylakis et al., 2020), biomassa acima do solo, concentração de pigmentos (Buchaillot et al., 2019; García-Martínez et al., 2020).

Os sensores remotos terrestres realizam o sensoriamento remoto proximal, que pode ser realizado com distância de 2 m do alvo (Viscarra Rossel et al., 2011), e podem ser classificados como portáteis ou online (os sensores online medem a refletância enquanto são conduzidos no campo por meio de trator ou quadriciclo) (Mouazen et al., 2020). Estes sensores podem ser usados na detecção de doenças (Muhammed e Larsolle, 2003), aplicação de nitrogênio em taxa variada (Dellinger et al., 2008; Solari et al., 2008; Shiratsuchi et al., 2011), além da detecção da saturação de índices de vegetação (Carneiro et al., 2019).

Previsão de variáveis agronômicas por sensoriamento remoto

Para o manejo eficiente da variabilidade espacial dos campos de produção, dados de alta resolução são necessários para o acesso de parâmetros relacionados ao crescimento de culturas e produtividade no sistema (Mouazen et al., 2020). Isto inclui dados de solo (Kuang et al., 2012), crescimento e saúde da cultura (Gebbers e Adamchuk, 2010), produtividade e dados auxiliares, tais como, condições climáticas e topografia (Nawar et al., 2017). Para o correto manejo essas informações devem ser quantificadas de forma rápida e a baixo custo para viabilizar respostas rápidas de manejo pelos produtores (Mouazen et al., 2020). Isto pode ser feito pelo uso de sensores remotos embarcados em plataformas aéreas, terrestres e orbitais.

Os comprimentos de onda mais utilizados na agricultura são a região do visível e infravermelho próximo (400 nm a 3000 nm). Inferências podem ser feitas dos alvos

agrícolas baseadas em seu comportamento espectral específico e sua interação com as radiações eletromagnéticas nos diferentes comprimentos de onda (Molin et al., 2015). Tipicamente, o sensoriamento remoto de plantas é baseado na mensuração da radiação emitida ou refletida das plantas ao invés da radiação transmitida ou absorvida (Mouazen et al., 2020). Esses processos de refletância, transmitância e absorptância da radiação podem emitir energia pela fluorescência (Apostol et al., 2003) ou emissão termal (Cohen et al., 2005). As plantas têm sua assinatura espectral influenciada principalmente pela clorofila, que apresenta maiores absorções na faixa do vermelho e do azul, enquanto na faixa do verde é onde ocorre sua maior reflexão dentro do espectro visível (Brandão et al., 2008).

A associação de relações funcionais entre variáveis agronômicas de plantas e sensoriamento remoto podem trazer benefícios aos produtores. Esforços científicos vem sendo aplicados para descobrir estas relações, principalmente na previsão de produtividade e estimativa de variáveis morfofisiológicas.

A previsão de produtividade pode ser feita usando diferentes abordagens. Uma das mais comuns é a utilização do sensoriamento remoto. O sensoriamento remoto oferece informações não destrutivas sobre as plantas a baixo custo e alta resolução temporal razoável (Battude et al., 2016; Foster et al., 2019; Veloso et al., 2017). No sensoriamento remoto, três principais modelos são usados: modelos de eficiência de uso da radiação, modelos de simulação de cultivo e modelos empíricos (Yu e Shang, 2018). Modelos empíricos são baseados na relação entre os índices de sensoriamento remoto e a produtividade (Yu e Shang, 2018). Modelos de simulação de cultivo, simulam o crescimento da cultura e a produtividade baseados em índices derivados de sensores remotos (Jin, et al., 2018; Xie et al., 2017). Modelos de eficiência do uso da radiação estimam a produtividade baseado na produtividade primaria bruta ou líquida derivadas de dados de sensoriamento remoto (Liu et al., 2010). Os modelos empíricos assumem que o vigor da planta, detectado pelo sensor remoto está diretamente relacionado a produtividade (Venancio, et al., 2019).

As técnicas de sensoriamento remoto são uma abordagem razoável para demonstrar a relação entre os parâmetros físicos de cultura e índices de vegetação (Carneiro, et al., 2019), e podem ser usadas para estimar parâmetros biofísicos (Myneni, et al., 1995). A altura de planta (Xavier, et al., 2006) e o diâmetro do dossel foliar podem ser estimados por sensoriamento remoto (Yang, et al., 2008).

Dentre as variáveis morfofisiológicas estudadas para demonstração de relações com o sensoriamento remoto está a maturação do amendoim, a maturação de amendoim como o estudo desenvolvido por Rowland et al. (2008), demonstrou existir correlação das bandas espectrais no infravermelho com a maturação de vagens de amendoim utilizando o sensoriamento remoto proximal. Em escala de parcelas Vellidis e Beasley (2013) demonstraram que o Índice de Vegetação Não-Linear (NLI) pode ser satisfatoriamente utilizado na estimativa da maturação da cultura do amendoim. Santos et al. (2019) demonstrou a existência de relação linear entre a maturação de amendoim e índices de vegetação modificados, em escala de campo, sendo o Modified Non-linear Index (MNLI) o índice mais promissor junto ao NLI.

Aprendizado de máquinas em agricultura de precisão

O Aprendizado de Máquinas ou *Machine Learning* (ML) aplicado à agricultura emergiu concomitantemente às tecnologias de *big data* e à computação de alto desempenho, criando oportunidade de extração de conhecimento em ambientes agrícolas (Liakos, et al., 2018). Os algoritmos de ML são abordagens promissoras para análise de dados em larga escala (Aparecido et al., 2020). O ML pode ser definido com o campo científico que fornece, às máquinas, a habilidade de aprender sem serem estritamente programadas (Samuel, 1959). O ML pode ser aplicado em diferentes áreas científicas, tais como, bioquímica (Richardson et al., 2016), medicina (Zhang, et al., 2017), meteorologia (Rhee e Jungho, 2017), ciências econômicas (Bohanec et al., 2018), robótica (Takahashi, et al., 2017), e segurança alimentar (Fragani, et al., 2018).

Os métodos de ML envolvem o processo de aprendizado para um fim específico baseado na experiência (dados de treinamento). Os dados de treinamento consistem em um conjunto de exemplos. Usualmente, um conjunto de atributos é um exemplo individual e estes atributos podem ser chamados de *features* (Liakos, et al., 2018). As *features* podem ser nominais, binárias, ordinais ou numéricas (Liakos, et al., 2018). Algumas terminologias são importantes para compreensão dos algoritmos de ML. Abaixo são descritas algumas terminologias (Tabela 1).

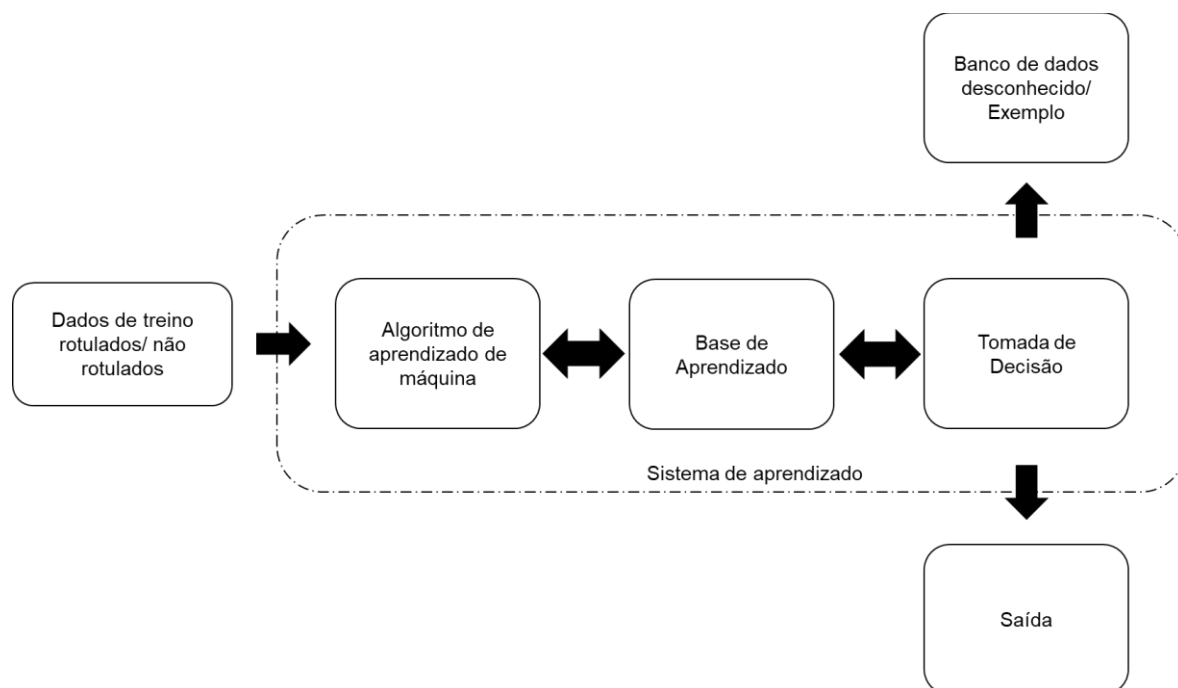
Tabela 1. Terminologias de ML

Termo	Definição
Exemplo (Example)	Instâncias dos dados usados para aprendizado.
Característica (Feature)	Conjunto de atributos ou vetores associados com o exemplo.
Hiperpâmetros (Hyperparameters)	Parâmetros especificados como entrada para o algoritmo de aprendizado.
Conjunto de hipóteses (Hypothesis set)	Conjunto de funções usadas para mapear um conjunto de características com o conjunto de rótulo.
Rótulo (Label)	Valor(es) atribuídos aos exemplos
Função de perda (Loss function)	Uma função que mede a diferença entre o valor real do rótulo e o rótulo predito.
Amostra de teste (Test sample)	Exemplos usados para avaliar o desempenho de um algoritmo de aprendizado.
Amostra de treino (Training sample)	Exemplos usados para o treinamento de algoritmo de aprendizado.
Amostra de validação (Validation sample)	Exemplos usados para ajuste de parâmetros de amostras do aprendizado.

Fonte: (Sharma et al., 2020).

O desenvolvimento de algoritmos de ML geralmente é feito em duas fases, sendo a primeira o treinamento e a segunda o teste. Após a etapa de treinamento o modelo treinado pode ser usado para classificação, predição, ou agrupamento de novos exemplos (fase de teste, Liakos et al., 2018). De forma geral a configuração de sistemas de aprendizado de máquinas pode ser expressa pelo esquema mostrado na Figura 2.

Figura 2. Configuração de sistema de aprendizado de máquina

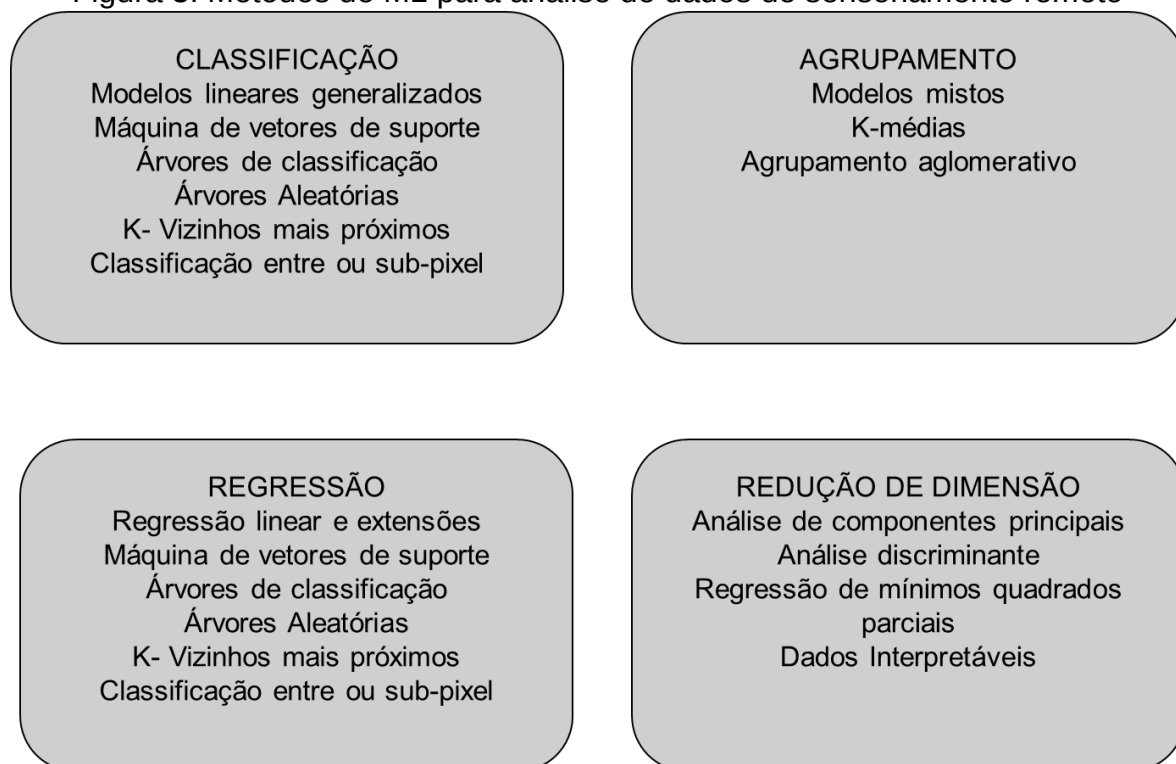


Fonte: (Adaptado de Hu et al., 2017).

O ML aplicado ao sensoriamento remoto pode ser agrupado em quatro principais categorias de acordo com seu objetivo, sendo: classificação, agrupamento, regressão e redução de dimensões (Holloway e Mengerson, 2018). Para cada categoria, existem algoritmos específicos (Figura 3).

Dependendo da tarefa de ML pode-se aplicar técnicas específicas de aprendizado. Existem duas principais categorias de aprendizado, as quais são, aprendizado supervisionado e não supervisionado. No aprendizado supervisionado, os dados são apresentados como exemplo de entrada e sua saída correspondente. O objetivo deste tipo de aprendizado é construir regras gerais entre as entradas e saídas. Já no aprendizado não supervisionado, não existe distinção entre os dados de treinamento e teste e os dados não possuem rótulos (Liakos, et al., 2018).

Figura 3. Métodos de ML para análise de dados de sensoriamento remoto



Fonte: (Adaptado de Holloway e Mengerson, 2018).

Durante o desenvolvimento e aplicação de ML podemos aplicar diversos tipos de algoritmos para determinados fins. Existem algoritmos para redução da dimensionalidade (RD) do banco de dados, a redução do banco de dados com informações representativas do fenômeno estudado visa prover uma representação mais compacta, com menor dimensão do banco de dados e preservar o máximo de informação possível. Estas técnicas podem ser aplicadas tanto no aprendizado supervisionado quanto no não supervisionado (Liakos et al., 2018). Os algoritmos de RD mais comuns são: Análise de componentes principais (Cox, 1958), regressão dos mínimos quadrados parciais (Efroymson, 1960) e análise discriminante linear (Craven e Islam, 2011).

Para análises de agrupamento, ou seja, encontrar grupos naturais de dados (clusters, Tryon, 1957), pode ser aplicado a técnica de K-médias (Lloyd, 1982), técnica hierárquica (Johnson, 1967) e a técnica de maximização de expectativa (Dempster, 1977).

Nas aplicações de ML usando aprendizado supervisionado em problemas de regressão existem diversos algoritmos os principais utilizados na agricultura de precisão são: modelos Bayesianos, árvore de decisão, redes neurais artificiais ANN,

máquinas de suporte de vetores, e modelos de aprendizagem por conjunto (*ensemble learning*) (Liakos et al., 2018).

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, B.M. et al. Divergência genética de milho transgênico em relação à produtividade de grãos e à qualidade nutricional. **Ciência Rural**, v. 45, n. 5, p. 884–891, 2015.

AMARAL, L.R. do et al. UAV applications in Agriculture 4.0. **Revista Ciência Agronômica**, v. 51, n. 5, 2020.

APARECIDO, L.E. de O. et al. Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. **International Journal of Biometeorology**, 2020.

APOSTOL, S. et al. Laser-induced fluorescence signatures as a tool for remote monitoring of water and nitrogen stresses in plants. **Canadian journal of remote sensing**, v. 29, n. 1, p. 57-65, 2003.

BALBINOT J.R., A. et al. Contribuição de componentes de rendimento na produtividade de grãos em variedades de polinização aberta de milho. **Current Agricultural Science and Technology**, v. 11, n. 2, 2005.

BATTUDE, M. et al. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. **Remote Sensing of Environment**, v. 184, p. 668-681, 2016.

BOHANEK, M.; BORŠTNAR, M.K.; ROBNIK-ŠIKONJA, M. Explaining machine learning models in sales predictions. **Expert Systems with Applications**, v. 71, p. 416-428, 2017.

BOUMAN, B.A.M. Crop modelling and remote sensing for yield prediction. **NJAS wageningen journal of life sciences**, v. 43, n. 2, p. 143-161, 1995.

BRANDÃO, Z.N. et al. Agricultura de precisão para gerenciamento do algodão. **AZEVEDO, DMP, BELTRÃO, NEM O Agronegócio do Algodão no Brasil. Cap**, v. 20, 2008.

BUCHAILLOT, M.L. et al. Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques. **Sensors**, v. 19, n. 8, 2019.

CARLEY, D.S. et al. Peanut response to planting date and potential of canopy reflectance as an indicator of pod maturation. **Agronomy Journal**, v. 100, n. 2, p. 376-380, 2008.

CARNEIRO, F.M. et al. Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. **Precision Agriculture**, p. 1-29, 2019.

COELHO, G. et al. Efeito de épocas de irrigação e de parcelamento de adubação sobre a produtividade do cafeeiro 'Catuaí'. **Ciência e Agrotecnologia**, v. 33, n. 1, p. 67-73, 2009.

COHEN, M.J.; PRENGER, J.P.; DEBUSK, W.F. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. **Journal of environmental quality**, v. 34, n. 4, p. 1422-1434, 2005.

CONAB. Companhia Nacional de Abastecimento. Central de informações agropecuárias. Disponível em: Acesso em: 05 mar. 2020.

COX, D.R. The regression analysis of binary sequences. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 20, n. 2, p. 215-232, 1958.

CRAVEN, B.D.; ISLAM, S.M.N. Ordinary least-squares regression. **The SAGE dictionary of quantitative management research**, p. 224-228, 2011.

DAMATTA, F.M. et al. Ecophysiology of coffee growth and production. **Brazilian journal of plant physiology**, v. 19, n. 4, p. 485-510, 2007.

DAVIS, A.P. et al. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. **Botanical Journal of the Linnean Society**, v. 167, n. 4, p. 357-377, 2011.

DELLINGER, A.E.; SCHMIDT, J.P.; BEEGLE, D.B. Developing nitrogen fertilizer recommendations for corn using an active sensor. **Agronomy Journal**, v. 100, n. 6, p. 1546-1552, 2008.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 1, p. 1-22, 1977.

DORNER, J.W. Relationship between kernel moisture content and water activity in different maturity stages of peanut. 2008.

EFROYMSON, M.A. Multiple regression analysis. **Mathematical methods for digital computers**, p. 191-203, 1960.

FAOSTAT. Food and Agriculture Organization of the United Nations. Crops 2018. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 01 de jan. 2021.

FOSTER, T. et al. Assessing landscape scale heterogeneity in irrigation water use with remote sensing and in situ monitoring. **Environmental Research Letters**, v. 14, n. 2, p. 024004, 2019.

FRAGNI, R. et al. Italian tomato-based products authentication by multi-element approach: A mineral elements database to distinguish the domestic provenance. **Food Control**, v. 93, p. 211-218, 2018.

GAO, F. et al. Assessing the variability of corn and soybean yields in central Iowa using high spatiotemporal resolution multi-satellite imagery. **Remote Sensing**, v. 10, n. 9, p. 1489, 2018.

GARCÍA-MARTÍNEZ, H. et al. Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and rgb images acquired with unmanned aerial vehicles. **Agriculture**, v. 10, n. 7, p. 1–24, 2020.

GEBBERS, R.; ADAMCHUK, V.I. Precision agriculture and food security. **Science**, v. 327, n. 5967, p. 828–831, 2010.

GIL, E. et al. Variable rate application of plant protection products in vineyard using ultrasonic sensors. **Crop Protection**, v. 26, n. 8, p. 1287–1297, 2007.

HAMID MUHAMMED, H.; LARSOLLE, A. Feature Vector Based Analysis of Hyperspectral Crop Reflectance Data for Discrimination and Quantification of Fungal Disease Severity in Wheat. **Biosystems Engineering**, v. 86, n. 2, p. 125–134, 2003.

HOLLOWAY, J.; Mengersen, K. Statistical machine learning methods and remote sensing for sustainable development goals: a review. **Remote Sensing**, v. 10, n. 9, p. 1365, 2018.

HUANG, Y. et al. Assimilating remotely sensed information with the wheatgrow model based on the ensemble square root filter for improving regional wheat yield forecasts. **Plant Production Science**, v. 16, n. 4, p. 352-364, 2013.

HU, H. et al. Differentiation of deciduous-calyx and persistent-calyx pears using hyperspectral reflectance imaging and multivariate analysis. **Computers and Electronics in Agriculture**, v. 137, p. 150-156, 2017.

JIN, X. et al. A review of data assimilation of remote sensing and crop models. **European Journal of Agronomy**, v. 92, n. November 2017, p. 141–152, 2018.

JOHNSON, S.C. Hierarchical clustering schemes. **Psychometrika**, v. 32, n. 3, p. 241-254, 1967.

KANDYLAKIS, Z. et al. Water Stress Estimation in Vineyards from Aerial SWIR and multispectral UAV data. **Remote Sensing**, v. 12, n. 15, 2020.

KUANG, B. et al. **Sensing soil properties in the laboratory, in situ, and on-line. A review.** 1. ed. [s.l.] Elsevier Inc., v. 114, 2012.

KRAPOVICKAS, A.; GREGORY, W.C. Taxonomia del gênero "arachis (leguminosae)". **Bonplandia**, p. 1-186, 1994.

LIAKOS, K.G. et al. Machine learning in agriculture: A review. **Sensors**, v. 18, n. 8, p. 2674, 2018.

LILLESAND, T.M.; KIEFER, R.W. Remote sensing and image interpretation. 3rd. ed. New York: John Wiley and Sons, 1994. 750 p.

LIU, J. et al. Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. **Remote Sensing of Environment**, v. 114, n. 6, p. 1167-1177, 2010.

LLOYD, Stuart. Least squares quantization in PCM. **IEEE transactions on information theory**, v. 28, n. 2, p. 129-137, 1982.

MATIELLO, J.B. et al. Cultura de café no Brasil: Novo Manual de Recomendações. Rio de Janeiro e Varginha: MAPA/PROCAFÉ, 2010. 542 p.

MATTHEWS, G.A. Developments in application technology. **Environmentalist**, Paris, v.28, n.1, p.19-24, 2008.

MATUO, T. Técnica de aplicação de defensivos agrícolas. Jaboticabal: Funep, 1990. 139p.

MKHABELA, M.S. et al. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. **Agricultural and Forest Meteorology**, v. 151, n. 3, p. 385-393, 2011.

MOLIN, J.P.; DO AMARAL, Lucas Rios; COLAÇO, André. **Agricultura de precisão.** Oficina de textos, 2015.

MOREIRA, M.A. **Fundamentos do sensoriamento remoto e metodologias de aplicação.** UFV, 2011.

MOREL, J. et al. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. **Remote Sensing**, v. 6, n. 7, p. 6620-6635, 2014.

MOUAZEN, A.M. et al. Monitoring. In: **Agricultural Internet of Things and Decision Support for Precision Smart Farming.** [s.l.] Elsevier Inc., 2020. p. 35–138.

MYNENI, R.B. et al. Optical remote sensing of vegetation: modeling, caveats, and algorithms. **Remote sensing of environment**, v. 51, n. 1, p. 169-188, 1995.

NAWAR, S. et al. Delineation of Soil Management Zones for Variable-Rate Fertilization. In: **Advances in Agronomy**. 1. ed. [s.l.] Elsevier Inc., 2017. 143p. 175–245.

ORMOND, A.T.S. et al. Tillage interference in the quality of peanut mechanized harvest. **Engenharia Agrícola**, v. 38, n. 2, p. 251-259, 2018.

REDE DE AGRICULTURA SUSTENTÁVEL. 2014. Norma para obtenção do selo de certificação Rainforest Alliance: Medidas tomadas para obtenção da certificação. Disponível em: <http://www.san.ag/biblioteca/docs/SAN-S-1-1.2_Sustainable_Agriculture_Standard.pdf>. Acesso em: 06 de Fev, 2021.

RESENDE, M.; FRANÇA, G.E.; ALVES, V.M.C. Cultura do milho irrigado. In: **Embrapa Milho e Sorgo-Artigo em anais de congresso (ALICE)**. In: SIMPÓSIO SOBRE FATORES QUE AFETAM A PRODUTIVIDADE DO MILHO E DO SORGO, 1990, Vitória, ES. Cultura do milho: fatores que afetam a produtividade. Piracicaba: POTAFOS, 1993. p. 237-248., 1993.

RHEE, J.; IM, J. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. **Agricultural and Forest Meteorology**, v. 237, p. 105-122, 2017.

RICHARDSON, A. et al. Clinical chemistry in higher dimensions: machine-learning and enhanced prediction from routine clinical chemistry data. **Clinical biochemistry**, v. 49, n. 16-17, p. 1213-1220, 2016.

Rowland, D.L.; Sorensen, R.B.; Butts, C.L.; Faircloth, W.H.; Sullivan, D.G. Canopy characteristics and their ability to predict peanut maturity. **Peanut Science**, v.1, n. 35, p. 43–54, 2008.

ROWLAND, D. L. et al. Determination of maturity and degree day indices and their success in predicting peanut maturity. **Peanut Science**, v. 33, n. 2, p. 125-136, 2006.

SAMUEL, A.L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, v. 3, n. 3, p. 210-229, 1959.

SANTINATO, F. et al. Análise quali-quantitativa da operação de colheita mecanizada de café em duas safras. *Coffee Science*, v.9, n.4, p.495-505, 2014. Disponível em: <http://www.coffeescience.ufla.br/index.php/Coffeescience/article/view/737/pdf_131> . Acesso em: 24 mar. 2021.

SANTOS, A. F. et al. Using remote sensing to map in-field variability of peanut maturity. In: **Precision Agriculture 2019-Papers Presented at the 12th European Conference on Precision Agriculture, ECPA 2019**. 2019. p. 605-611.

SANTOS, E.P. dos et al. Produtividade e perdas de amendoim em cinco diferentes datas de arranquio. **Revista Ciência Agronômica**, v. 44, n. 4, p. 695-702, 2013.

SCHMITT, A. **Arranjo de plantas para maximizar o desempenho agrônomo do milho em ambientes de alto manejo**. 2014. 226p. Tese (Doutorado em Produção Vegetal) - Universidade do Estado de Santa Catarina, Lages, 2014.

SHARMA, R. et al. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. **Computers & Operations Research**, v. 119, p. 104926, 2020.

SHIRATSUCHI, L.S. et al. Desenvolvimento de um algoritmo baseado em sensores ativos de dossel para recomendação da adubação nitrogenada em taxas variáveis. **Área de Informação da Sede-Capítulo em livro científico (ALICE)**, 2011.

SILVA, B. M. et al. Deposição da calda de pulverização aplicada com pontas de jato plano em diferentes partes da planta de soja (*Glycine max*) e milho (*Zea mays*). **Engenharia na Agricultura**, Viçosa, v. 22, n. 1, p. 17- 24, 2014.

SOLARI, F. et al. Active sensor reflectance measurements of corn nitrogen status and yield potential. **Agronomy Journal**, v. 100, n. 3, p. 571-579, 2008.

SOUSA JÚNIOR, J. M.; RUAS, R. A. A.; SILVA, C. D.; FARIA, V. R.; CARVALHO FILHO, A.; VIEIRA, L. C. Determinação do índice de volume de pulverização para a cultura do café. **Coffee Science**, Lavras, v. 12, n. 1, p. 82-90, 2017.

SOUZA, J. A. de; BUZETTI, S.; TARSITANO, M. A. A.; VALDERRAMA, M. Lucratividade do milho em razão das fontes, doses e épocas de aplicação de nitrogênio. **Revista Ceres**, Viçosa, v. 59, n. 3, p. 321–329, 2012.

SUASSUNA, T. M. F.; ANDRADE, F. P.; SUASSUNA, N. D. (2014). Regiões produtoras de amendoim no Brasil. In: SUASSUNA, T. M. F. Sistema de Produção de Amendoim. Embrapa Algodão. Disponível em: <<https://www.spo.cnptia.embrapa.br/>>. Acesso em: 02 janeiro 2021.

TAKAHASHI, K. et al. Tool-body assimilation model considering grasping motion through deep learning. **Robotics and Autonomous Systems**, v. 91, p. 115-127, 2017.

TRYON, R.C. Communalities of a variable: Formulation by cluster analysis. **Psychometrika**, v. 22, n. 3, p. 241-260, 1957.

TORRES, F. E.; SOUZA, L. C. F.; ANDRADE, L. H. L.; PEDROSO, F. F.; MATOSO, A. O.; TORRES, L. D.; BENETT, C. G. S.; BENETT, K. S. S. Influência da cobertura do solo e doses de nitrogênio na cultura do milho safrinha. **Revista Brasileira de Ciências Agrárias**, Dois Irmãos, v. 9, n. 1, p. 36–41, 27, 2014.

TSOUROS, D.C. et al. A review on UAV-based applications for precision agriculture. **Information**, v. 10, 2019.

UNITED STATES DEPARTMENT OF AGRICULTURE. USDA. Coffee: World Markets and Trade. Foreign Agricultural Service December 2017. Disponível em: <<https://apps.fas.usda.gov/psdonline/circular/coffee.pdf>>. Acesso em: 01 jan. 2021.

USDA. United States Department of Agriculture (2019). Foreign Agricultural Service. World Agricultural Production. **Circular Series**, WAP 4-19, April 2019. Disponível em: <<https://apps.fas.usda.gov/PSDOnline/Circulares/2019/04production.pdf>>. Acesso em: 16 abr. 2019.

VELLIDIS, G.; BEASLEY, J. Using vegetation indices to determine peanut maturity. **Report to the Georgia Agricultural Commodity Commission for Peanuts. GPC**, 2013.

VEIGA, R.F. DE A. et al. Caracterização morfológica de acessos de germoplasma de quatro espécies brasileiras de amendoim-silvestre. **Bragantia**, v. 60, n. 3, p. 167-176, 2001.

VENANCIO, L. P. et al. Forecasting corn yield at the farm level in Brazil based on the FAO-66 approach and soil-adjusted vegetation index (SAVI). **Agricultural Water Management**, v. 225, p. 105779, 2019.

VELOSO, A. et al. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. **Remote sensing of environment**, v. 199, p. 415-426, 2017.

VIAN, A.L. et al. Variabilidade espacial da produtividade de milho irrigado e sua correlação com variáveis explicativas de planta. **Ciência Rural**, v. 46, n. 3, p. 464-471, 2016.

ROSSEL, R.A.V. et al. Proximal soil sensing: An effective approach for soil measurements in space and time. **Advances in agronomy**, v. 113, p. 243-291, 2011.

WERLE, A. J. K.; NICOLAY, R. J.; SANTOS, R. F.; BORSOI, A.; SECO, D. Avaliação de híbridos de milho convencional e transgênico (Bt), com diferentes aplicações de inseticida em cultivo safrinha. **Revista Brasileira de Tecnologia Aplicada nas Ciências Agrárias, Guarapuava**, v. 4, n. 1, p. 150-159, 2011.

WILLIAMS, E.J.; DREXLER, J. Stanley. A non-destructive method for determining peanut pod maturity. **Peanut Science**, v. 8, n. 2, p. 134-141, 1981.

XAVIER, A.C. et al. Hyperspectral field reflectance measurements to estimate wheat grain yield and plant height. **Scientia Agricola**, v. 63, n. 2, p. 130-138, 2006.

XIE, Y. et al. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. **Agricultural and Forest Meteorology**, v. 246, p. 194-206, 2017.

YANG, C.; LIU, T.; EVERITT, J.H. Estimating cabbage physical parameters using remote sensing technology. **Crop protection**, v. 27, n. 1, p. 25-35, 2008.

YAO, F. et al. Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China Plain. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 87, p. 142-152, 2015.

YU, B.; SHANG, S. Multi-year mapping of major crop yields in an irrigation district from high spatial and temporal resolution vegetation index. **Sensors**, v. 18, n. 11, p. 3787, 2018.

ZHANG, B.z et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. **Cancer letters**, v. 403, p. 21-27, 2017.

CAPÍTULO 2 - FORECASTING CORN YIELD AT THE MANAGEMENT ZONE LEVEL

Abstract

Methods using remote sensing associated with artificial intelligence to forecast corn yield at the management level zone (MZ) can help farmers to understand the spatial variability of yield before harvesting. Here, spectral bands, topographic wetness index (TWI) and topographic position index (TPI) were integrated to predict corn yield at the management zone level using machine learning approaches, e.g. extremely randomized trees, gradient boosting machine (GBM), XGBoost algorithms, and stacked ensemble models. We tested the four approaches: only spectral bands, spectral bands + TPI, spectral bands + TWI, and spectral bands + TPI + TWI. We also explored two approaches for model calibration: the whole field approach, and the site-specific model at the MZ level. Models performance were evaluated in terms of accuracy (mean absolute error) and tendency (estimated mean error). The results showed that it is possible to predict corn yield with reasonable accuracy using spectral crop information associated with TWI and TPI during the flowering growth stage. Site-specific models are required to increase the accuracy and reduce the tendency of corn yield forecasting on MZs with high, low, and intermediate yields. The most important features for modeling corn yield were spectral bands + TPI + TWI.

Introduction

Forecasting crop yield at the field level is essential for better management decisions, as well as to assess the economic returns (Venancio et al., 2019) and to make informed decisions regarding soil and plant management (Peralta et al., 2016). Forecast yield is a complex task because yield presents high temporal and spatial variability (Kayad et al., 2019). Soil physical and chemical properties (Al-Gaadi et al., 2018), agricultural management (Yu e Shang, 2018), irrigation, fertilizer management, and topography (Chlingaryan; Sukkarieh; Whelan, 2018) are the main factors associated with yield variability.

Topography is a first-order control of spatial variation in hydrological conditions (Sørensen; Zinko; Seibert, 2006) it affects not only the spatial distribution of soil moisture and groundwater flows (Zinko et al., 2005), but also the water available for

plants (McCann et al., 1996). These hydrological processes associated with terrain attributes directly affect crop yield (Kaspar et al., 2003). One method for describing spatial soil moisture patterns is the topographical indexes (Burt; Butcher, 1986; Moore; Grayson; Ladson, 1991). Different topographical indexes have been proposed, such as the topographic position index (TPI) and the topographic wetness index (TWI) (Moore et al., 1993). TWI describes the water accumulation phenomenon based on the landscape position (Qin et al., 2011; Silva e Alexandre, 2005) and is highly correlated with crop yield (Maestrini e Basso, 2018; Silva e Alexandre, 2005). Thus, TWI can be an important feature for describing the hydrological process in crop yield forecasting.

Another well-known approach that can be used to forecast crop yield, is remote sensing. Remote sensing provides nondestructive information on plants at a low cost and reasonable temporal resolution (Battude et al., 2016; Foster et al., 2019; Veloso et al., 2017). In remote sensing, three main types of models are used: radiation use efficiency models, crop growth models, and empirical models (Yu; Shang, 2018). Empirical models are based on the relationship between remote sensing indexes and crop yields (Yu; Shang, 2018). Crop growth models simulate crop growth and yield based on indexes derived from remote sensing sensors (Jin et al., 2018; Xie et al., 2017). Radiation use efficiency models to estimate crop yield based on gross primary productivity or net primary productivity derived from remote sensing data (Liu et al., 2010). Empirical models assume that the vigor of the plant, detected by the remote sensor, is directly related to the yield (Venancio et al., 2019). These empirical models have spatial and temporal constraints for application in another field or season (Lobell, 2013; Lopresti; Di Bella; Degioanni, 2015).

An alternative method for creating agricultural models based on remote sensing is using machine learning (ML) algorithms (Mas e Flores, 2008; Yuan et al., 2017)^{25,26}. ML is more flexible than traditional models because it makes no assumptions in the underlying function (Stuart e Peter, 2016). Among the ML algorithms, random forest and support vector machine algorithms have been used with reasonable success when using remote sensing data to forecast crop yield because they are not affected by collinearity and non-normal distribution of the variables and can handle overfitting (Kayad et al., 2019), which are common characteristics of spectral bands in satellite images. ML methods are a flexible approach for utilizing many inputs using remote sensing data to forecast crop yield (Ali et al., 2015).

New technologies involving remote sensing, proximal sensing, and geospatial analyses supported by the global position system have led farmers to identify and analyze the temporal and spatial variability of crop fields to optimize resources (Duffera; White; Weisz, 2007; Li et al., 2008). With the knowledge of field variability, subregions with broad similarity can be defined as management zones (MZ) and can support decisions based on this known field variability (Nawar et al., 2017). A model based on a specific MZ might have better accuracy than a model developed for an entire field because with the MZ concept, the model is trained and validated using data from a region where the spatial variability is lower than that of the entire field.

Following this rationale, an ML algorithm based on spectral and topographical features using data from known MZ was used to determine whether it is possible to achieve more accurate predictions than that from a model developed using data from an entire field. To answer this hypothesis, experiments were undertaken to predict crop yield at the MZ level using ML algorithms associated with remote sensing and topographical data.

Materials and Methods

Study area

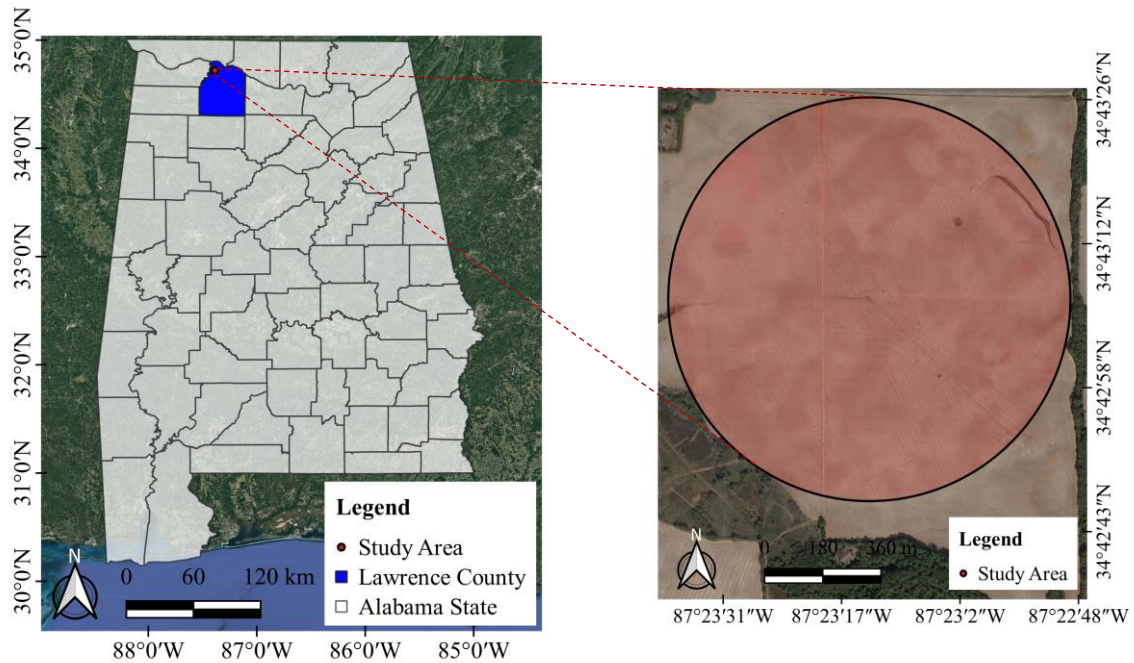
The study was conducted in a commercial cornfield in Lawrence County, Town Creek, Alabama (latitude 34°72' N; longitude 87°39' W) to evaluate the potential of using artificial neural networks associated with high-resolution satellite data and topographical indexes to predict corn yield during the 2018–2019 growing seasons. The predominant soil type in the study field is Decatur silty clay loam with 2%–10% slope variation, and the elevation varies between 169 and 180 m. The field was irrigated using a Reinke® center pivot irrigation system of 623 m length over 125 ha⁻¹ (Figure 1). A description of the sowing, tasseling, and harvest dates for 2018 and 2019 is shown in Table 1.

Table 1. Description of the dates to operation and growth stage of the corn for the years 2018 and 2019.

Year

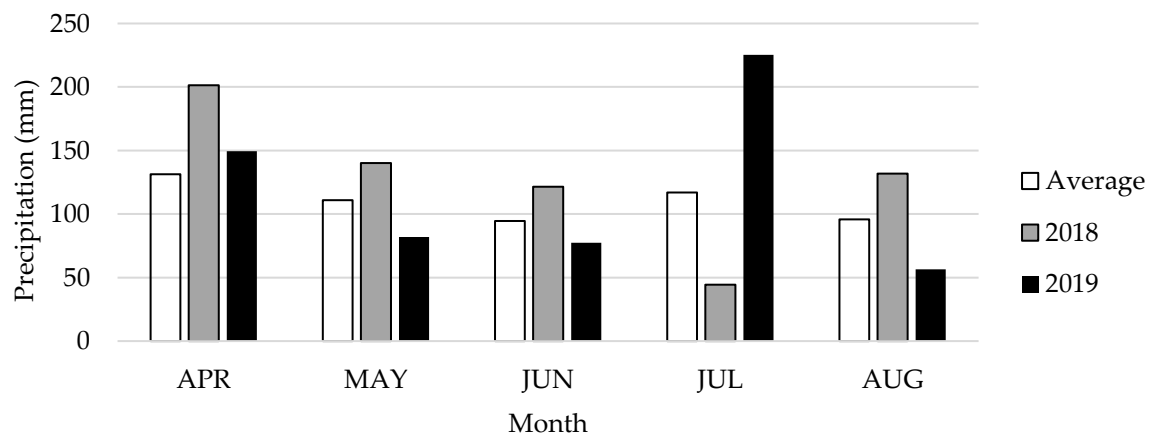
	2018	2019
Sowing	10 / April	27/ March
Tasseling	23 / June	09 / June
Harvest	03 / September	29 / August
Corn Hybrid	Dekalb® DKC 66-97	Dekalb® DKC 66-97
Row spacing	0.76 m	0.76 m
Seed population	84,000 seed/ha ⁻¹	84,000 seed/ha ⁻¹

Figure 1. Study field in Lawrence County, Alabama, US.



The historic average precipitation for the season (April to August) was 520 mm. The total precipitation was above the historic average for 2018 (640 mm) and 2019 (590 mm). In July 2018, and May, June, and August 2019, the precipitation distribution was below the historic average (Figure 2).

Figure 2. Total precipitation (Average from 1999 - 2019, 2018 and 2019) during the growing seasons.



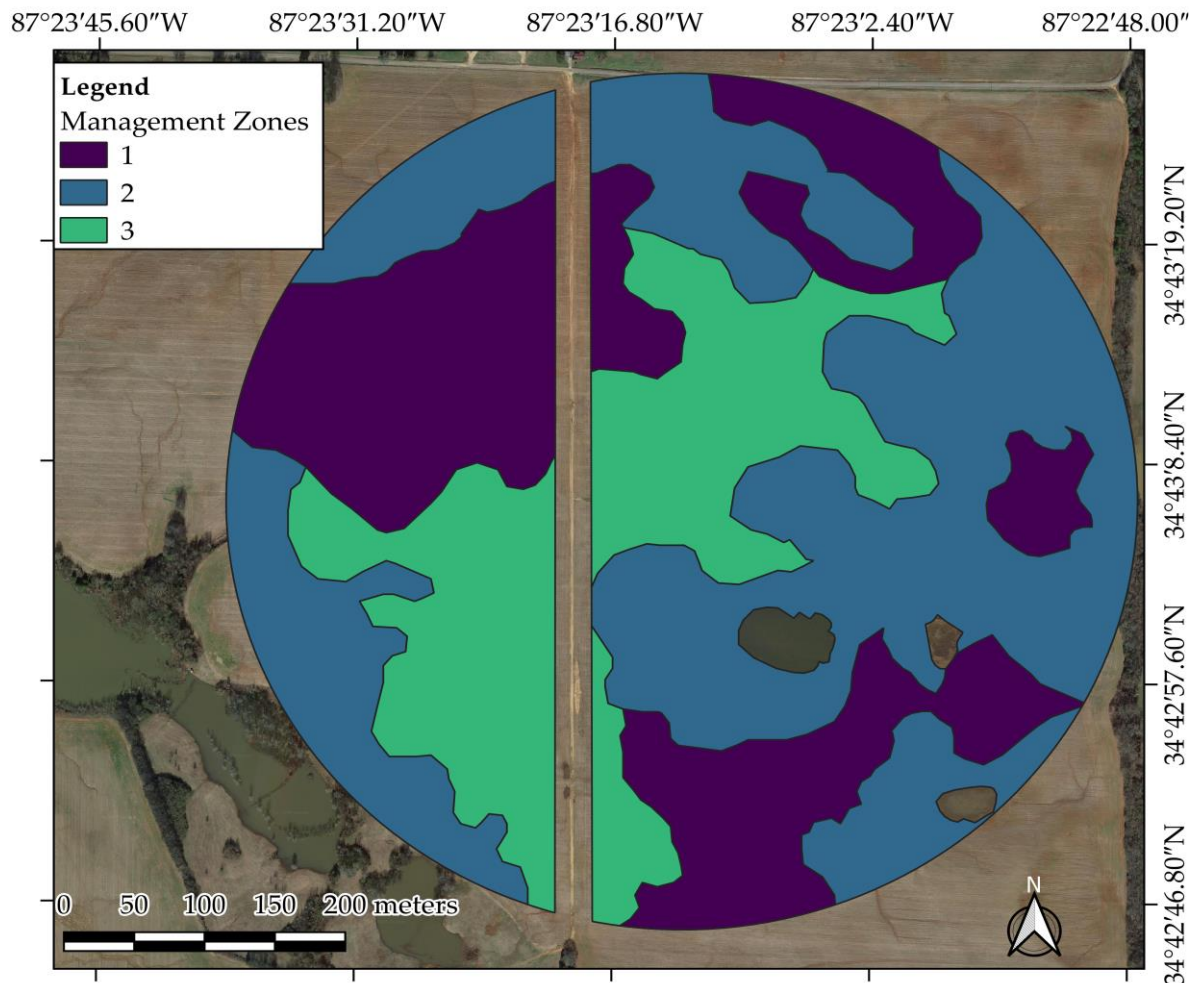
Three MZs were delineated using information about elevation, slope, ten years of normalized yield map and soil texture (Figure 3), this information were used as input into the MZ analyst (MZA) software. This software is used to evaluate the number of homogeneous zones and its delineation (Fridgen et al., 2004). We used the unsupervised algorithm fuzzy c-means to divide the field into three cluster classes. After the delineation of the MZ`s using the variable mentioned previously, they were adjusted based on farmer`s and researcher`s knowledge offield variability, a description of the all methods and validation of the MZ can be accessed in Morata (2020). A description of the soil type and particle size distribution of each MZ is presented in Table 2.

Table 2. Soil proprieties of the study field.

Management Zone	Soil type	Sand%	Silt%	Clay%
MZ01 ¹	Silty Clay	13.07	40.13	46.80
MZ02 ²	Clay Loam	23.20	45.47	31.33
MZ03 ³	Clay Loam	34.40	31.60	34.00

¹ Management zone one, ² Management zone two, ³ Management zone three

Figure 3. Polygons of the management zones delineated on the study area.



Satellite imagery

The bottom of the atmosphere reflectance Ortho Scene product was acquired from PlanetScope, Planet Labs, Inc., San Francisco, USA (Planet, 2016). Cloud Planet Scope satellite data provide 3 m spatial resolution images. The satellite data were collected during the critical period for corn yield determination (approximately 20 days before and 20 days after flowering) (Johnson e Mueller, 2010; Peralta et al., 2016; Sakamoto; Gitelson; Arkebauer, 2014). The PlanetScope satellite data used had four spectral bands: blue (455–515 nm), green (500–590 nm), red (590–670 nm), and near-infrared (NIR, 780 – 860 nm) in a 16-bit GeoTiff format.

The spectral bands images were carefully selected for days with 0% cloud over the study area. The images were used as input to calibrate and validate the models.

Data processing and building the dataset

Yield data

We developed an approach to correct and select an adequate quality yield monitoring and satellite data. The yield data were submitted to a filter process to remove global and local outliers. Outlier yield values were those out of the mean ± 3 standard deviation (SD) range. Other filter procedures such as flow and moisture delay, maximum and minimum yield, and velocity, and start and end pass delays were performed using Yield Editor (Sudduth and Drummond 2007). In a commercial cornfield, it is almost impossible to harvest the entire field using only one combine when the farmer has more than one combine. In our study area, the field was harvested using three combines (C1, C2, and C3), and we developed a method to correct the data. First, we collected data to represent the spatial variability of the yield and the parts of the entire area. All data were collected in parallel passes in which all three combines were harvested side by side. To achieve the best quality, we applied statistical control process techniques. This method has been applied previously to evaluate the mechanical harvest process (Menezes et al., 2018; Girio et al., 2019; Tavares et al., 2018)^{34,36,38}. The quality of the combines was evaluated using Shewhart's individual-moving range control charts, in which the central lines correspond to the mean of the calculated values, and the upper and lower control limits (UCL and LCL, respectively) were calculated based on the SDs using the following equations:

$$UCL = \bar{X} + 3\sigma, \quad (1)$$

$$\bar{X} = \frac{(X1+X2+X3....Xn)}{N}, \quad (2)$$

$$LCL = \bar{X} - 3\sigma, \quad (3)$$

where, UCL is the upper control limit, LCL is the lower control limit, $\bar{X}$ is the overall mean, X_n is the value of the sample "n," N is the total sample number, and σ is the standard deviation. When LCL was negative for I charts, a null value was assigned (LCL= 0).

Topographical data

One of the main pieces of information to generate the TWI and TPI is elevation from the study area. This information was obtained using a global navigation satellite system with real-time kinematic signal correction. This system provides an accuracy close to 2.5 cm on the horizontal coordinates and 5 cm on the vertical coordinates. The global navigation satellite system provided geo-object points with X and Y coordinates and elevation. To transform the geo-object into the geo-field (digital elevation model), we used kriging interpolation for the elevation attribute using ArcMap software (version 10.3.1; ESRI®).

The TWI and TPI were calculated using the following equations, whereas the catchment area and slope that are required for the TWI equation were obtained using the interpolated map of elevation in the System for Automated Geoscientific Analyses (version 2.3.2):

$$TWI = \ln \left(\frac{A_s}{\tan \beta} \right), \quad (4)$$

$$TPI = z_0 - \bar{z}, \quad (5)$$

where, A_s is the specific catchment area ($\text{m}^2 \text{m}^{-1}$); β is the slope angle (degrees); z_0 is the central point; and $\bar{z}$ is the average elevation around the central point.

Dataset

After correcting the harvester yield data, the data were interpolated using ordinary kriging to a raster file with the same spatial resolution as the satellite images used (3 m resolution). Then, we collected the centroid of each pixel to build the dataset with the spectral bands and interpolated yield. All pixels were labeled with their respective MZ. To minimize the uncertain transitional areas between zones, a buffer of 10 m was applied. All procedures related to this part of the research were performed using the qGIS software.

Auto-ML

The open-source auto-ML library H2O for the Python programming language was used to solve the fundamental problem of deciding which ML algorithm to use in our experiments (H2O 2020). Extremely randomized trees, gradient boosting machine (GBM), XGBoost algorithms, and stacked ensemble models were tested with grids of hyperparameters defined by the auto-ML algorithm. For more details regarding each algorithm and how the auto-ML approach works, authors are referred to the library documentation (Hall et al., 2019).

The training phase was performed using 5-fold cross-validation of the data. The dataset was divided into 80% for training and 20% for validation of the models. The model with the lowest mean absolute error during the validation phase was chosen as the best model for each experiment. A total of 50 models were trained per experiment, and the best model was chosen to show the results.

Model performance analysis

Experiments were developed to define the optimal features for input in the models. We tested the following four approaches: only spectral bands, spectral bands + TPI, spectral bands + TWI, and spectral bands + TPI + TWI. We also explored two approaches for model calibration: the whole field approach, which involved calibrating and validating a model using data from the entire field, and the site-specific model at the MZ level, which involved calibrating and validating a model using data from constrained within eachone MZ. These experiments were developed in the following three different temporal scenarios: data from the two seasons together (2018 and 2019), only data from 2018, and only data from 2019.

The accuracy and tendency of the proposed models were evaluated by calculating the mean absolute error (MAE) and estimated mean error (EME) according to equations 6 and 7, respectively:

$$MAE = \frac{\sum_{i=1}^n (Y_{est_i} - Y_{obs_i})}{n}, \quad (6)$$

$$EME = \frac{\sum_{i=1}^N (Yobs_i - \bar{Y})}{N} \quad (7)$$

where, n is the number of data, $Yest_i$ is the value of the variable estimated by the network, $Yobs_i$ is the value of the observed variable, and $\bar{Y}$ is the mean estimated value.

The relative percentage error was calculated to understand the accuracy achieved by the addition of topographic components in the models, expressed by the following equation:

$$X = \left| \frac{x1-y}{x1} \right| * 100 \quad (8)$$

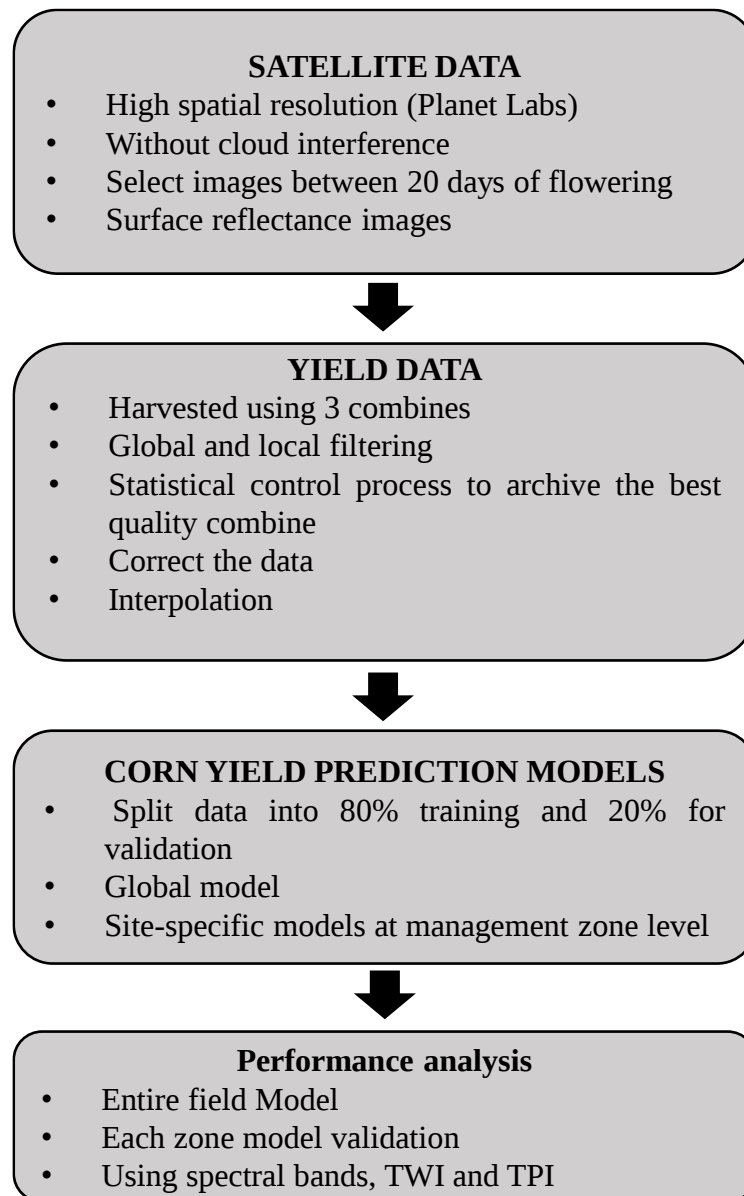
where, X is the relative percentage error, x1 is the MAE of a model with only spectral bands as features, and y is the MAE of the chosen model.

Performance analyses were performed using a scatter plot of predicted values as the dependent variable (y-axis) and the observed values as the independent variable (x-axis), and a line of 1:1 was plotted to determine which model showed values near the line.

Theoretical framework

For better understanding of the global approach used to predict the corn yield, a theoretical framework was developed to illustrate the steps developed in this research (Figure 4).

Figure 4. Theoretical Framework to illustrate the approach developed in this research to predict corn yield.



Results

The following sections present the results from the approach to predict corn yield at the MZ level using spectral bands and topographical indexes. Descriptive statistics were used to characterize the variables during 2018 and 2019. The models were ranked based on the accuracy metric MAE and then the performance of the best models for each experiment was analyzed. The best model from each experiment was used to generate the predicted yield maps to show the spatial variability of the yield.

Descriptive statistics of corn yield, spectral bands, and TWI

Descriptive statistics of corn yield, spectral bands, and topographic indexes for the 2018 and 2019 corn growing are shown in Table 3. Mean yield values among all MZ's was 14.19 Mg ha⁻¹ in 2019 and 13.25 Mg ha⁻¹ in 2018. The highest mean yield was observed in MZ1 (13.25 and 14.19 Mg ha⁻¹ for 2018 and 2019, respectively), whereas MZ2 showed the lowest mean yield (11.37 and 13.05 Mg ha⁻¹ for 2018 and 2019, respectively). Based on these results, we classified each MZ as: MZ1 - high-yielding zone, MZ2 - low-yielding zone, and MZ3 - intermediate yielding zone, independent of the year analyzed, showing the temporal stability of the MZ. It is important noting that MZ2 presented the highest variability in terms of yield (mean C.V. of 14% in both seasons).

Table 3. Descriptive analysis for spectral bands, TWI and yield for the years 2018 and 2019.

2018						2019					
MZ1											
	mean	Std	C.V(%)	min	max		mean	std	C.V(%)	min	max
Blue	5509	72	1	5172	6189	Blue	499	27	5	410	723
Green	4773	58	1	4554	5398	Green	588	32	5	504	819
Red	3238	85	3	2959	4380	Red	557	52	9	443	909
NIR	11441	290	3	8590	12435	NIR	3721	159	4	2983	4210
TWI	6	1	20	2	10	TWI	6	1	18	2	10
TPI	-1	1	-158	-5	5	TPI	-1	1	-151	-5	5
Yield	13.25	1.28	9	5.31	17.49	Yield	14.19	1.14	8	2.89	19.77
MZ2											
Blue	5576	99	2	5231	6904	Blue	529	35	7	382	702
Green	4857	104	2	4567	5897	Green	626	42	7	473	818
Red	3368	168	5	3031	5101	Red	627	79	13	412	981
NIR	11126	423	4	5476	12717	NIR	3686	180	5	1962	4277
TWI	4	1	23	2	9	TWI	4	1	23	2	8

TPI	0	2	413	-9	6	TPI	1	2	309	-9	6
Yield	11.37	1.68	15	3.97	17.75	Yield	13.05	1.68	13	2.35	19.44
MZ3											
Blue	5513	79	1	5176	6965	Blue	510	26	5	431	696
Green	4784	68	1	4574	6136	Green	604	31	5	507	808
Red	3255	96	3	3005	5453	Red	579	50	9	453	979
NIR	11366	273	2	9093	12499	NIR	3715	121	3	3146	4210
TWI	6	1	15	1	9	TWI	6	1	15	1	9
TPI	0	1	-2218	-3	4	TPI	0	1	-2081	-3	4
Yield	13.05	1.14	9	1.41	17.96	Yield	13.99	1.08	8	3.23	19.17

std is standard deviation, C.V. is coefficient of variation, min is the minimum value, and max is the maximum value.

Auto-ML for predicting corn yield using TWI and spectral bands

For the model that considered the dataset from the entire field, the stacked ensemble model type showed better accuracy for predicting corn yield in all scenarios, except in 2019 in which the model type XGBoost best predicted yield, when spectral bands were used as features in the model; this same pattern was also observed for MZ2. For MZ1 and MZ3, GBM was more accurate for predicting corn yield when using only spectral bands as features in all scenarios. In contrast, for MZ2 using other features and scenarios, the stacked ensemble model type was more accurate, except when both years were combined and spectral bands + TPI were used as features, in which case the XGBoost model was preferable (Table 4).

The stacked ensemble models showed better accuracy for predicting corn yield independently of the approach of a general model for all areas and specific models for each MZ when spectral bands + TWI were used as features (Table 4). Using only spectral bands as features in the models, for MZ1 and MZ3, the GBM algorithm showed better accuracy, whereas for MZ2, MZ3, and all areas, the stacked ensemble algorithm showed better accuracy (Table 4). When using only spectral bands as features for the model, the GBM algorithm best predicted yield in MZ1 and MZ3 for each year and when considered both years. For MZ2, however, the best models for

yield prediction were SE_BF, XGBoost, and SE in 2018, 2019, and the two years combined, respectively. Additionally, MZ2 was found to present higher MAE values for 2018, 2019, and both years when compared to the other MZs using only spectral band as feature. These findings could be associated with the high variability of MZ2 that the prediction yield become less accurate. Accuracy was increased (lower MAE) when at least one topographic index was added as a feature alongside the spectral bands, with the highest accuracy found for the stacked ensemble model in MZ1 for spectral bands + TWI + TPI as features in 2018 (MAE = 1.43).

Table 4. Results for training and validation of best models in terms of Mean Absolut Error (MAE) using data of 2018, 2019 and both years combined.

2018				2019			All years		
Whole Field									
	Model Type	Training	Test	Model Type	Training	Test	Model Type	Training	Test
Bands	SE_BF	12.38	12.65	XGBoost	11.62	11.97	SE	12.08	12.43
Bands + TWI	SE	5.90	10.85	SE	6.94	10.83	SE	9.52	11.33
Bands +TPI	SE	6.58	11.11	SE	6.73	10.90	SE	9.38	11.42
Bands + TPI + TWI	SE	7.50	9.68	SE	6.24	9.80	SE	8.50	10.14
MZ1									
Bands	GBM	10.86	11.33	GBM	10.03	10.27	GBM	10.42	10.93
Bands + TWI	SE	7.28	9.54	SE	5.76	9.28	SE	5.33	9.43
Bands +TPI	SE	2.85	9.19	SE	4.52	9.25	XGBoost	5.31	9.32
Bands + TPI + TWI	SE	1.43	7.02	SE	2.45	7.76	SE	2.81	7.43
MZ2									
Bands	SE_BF	12.37	13.13	XGBoost	13.55	14.36	SE	12.61	13.71

Bands + TWI	SE	5.64	11.22	SE	6.56	12.64	SE	6.66	11.70
Bands +TPI	SE	5.91	11.04	SE	6.11	12.58	SE	6.83	11.64
Bands + TPI + TWI	SE	2.30	8.80	SE	3.29	10.66	SE	6.73	10.09

MZ3

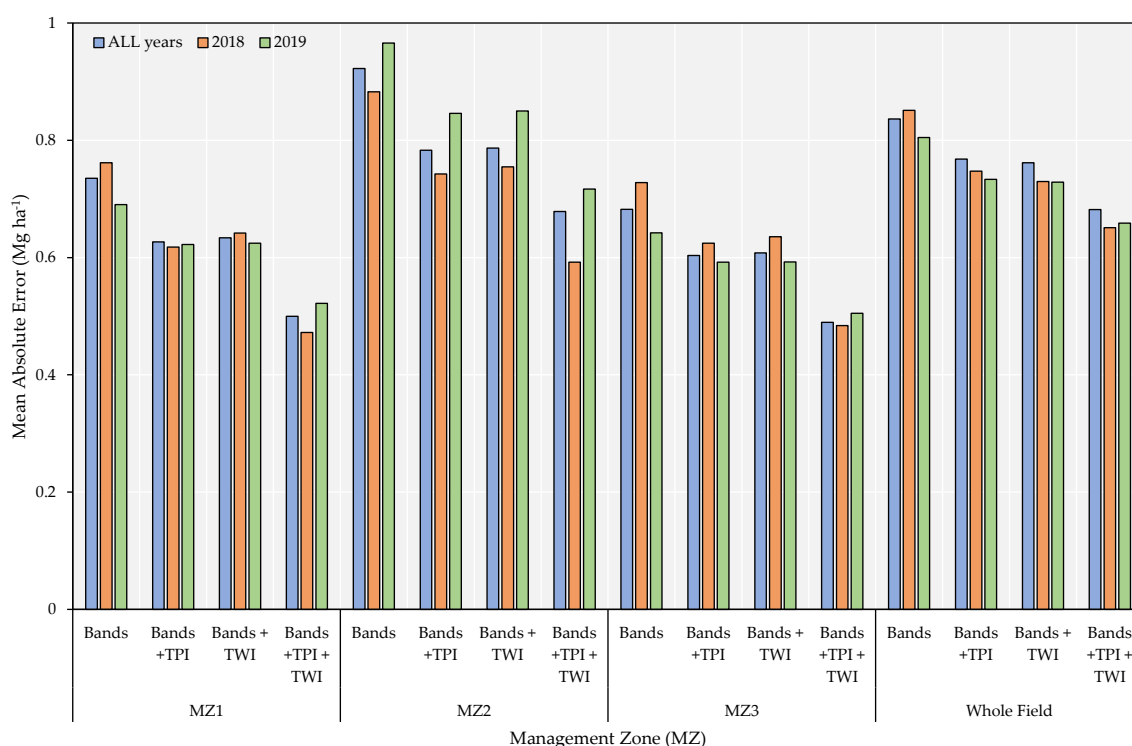
Bands	GBM	10.64	10.83	GBM	9.42	9.55	GBM	10.01	10.15
Bands + TWI	XGBoost	5.72	9.45	XGBoost	4.78	8.81	XGBoost	4.57	9.04
Bands +TPI	SE	3.75	9.29	XGBoost	4.95	8.81	XGBoost	4.77	8.98
Bands + TPI + TWI	SE	1.70	7.20	XGBoost	3.76	7.51	SE	2.85	7.28

SE, stacked ensemble model; GBM, gradient boost machine; TWI, topographic wetness index; TPI, topographic position index; MZ1, MZ2, and MZ3, high, low and intermedium yield zones, respectively.

Comparison of models using different features in three scenarios in terms of accuracy, relative error, and tendency

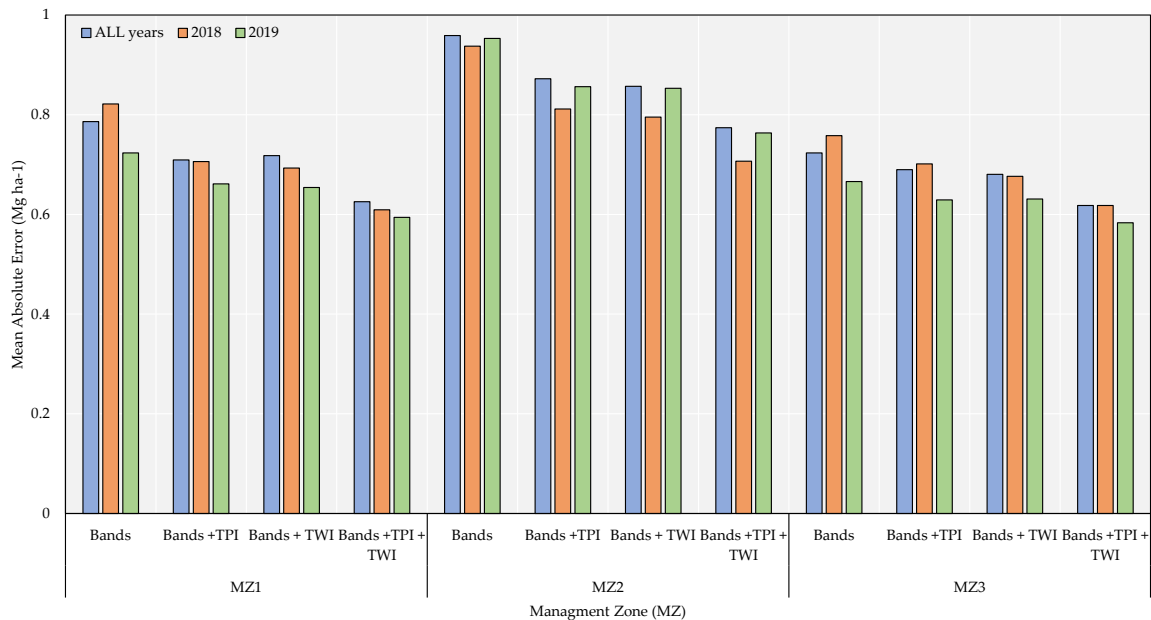
Figure 5 shows the results of the accuracy metric (MAE), as reported in Equation 6. MAE is a parameter that be use to distinguish which feature combination was the most accurate for predicting corn yield. Based on this metric, the models had the lowest accuracy (highest MAE) when only the spectral bands were used, whereas when the models used spectral bands, TWI, and TPI as features, the highest accuracy (lowest MAE) for predicting the corn yield for all scenarios (all years combined, and 2018 and 2019 seasons) was obtained. When the models used TWI or TPI as features, the accuracy was similar within each scenario. The whole field model showed lower accuracy for the feature combinations of spectral bands + TPI, spectral bands + TWI, and spectral bands + TPI + TWI in all scenarios analyzed compared to the MZ1 and MZ3 site-specific models. The MZ2 site-specific model showed a similar accuracy to that of the whole field model.

Figure 5. Mean absolute error (MAE) of the validation phase for models developed for isolated management zones and for whole field, using several features (spectral bands and topographic indexes) in three scenarios (2018, 2019 and all years combined).



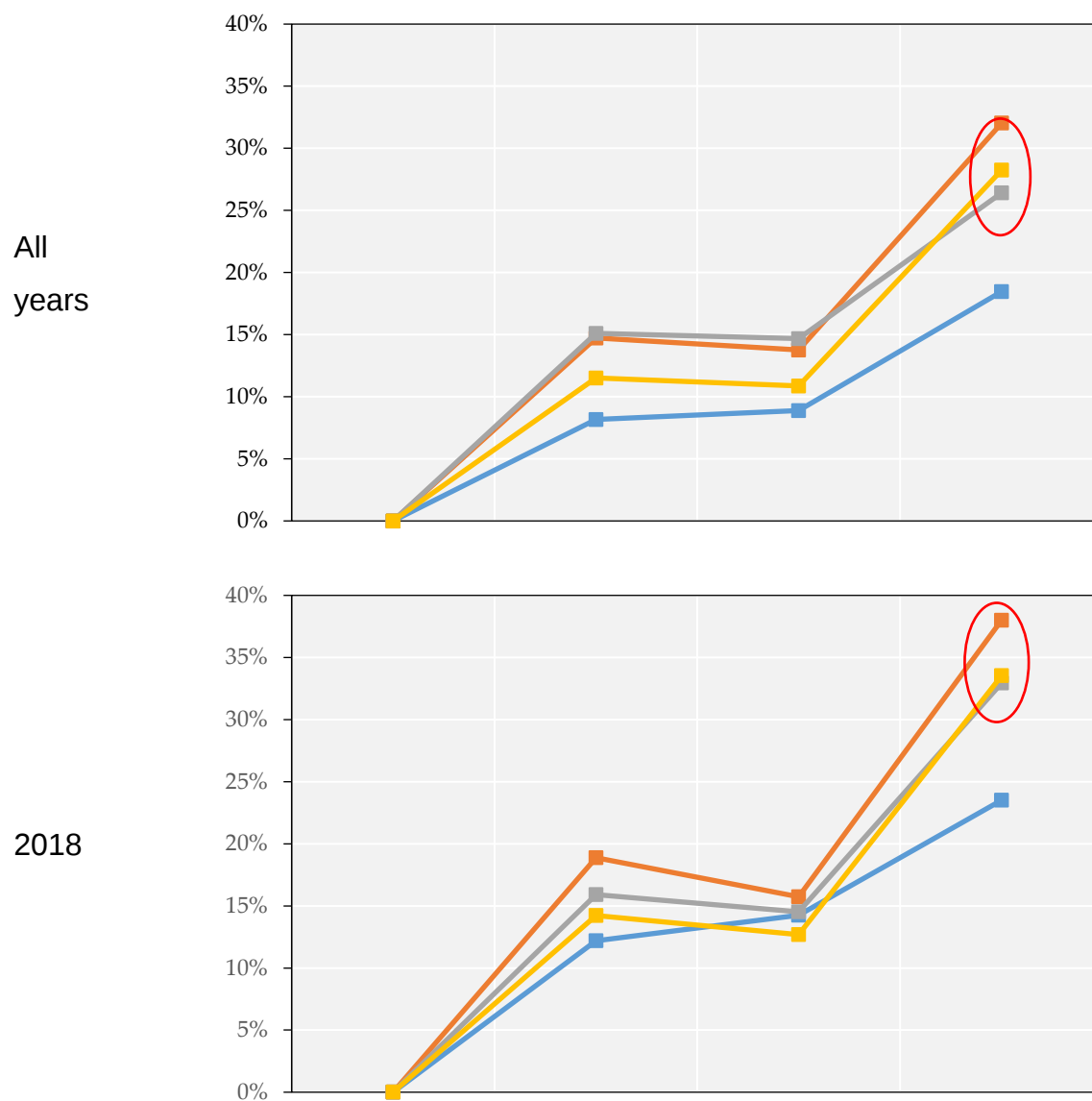
The specific accuracy of the whole field model was less accurate in MZ2 compared to MZ1 and MZ3. These models had difficulty predicting corn yield in low-yield zones (Figure 6). The spectral bands + TPI + TWI features had a higher accuracy than those of the other combination of features.

Figure 6. Accuracy (MAE) of the validation phase the model developed using whole field data, using several features (spectral bands and topographic indexes) in three scenarios (2018, 2019 and all years combined), the accuracy is showed in each management zone. TWI, topographic wetness index; TPI, topographic position index; MZ1, MZ2 and MZ3, management zones one (high yield), two (low yield) and three (intermedium yield) respectively.



To determine whether or not the association between topographical indexes and spectral bands could promote accuracy increase of the yield forecasting, we calculated the relative accuracy increase of the models when only spectral bands were used as features (Figure 7). Analyzing all scenarios, isolated season (2018 and 2019) and all years combined, using spectral bands + TWI + TPI as features obtained a better accuracy, whereas accuracy was higher for site-specific models (Figure 7). The model accuracy was consistently higher not only when both topographic indexes were considered for yield prediction when compared with only spectral band, but also when the simulation was performed for MZ compared with the entire field. For the scenario of the isolated season and site-specific models, the models using spectral bands +

TWI and spectral bands + TPI showed a similar increase in accuracy to that of the models using only spectral bands.



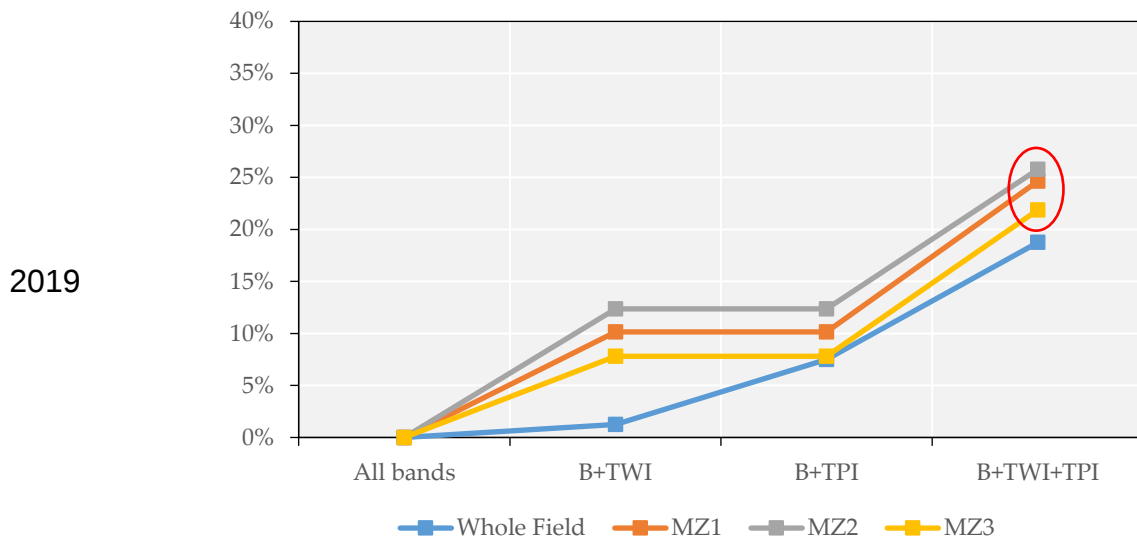
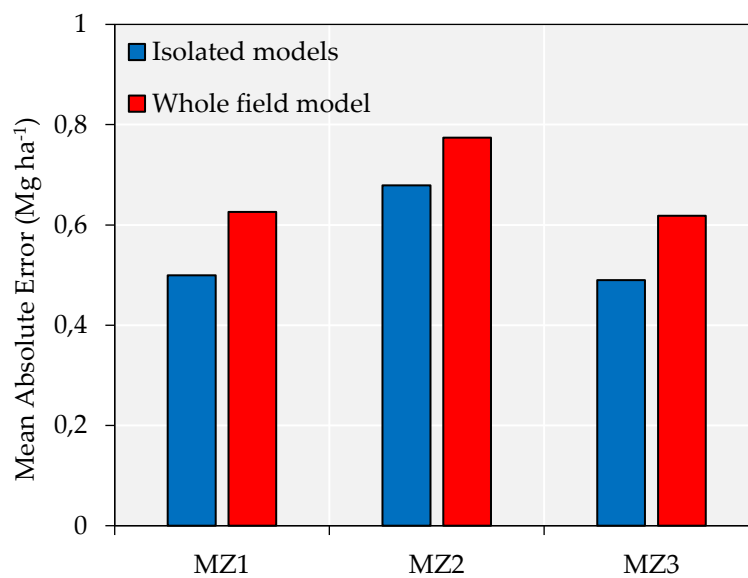


Figure 7. Relative increase accuracy for models developed for isolated management zones and for whole field, using several features (spectral bands and topographic indexes) in three scenarios (all years combined, 2018 and 2019 season). B, spectral bands; TWI, topographic wetness index; TPI, topographic position index; MZ1, MZ2 and MZ3, management zones one (high yield), two (low yield) and three (intermedium yield) respectively. Red open circles represent the isolated models.

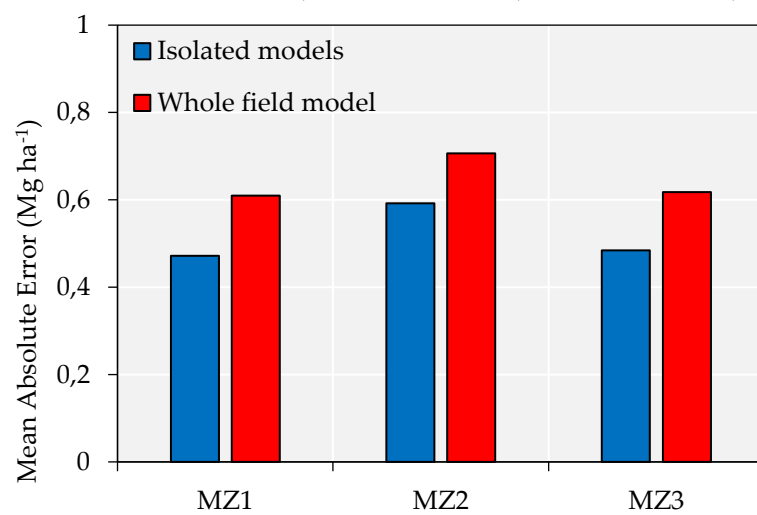
To compare the accuracy of the whole field model and isolated model, we analyzed the specific accuracy of the best feature combination (spectral bands + TPI + TWI) against the accuracy of the specific models for each MZ (high, low, and intermediate yields; Figure 8) for the two scenarios (all years combined and specific season, 2018 and 2019). For all scenarios analyzed, the specific models showed better accuracy (lower MAE) than the whole field model, but for better understanding not only look at accuracy, it is enough to decide if we need site-specific models or a whole field model.

Figure 8. Comparison between the accuracy of the whole field model and specific models for each management zone in three scenarios (all years combine, 2018 and 2019 season). MZ1, MZ2, and MZ3 are, management zones one (high yield), two (low yield), and three (intermedium yield), respectively.

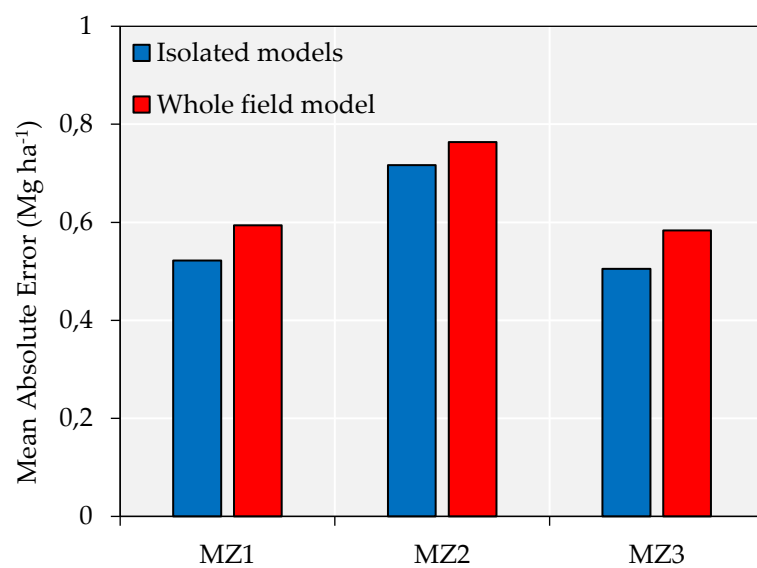
All Years



2018

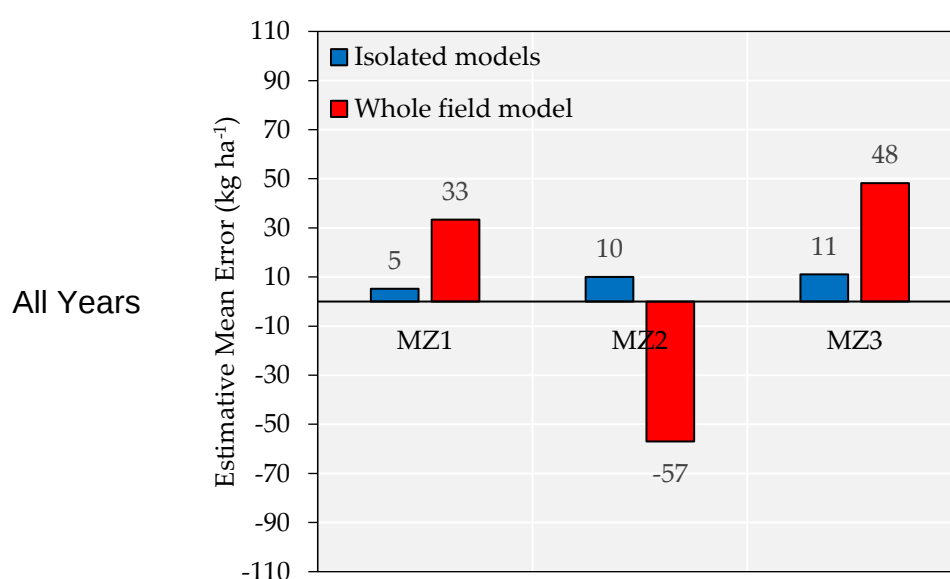


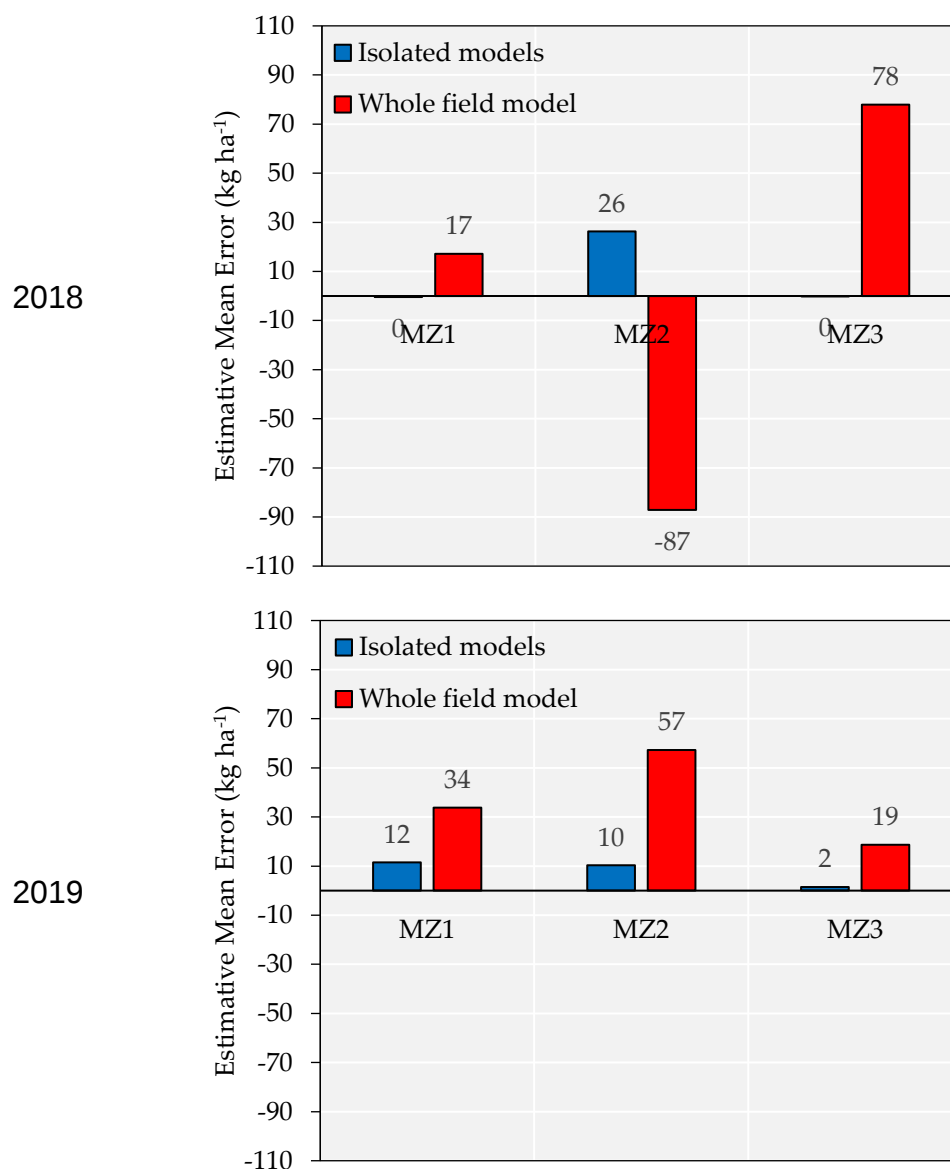
2019



To analyze the tendency of the models, the EME was calculated and compared in the three scenarios (all years combined, 2018, and 2019) using models with spectral bands + TWI + TPI as features. All the comparisons were performed using MZ-based models compared to the whole field-based model, using its tendency in each MZ. The tendency metric allowed us to understand whether the model was underestimating or overestimating the yield prediction and infer which approach had a lower tendency. For all years, the whole field model underestimated the corn yield (-57 kg ha^{-1}) in MZ2 (low-yield zone), and overestimated corn yield in MZ1 (33 kg ha^{-1}) and MZ3 (48 kg ha^{-1}). The MZ1 model (high-yield zone) showed approximately 55% lower tendency than the whole field model in MZ1. This value had an approximately 77% lower tendency for the model trained and validated using only the MZ3 data compared to the whole field model (Figure 9). Underestimation of MZ2 was observed in the 2018 scenario (-87 kg ha^{-1}), whereas in 2019 the whole field model showed an overestimation (57 kg ha^{-1}). The lowest tendency was observed in 2018 using site-specific models for MZ1 and MZ3, which was approximately zero. Based on these results, using site-specific models had lower tendency than using the whole field model.

Figure 9. Comparison between the tendency of the whole field model and specific models for each management zone in three scenarios (all years combine, 2018 and 2019 season). MZ1, MZ2 and MZ3, management zones one (high yield), two (low yield) and three (intermedium yield) respectively.





To analyze the performance of the chosen models, the corn yield forecasted by the resultant model during the validation phase was compared with the observed yield (Figures 10, 11, and 12). The results for the models using combined data and the separate seasons (2018 and 2019) suggested that it was possible to forecast corn yield with reasonable accuracy using data combining spectral bands + TPI + TWI as features. A general behavior, independently of using data from a specific MZ or the entire field, was that the accuracy of the models was lower when predicting the yield in a low-yield area (MZ2). Building a model with data from a high-yield zone (MZ1) and using spectral bands, TPI, and TWI as features was the most accurate for forecasting corn yield (MAE = 0.47).

Our second observation, supporting the hypothesis that site-specific models using spectral bands, TPI, and TWI as features had different accuracies depending on the MZ used for the calibration and validation (MAE MZ1 = 0.50, MAE MZ2 = 0.68, and MAE MZ3 = 0.49), led us to infer that for MZs that represent low-yield zones (MZ2), the corn yield forecasting accuracy was compromised due to lower accuracy (higher MAE and EME) (Figure 10).

Figure 10. Performance analysis of test dataset with the features Spectral bands + TPI + TWI in two approaches of model calibration (a) whole field and (b) site-specific models in the scenario of the all season combined. Observation: MAE (mean absolute error) is expressed in Mg ha^{-1} and EME (estimative mean error) in Kg ha^{-1}

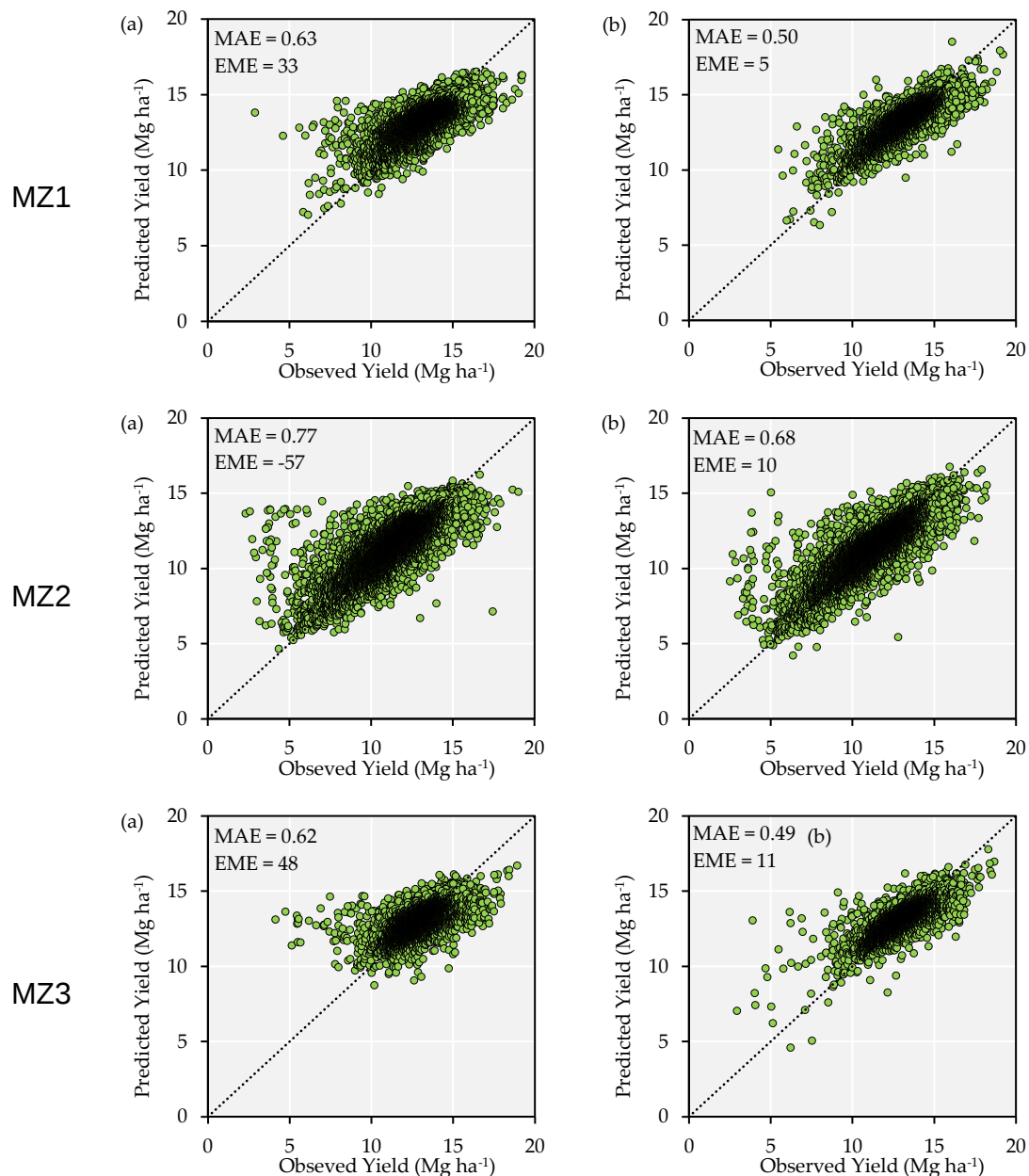


Figure 11. Performance analysis of test dataset with the features Spectral bands + TPI + TWI in two approaches of model calibration (a) whole field and (b) site-specific models in the scenario of the season 2018. Observation: MAE (mean absolute error) is expressed in Mg ha^{-1} and EME (estimative mean error) in Kg ha^{-1}

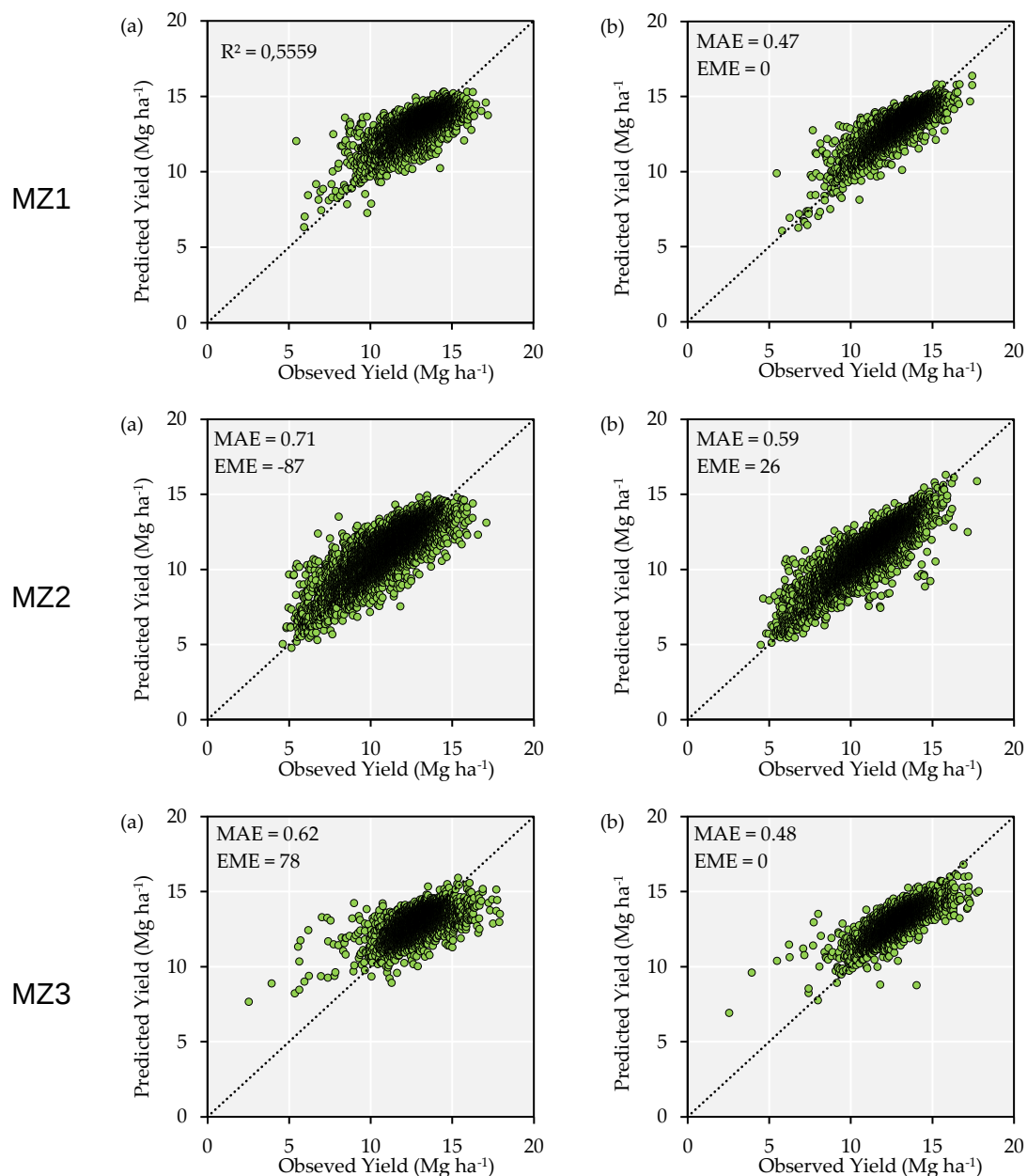
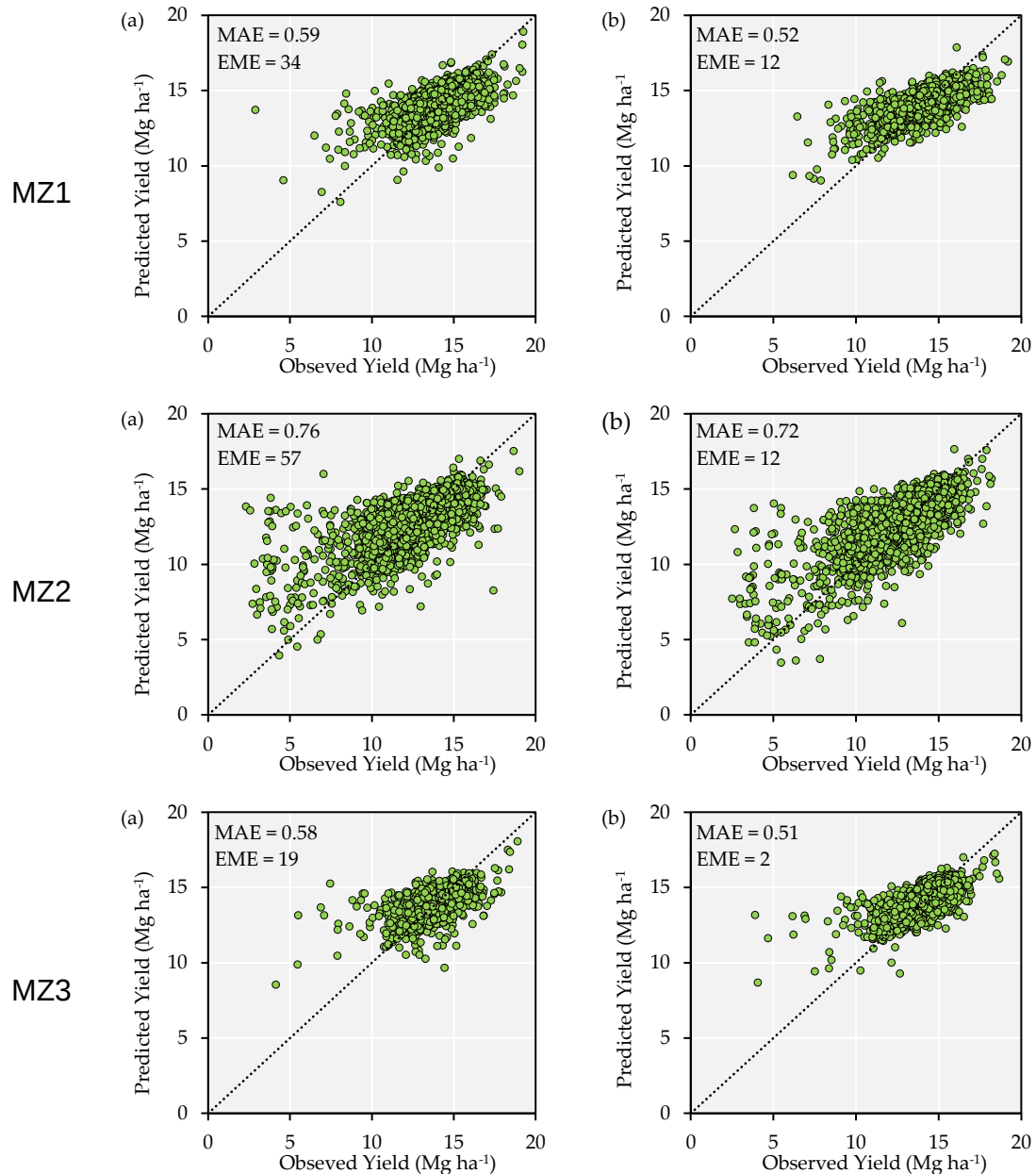


Figure 12. Performance analysis of test dataset with the features Spectral bands + TPI + TWI in two approaches of model calibration (a) whole field and (b) site-specific models in the scenario of the season 2019. Observation: MAE (mean absolute error) is expressed in Mg ha^{-1} and EME (estimative mean error) in Kg ha^{-1}



Discussion

In the present study, a method for developing site-specific models for corn yield modeling at the MZ level was proposed and the results showed that it is possible to predict corn yield with reasonable accuracy using spectral crop information associated with TWI and TPI during the flowering growth stage.

The first and most important outcome from the present study was that it was possible to predict corn yield using only spectral bands. However, to obtain a more accurate model, spectral bands associated with TPI and TWI should be used, which are surface-related variables (Figure 5). The importance of surface structure (slope) variables on crop yield prediction has been shown by (Kross et al., 2020). Surface variables explain yield variability components such as the spatial patterns of soil, water, and nutrient distribution (Turpin et al., 2005). The increase in accuracy from adding the topographical indexes (spectral bands + TPI + TWI as features, Figure 7) can be explained by the addition of deterministic variables (equations of the topographical indexes), which are consistent throughout time. These topographical indexes have been generated from elevation data arising from yield monitoring, and the field elevations are the same unless the farmer modifies the field surface, which is not common. Another fact that can influence the increase in accuracy is that topography influences hydrological processes (Sørensen; Zinko; Seibert, 2006). These processes are related to the spatial distribution of groundwater flow and soil moisture (Zinko et al., 2005). Soil moisture is correlated with TWI reflecting the water availability for crop growth and development (Yang et al., 2021) and can ultimately affect yield. The TPI is a topography-derived index that considers the local topography for a given neighborhood (Szypuła, 2019) and is also correlated with yield (Mieza et al., 2016; Kross et al., 2020). Highlighted the necessity of exploring the inclusion of micro-topography metrics derived from slope data to predict within-field variability accurately during a growing season, and our approach could be a promising application in this field.

The second result from our analysis showed the potential for the development of site-specific models at the MZ level due to a lower tendency (Figure 9) and higher accuracy (Figure 8). Site-specific crop management in terms of resource use efficiency can be enhanced by identifying homogenous MZs (Longchamps e Khosla, 2015). Crop management is changing from large-scale operations to the use of precision agriculture, which requires technologies to forecast yield by considering within-field variation (Peralta et al., 2016). Our approach to modeling corn yield could provide assistance for precision agriculture technologies. These models could predict spatially within-field yield variability. Models that predict within-field yield can be used to characterize the environmental and management factors that contribute to yield variability and are also suitable for capturing critical site-specific factors that drive inter-

region or in-field variability (Kross et al., 2020). Our results in terms of modeling crop yield at the field level agree with several previous studies (Kayad et al., 2019; Kross et al., 2020; Peralta et al., 2016; Schwalbert et al., 2018) that have successfully developed models to forecast yield at the field level. To the best of our knowledge, no other study has attempted to forecast corn yield associated with TWI, TPI, spectral bands, and ML algorithms. There is a demand for crop modeling tools that involve multiple sources of data to improve the prediction accuracy of the crop yield.

From a practical standpoint, farmers can use high-quality data that have already been collected during the harvest (yield and elevation) and derive information regarding the micro-topography (TWI and TPI) associated with this information with spectral bands in high-resolution imagery to predict corn yield during the flowering period. This prediction might allow farmers to determine the spatial within-field variability of the yield before harvesting and make management-informed decisions based on the yield forecast report. The predictive yield maps at the MZ level would be valuable for farmers when planning their MZ land use to achieve their production goals and define in advance the harvester settings. Farmers might also use the higher-quality combine to harvest a high-yield zone.

Conclusions

The present study investigated the potential of using spectral bands of high-resolution satellite images and topographical indexes to create site-specific yield prediction models using auto-ML algorithms. Assuming that the topographical indexes can explain site-specific yield variability, such indexes are important not only for crop yield modeling but also to support management decisions based on yield forecasting. As studies within yield forecasting have focused on large-scale approaches, the present study provides evidence that yield prediction accuracy was improved at the MZ level.

The use of one or more topographical indexes associated with spectral bands as features increased the accuracy of corn yield forecasting when of the auto-ML algorithms models.

Site-specific models are required to increase the accuracy and reduce the tendency of corn yield forecasting on MZs with high, low, and intermediate yield

potentials. The most important features for modeling corn yield were spectral bands + TPI + TWI.

Auto-ML using the stacked ensemble algorithm can be used to forecast corn yield before harvesting using combined data from different seasons.

References

AL-GAADI, K. A.; HASSABALLA, A. A.; TOLA, E. K.; KAYAD, A. G.; MADUGUNDU, R.; ASSIRI, F.; ALBLEWI, B. Characterization of the spatial variability of surface topography and moisture content and its influence on potato crop yield. **International Journal of Remote Sensing**, [s. l.], v. 39, n. 23, p. 8572–8590, 2018.

ALI, I.; GREIFENEDER, F.; STAMENKOVIC, J.; NEUMANN, M.; NOTARNICOLA, C. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. **Remote Sensing**, [s. l.], v. 7, n. 12, p. 16398–16421, 2015.

BATTUDE, M.; AL BITAR, A.; MORIN, D.; CROS, J.; HUC, M.; MARAIS SICRE, C.; LE DANTEC, V.; DEMAREZ, V. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. **Remote Sensing of Environment**, [s. l.], v. 184, p. 668–681, 2016.

BURT, T.; BUTCHER, D. Stimulation from simulation? A teaching model of hillslope hydrology for use on microcomputers. **Journal of Geography in Higher Education**, [s. l.], v. 10, n. 1, p. 23–39, 1986.

CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. **Computers and Electronics in Agriculture**, [s. l.], v. 151, n. November 2017, p. 61–69, 2018.

DUFFERA, M.; WHITE, J. G.; WEISZ, R. Spatial variability of Southeastern U.S. Coastal Plain soil physical properties: Implications for site-specific management. **Geoderma**, [s. l.], v. 137, n. 3–4, p. 327–339, 2007.

FOSTER, T.; GONÇALVES, I. Z.; CAMPOS, I.; NEALE, C. M. U.; BROZOVIĆ, N. Assessing landscape scale heterogeneity in irrigation water use with remote sensing and in situ monitoring. **Environmental Research Letters**, [s. l.], v. 14, n. 2, 2019.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management Zone Analyst (MZA). **Agronomy Journal**, [s. l.], v. 96, n. 1, p. 100, 2004.

GÍRIO, L. A. da S.; SILVA, R. P.; MENEZES, P. C.; CARNEIRO, F. M.; ZERBATO, C.; ORMOND, A. T. S. Quality of multi-row harvesting in sugarcane plantations established from pre-sprouted seedlings and billets. **Industrial Crops and Products**, [s. l.], v. 142, n. October, p. 111831, 2019. HALL, P.; GILL, N.; KURKA, M.; PHAN, W.; BARTZ, A. **Machine Learning Interpretability with H2O Driverless AI: First Edition** Machine Learning Interpretability with H2O Driverless AI. 2019. Disponível em:

<<http://docs.h2o.ai>>.

JIN, X.; KUMAR, L.; LI, Z.; FENG, H.; XU, X.; YANG, G.; WANG, J. A review of data assimilation of remote sensing and crop models. **European Journal of Agronomy**, [s. l.], v. 92, n. November 2017, p. 141–152, 2018.

JOHNSON, D. M.; MUELLER, R. The 2007 Cropland Data Layer. [s. l.], v. 76, n. 11, p. 1201–1205, 2010.

KASPAR, T. C.; COLVIN, T. S.; JAYNES, D. B.; KARLEN, D. L.; JAMES, D. E.; MEEK, D. W.; PULIDO, D.; BUTLER, H. Relationship between six years of corn yields and terrain attributes. **Precision Agriculture**, [s. l.], v. 4, n. 1, p. 87–101, 2003.

KAYAD, A.; SOZZI, M.; GATTO, S.; MARINELLO, F.; PIROTTI, F. Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. **Remote Sensing**, [s. l.], v. 11, n. 23, p. 2873, 2019.

KROSS, A.; ZNOJ, E.; CALLEGARI, D.; KAUR, G.; SUNOHARA, M.; LAPEN, D. R.; MCNAIRN, H. Using artificial neural networks and remotely sensed data to evaluate the relative importance of variables for prediction of within-field corn and soybean yields. **Remote Sensing**, [s. l.], v. 12, n. 14, p. 1–27, 2020.

LAB, P. **PLANET IMAGERY PRODUCT SPECIFICATION: PLANETSCOPE & RAPIDEYE**. 2016. Disponível em: <https://www.planet.com/products/satellite-imagery/files/1610.06_Spec_Sheet_Combined_Imagery_Product_Letter_ENGv1.pdf>.

LI, Y.; SHI, Z.; WU, C.; LI, H.; LI, F. Determination of potential management zones from soil electrical conductivity, yield and crop data. **Journal of Zhejiang University SCIENCE B**, [s. l.], v. 9, n. 1, p. 68–76, 2008.

LIU, J.; PATTEY, E.; MILLER, J. R.; MCNAIRN, H.; SMITH, A.; HU, B. Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. **Remote Sensing of Environment**, [s. l.], v. 114, n. 6, p. 1167–1177, 2010.

LOBELL, D. B. The use of satellite data for crop yield gap analysis. **Field Crops Research**, [s. l.], v. 143, p. 56–64, 2013.

LONGCHAMPS, L.; KHOSLA, R. **Improving N use efficiency by integrating soil and crop properties for variable rate N management**. 15. ed. The Netherlands: Wageningen Academic Publishers, 2015.

LOPRESTI, M. F.; DI BELLA, C. M.; DEGIOANNI, A. J. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. **Information Processing in Agriculture**, [s. l.], v. 2, n. 2, p. 73–84, 2015.

MAESTRINI, B.; BASSO, B. Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. **Scientific Reports**, [s. l.], v. 8, n. 1, p. 1–9, 2018.

MAS, J. F.; FLORES, J. J. The application of artificial neural networks to the analysis of remotely sensed data. **International Journal of Remote Sensing**, [s. l.], v. 29, n. 3, p. 617–663, 2008.

MCCANN, B. L.; PENNOCK, D. J.; VAN KESSEL, C.; WALLEY, F. L. The Development of Management Units for Site-Specific Farming. In: (P. C. Robert, R. H.

Rust, W. E. Larson, Eds.) PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE 1996, Madison. **Anais...** Madison: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 1996.

MENEZES, P. C. De; SILVA, R. P. Da; CARNEIRO, F. M.; GIRIO, L. A. da S.; OLIVEIRA, M. F. De; VOLTARELLI, M. A. Can combine headers and travel speeds affect the quality of soybean harvesting operations? **Revista Brasileira de Engenharia Agrícola e Ambiental**, [s. l.], v. 22, n. 10, p. 732–738, 2018.

MIEZA, M. S.; CRAVERO, W. R.; KOVAC, F. D.; BARGIANO, P. G. Delineation of site-specific management units for operational applications using the topographic position index in La Pampa, Argentina. **Computers and Electronics in Agriculture**, [s. l.], v. 127, p. 158–167, 2016.

MOORE, I. D.; GESSLER, P. E.; NIELSEN, G. A.; PETERSON, G. A. Soil Attribute Prediction Using Terrain Analysis. **Soil Science Society of America Journal**, [s. l.], v. 57, n. 2, p. NP-NP, 1993.

MOORE, I. D.; GRAYSON, R. B.; LADSON, A. R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, [s. l.], v. 5, n. 1, p. 3–30, 1991.

NAWAR, S.; CORSTANJE, R.; HALCRO, G.; MULLA, D.; MOUAZEN, A. M. Delineation of Soil Management Zones for Variable-Rate Fertilization. In: **Advances in Agronomy**. 1. ed. [s.l.] : Elsevier Inc., 2017. v. 143p. 175–245.

PERALTA, N.; ASSEFA, Y.; DU, J.; BARDEN, C.; CIAMPITTI, I. Mid-Season High-Resolution Satellite Imagery for Forecasting Site-Specific Corn Yield. **Remote Sensing**, [s. l.], v. 8, n. 10, p. 848, 2016.

QIN, C.-Z.; ZHU, A.-X.; PEI, T.; LI, B.-L.; SCHOLTEN, T.; BEHRENS, T.; ZHOU, C.-H. An approach to computing topographic wetness index based on maximum downslope gradient. **Precision Agriculture**, [s. l.], v. 12, n. 1, p. 32–43, 2011.

SAKAMOTO, T.; GITELSON, A. A.; ARKEBAUER, T. J. Near real-time prediction of U.S. corn yields based on time-series MODIS data. **Remote Sensing of Environment**, [s. l.], v. 147, p. 219–231, 2014.

SCHWALBERT, R. A.; AMADO, T. J. C.; NIETO, L.; VARELA, S.; CORASSA, G. M.; HORBE, T. A. N.; RICE, C. W.; PERALTA, N. R.; CIAMPITTI, I. A. Forecasting maize yield at field scale based on high-resolution satellite imagery. **Biosystems Engineering**, [s. l.], v. 171, p. 179–192, 2018.

SILVA, J. R. M. Da; ALEXANDRE, C. Spatial Variability of Irrigated Corn Yield in Relation to Field Topography and Soil Chemical Characteristics. **Precision Agriculture**, [s. l.], v. 6, n. 5, p. 453–466, 2005.

SØRENSEN, R.; ZINKO, U.; SEIBERT, J. On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. **Hydrology and Earth System Sciences**, [s. l.], v. 10, n. 1, p. 101–112, 2006.

STUART, R.; PETER, N. **Artificial intelligence-a modern approach 3rd ed.** [s.l.] : Berkeley, 2016.

SZYPULA, B. Quality assessment of DEM derived from topographic maps for

geomorphometric purposes. **Open Geosciences**, [s. l.], v. 11, n. 1, p. 843–865, 2019.

TAVARES, T. de O.; BORBA, M. A. de P.; DE OLIVEIRA, B. R.; DA SILVA, R. P.; VOLTARELLI, M. A.; ORMOND, A. T. S. Effect of soil management practices on the sweeping operation during coffee harvest. **Agronomy Journal**, [s. l.], v. 110, n. 5, p. 1689–1696, 2018.

TURPIN, K. M.; LAPEN, D. R.; GREGORICH, E. G.; TOPP, G. C.; EDWARDS, M.; MCLAUGHLIN, N. B.; CURNOE, W. E.; ROBIN, M. J. L. Using multivariate adaptive regression splines (MARS) to identify relationships between soil and corn (*Zea mays* L.) production properties. **Canadian Journal of Soil Science**, [s. l.], v. 85, n. 5, p. 625–636, 2005.

VELOSO, A.; MERMOZ, S.; BOUVET, A.; LE TOAN, T.; PLANELLS, M.; DEJOUX, J. F.; CESCIA, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. **Remote Sensing of Environment**, [s. l.], v. 199, p. 415–426, 2017.

VENANCIO, L. P.; MANTOVANI, E. C.; DO AMARAL, C. H.; USHER NEALE, C. M.; GONÇALVES, I. Z.; FILGUEIRAS, R.; CAMPOS, I. Forecasting corn yield at the farm level in Brazil based on the FAO-66 approach and soil-adjusted vegetation index (SAVI). **Agricultural Water Management**, [s. l.], v. 225, n. September 2019, p. 105779, 2019.

XIE, Y.; WANG, P.; BAI, X.; KHAN, J.; ZHANG, S.; LI, L.; WANG, L. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. **Agricultural and Forest Meteorology**, [s. l.], v. 246, n. 17, p. 194–206, 2017.

YANG, M.; WANG, G.; LAZIN, R.; SHEN, X.; ANAGNOSTOU, E. Impact of planting time soil moisture on cereal crop yield in the Upper Blue Nile Basin: A novel insight towards agricultural water management. **Agricultural Water Management**, [s. l.], v. 243, n. February 2020, p. 106430, 2021.

YU, B.; SHANG, S. Multi-year mapping of major crop yields in an irrigation district from high spatial and temporal resolution vegetation index. **Sensors (Switzerland)**, [s. l.], v. 18, n. 11, p. 1–15, 2018.

YUAN, H.; YANG, G.; LI, C.; WANG, Y.; LIU, J.; YU, H.; FENG, H.; XU, B.; ZHAO, X.; YANG, X. Retrieving Soybean Leaf Area Index from Unmanned Aerial Vehicle Hyperspectral Remote Sensing: Analysis of RF, ANN, and SVM Regression Models. **Remote Sensing**, [s. l.], v. 9, n. 4, p. 309, 2017.

ZHU, H. D.; SHI, Z. H.; FANG, N. F.; WU, G. L.; GUO, Z. L.; ZHANG, Y. Soil moisture response to environmental factors following precipitation events in a small catchment. **Catena**, [s. l.], v. 120, p. 73–80, 2014.

ZINKO, U.; SEIBERT, J.; DYNESIUS, M.; NILSSON, C. Plant species numbers predicted by a topography-based groundwater flow index. **Ecosystems**, [s. l.], v. 8, n. 4, p. 430–441, 2005.

CAPÍTULO 3 - DETERMINATION OF APPLICATION VOLUME FOR COFFEE PLANTATIONS USING ARTIFICIAL NEURAL NETWORKS AND REMOTE SENSING¹

Abstract

Methods for optimizing the application of phytosanitary products can be an alternative for sustainable agriculture. Such methods can be achieved with the use of artificial intelligence and remote sensing techniques. Our experiments were carried out in a commercial coffee plantation, where morphological variables (height and diameter) and vegetation indexes (normalized difference vegetation index, NDVI and normalized difference red edge, NDRE) were collected in the upper, medium, and lower thirds of the coffee plant. From the remote sensing data, experiments were developed to determine the best neural network topology, in terms of accuracy (RMSE) and precision (R^2) and type (Multilayer Perceptron “MLP” and Radial Basis Function “RBF”), to estimate morphological variables. From these results, we evaluated the possibility of applying pesticides at a variable rate, using the tree row volume principle. The results show that, using remote sensing and artificial neural networks (MLP), it is possible to estimate coffee tree volume with reasonable accuracy. This can be done using a multi-layer perceptron model to estimate coffee tree height and diameter using vegetation indexes of different parts of the plant as input.

Keywords: coffee canopy; vegetation index; variable rate spraying; machine learning; digital agriculture.

‘Introduction

Coffee is a heavily traded global commodity, surpassed only by crude oil. In many countries, its cultivation, processing, and commercialization supports millions of jobs (Martins et al., 2017). In this regard, plant protection is an important segment that guarantees the potential yield (Shilin et al., 2017). Among the plant protection management strategies, the use of pesticides is important for the crop to express its complete genetic potential through the control of pests, weeds, and diseases (Cunha et al., 2019).

Pesticide applications are commonly done by using sprayers usually equipped with some means of air-assistance to carry droplets into the canopy. Air blast sprayers have higher flexibility in the fluid and air delivery that allow the application in variable rate volume (Manandhar et al., 2020). A common method to vary the spray volume is to use the volume of the canopy to adjust the rate. This method is called the tree row

¹ Paper accepted to publication in the journal Computers and Electronics Agricultural.

volume (TRV) method and was developed by Byers et al. (1971). This approach is based on two morphological parameters (plant height and diameter) and one agronomic factor (the distance between rows). With this method, researchers have rendered reasonable results in orchards (Gil et al., 2007; Siegfried et al., 2007).

The accurate measurement of the morphological variables is a key factor of the TRV method (Llorens et al., 2010) because the volume of the canopy is calculated based on these variables. However, these variables are commonly measured by manual methods, which is not very practical because it is difficult for farmers to determine these variables under field conditions, and the TRV method has had limited adoption because it is time-consuming (Cunha et al., 2019). Determining the vegetation volume from manual measurements is laborious and does not always yield satisfactory results due to poor accuracy and low sampling rates (Torres Sánchez et al., 2015). To overcome these problems and apply this method at a commercial scale, it is necessary to develop methods to estimate these physical parameters to generate the within-field volume, which may allow the capture of spatial variability and to vary the rate application based on tree volume.

Remote sensing techniques are a reasonable approach to demonstrate the relationship between crop physical parameters and vegetation indexes (Carneiro et al., 2019), and can be used to estimate biophysical parameters (Myneni et al., 1995). Plant height (Xavier et al., 2006) and diameter can be estimated by remote sensing (Yang et al., 2008). However, for coffee trees, it is still unknown if remote sensing techniques can be used to estimate its physical parameters. The physical parameters (height and diameter) of the coffee tree are complex and can be influenced by several factors, such as hydric regimes, fertigation dosages (Rezende et al., 2010), irrigation water depths (Alves et al., 2000), and other factors. Therefore, a technique with the ability to learn this complex pattern should be applied to estimate these parameters associated with remote sensing.

A feasible approach to learn complex patterns is the use of artificial neural networks (ANN), which is a robust method for finding patterns in complex systems. Research involving ANN associated with remote sensing generally focuses on predicting yield (Yang et al., 2019), biomass (Khan et al., 2020), and biophysical parameters (Xie et al., 2019). However, methods involving ANN associated with remote sensing to estimate the physical parameters of a coffee tree still remain

unknown. A method to estimate within-field physical parameters can be applied to develop a variable rate application in coffee orchards.

Taking all this into consideration, the aim of this research was to investigate the possibility of determining the spray volume, using the tree row volume principle, based on the assessment of the physical characteristics of coffee trees.

Materials and Methods

Description of the experimental site

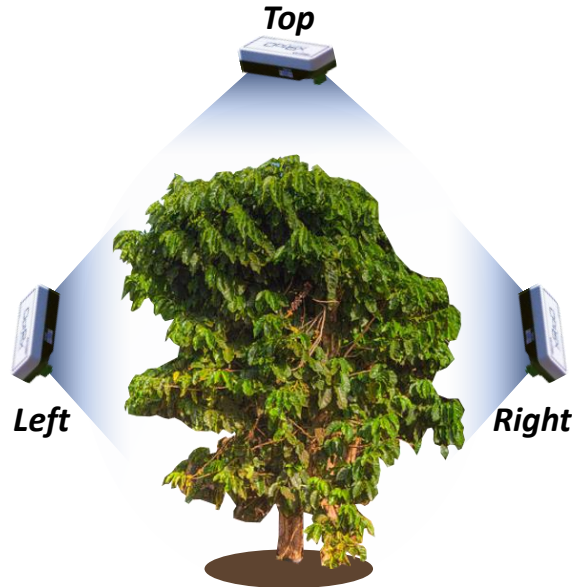
The experiment was carried out in a commercial coffee (*Coffea arabica*) plantation in the first harvest (2 years), with the Catuaí Amarelo IAC 62 variety and an area of 43 ha. The crop was planted with a spatial planting arrangement of 3.5 m between rows, 0.7 m between plants, and east-west planting orientation. For this research, 50 plants were sampled spaced at a grid of 50 m each. The study field was located in the municipality of Pedregulho, São Paulo State, Brazil, near the coordinate latitude 20°8'2.07"S and longitude 47°29'29.49"W (WGS 84).

Proximal remote sensing variables

Vegetation indexes (VI) are mathematical formulas based on various band combinations within the electromagnetic spectrum (Carneiro et al., 2019). Worldwide, the normalized difference vegetation index (NDVI) is the most widely used (Tucker, 1979). The sensitivity of the NDVI can be increased by including a red channel (near 700 nm, normalized difference red edge, NDRE) (Gitelson and Merzlyak, 1997). Studies using these VI have been published and, for instance, correlations have been found with plant canopy structure (Candiago et al., 2015), nitrogen supply, and biomass (Portz et al., 2012).

The variables analyzed were NDVI and NDRE, measured on the right (north), left (south), and top of the plant. For acquisition of the indices, an OptRx® (AgLeader, 2202, South River Side Drive Ames, IOWA 50010, USA) crop sensor system was used at 0.6 m distance from the canopy of the plant. An example is shown in Figure 1.

Figure 1. Remote sensing of the coffee plant canopy. In this research we have adopted the following rule to nominate the variables: first we have the name of the vegetation index and second the face of the plant, e.g., NDVI.Left.



This sensor emits waves in the spectrum of the visible (red, 670 nm) and invisible (near infrared, 730 nm and red-edge, 780 nm) bands, which allow the generation of VI (Table 1).

Table 1. Equations of the used for calculate the VI's.

VI equation*	Reference	Equation
$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$	(Rouse et al., 1974)	(1)
$NDRE = \frac{(NIR - RE)}{(NIR + RE)}$	(Gitelson et al., 1996)	(2)

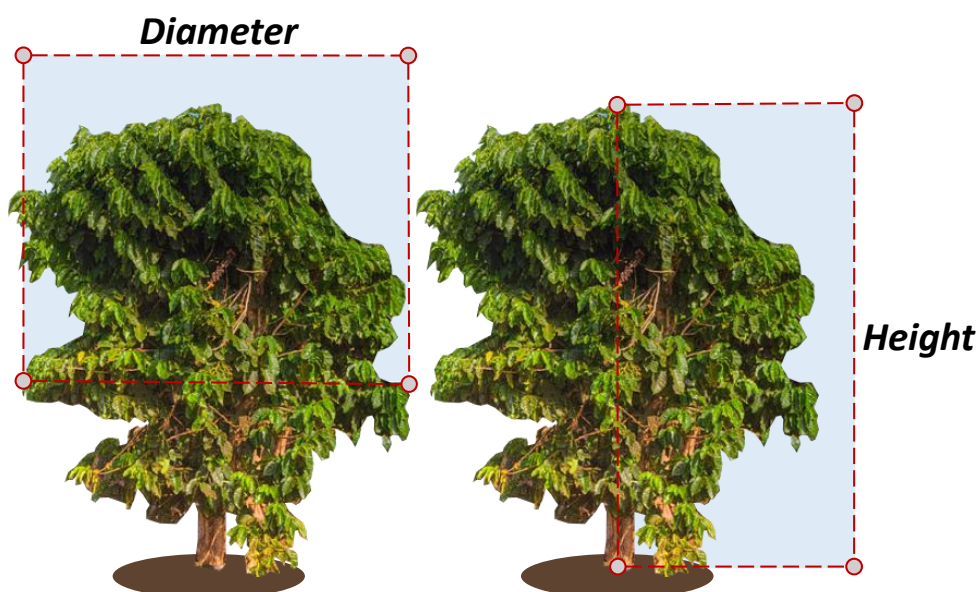
*NDVI is the normalized difference vegetation index, NDRE is the normalized difference red edge, and NIR, RED, and RE are the reflectances in the near infrared, red, and red-edge bands.

Plant physical variables

Considering that the spraying targets the canopy of the plants, and in the case of the coffee plant, the canopy can vary according to the height and diameter of the plants, these two parameters were measured using a measuring tape. For plant height, the measuring tape was attached to a rigid structure and inserted close to the

orthotropic branch of the coffee plant, in order to consider the entire vegetative part of the soil up to the top of the plant, and the diameter of the plants was obtained considering the ends of the plagiotropic branches on both sides (right and left), as shown in Figure 2. Measuring these variables is time-consuming and laborious, necessitating the need for indirect methods capable of predicting diameter and height.

Figure 2. An example of the measurements of the height and diameter from coffee canopy.



Morphological characteristics are not regular across the field; for example, height showed values ranging from 100 cm to 134 cm and diameters ranging from 129 cm to 151 cm, which led to different canopy volume needing different spray volume.

Table 2. Descriptive statistics of coffee plants morphological characteristics.

	Height (cm)	Diameter (cm)
Maximum	134	151
Minimum	100	108
Average	115	129
Standard Deviation	7	10

Neural Network Model Description

For faster, more efficient, and accurate large scale analysis, machine learning algorithms are a very promising approach (Aparecido et al., 2020). To develop nonlinear regression the most common approaches among machine learning algorithms is artificial neural networks (ANN)(Haykin, 1998). ANN can mimic the complex and functional relationship between input and output training data, but users have to spend a considerable amount of time in the process of parameter training and validation (Khan et al., 2020). The approach has gained popularity in various types of studies owing to its strong capability of extracting target information (Yang et al., 2017).

Multi-layer perceptron (MLP) neural network

MLP networks are interconnected by connecting forces represented by values that are called synaptic weights, which are responsible for storing the acquired knowledge. The values used in the input layers were normalized according to equation 3.

$$y_i = \frac{x_i + x_{\max}}{x_i + x_{\min}}, \quad (3)$$

where x_i is the value of the input vector (e.g., NDVI), $x_{\min}$ is the minimum value, and $x_{\max}$ is the maximum value.

The output value of each neuron in the k layer is expressed by $y_k = g(a_k)$, where, g is the activation function of a_k and a_k is the synaptic function, which is a linear combination of the normalized input values and the synaptic weights, as shown in equation 4.

$$a_k = \sum_j y_j w_{kj}, \quad (4)$$

where w_{kj} are the synaptic weights linking the y_j input values with each k neuron.

The transfer or activation function in the neurons of each hidden layer was the hyperbolic function, as shown in equation 5.

$$g(a_k) = \frac{e^{a_k} - -e^{-a_k}}{e^{a_k} + e^{-a_k}}, \quad (5)$$

where, “e” is the Neperian algorithm.

Radial basis function (RBF)

The RBF networks have only one hidden layer, and each neuron contains a radial-based activation function. In each neuron, the Gaussian function was used as a radial base function (Bishop et al., 1995), and the distance (deviation) values of this function increase or decrease the relation to the central point (Haykin, 1998). The Gaussian function is presented by equation 6:

$$\varphi = \exp\left(-\frac{v^2}{2\sigma^2}\right), \quad (6)$$

where, φ is the Euclidean distance between the input vector and the center μ of the Gaussian function, and σ is the width. The Euclidian distance of the input vector to the center μ is the input to a Gaussian function, which gives the activation value of the radial unit.

The output of each neuron in the output layer is presented in equation 7:

$$y_i = \sum_{h=1}^H \omega_{jh} \varphi_h(x_s), \quad (7)$$

where, ω_{jh} is the synaptic weight between the neuron h of the hidden layer and the neuron of the output layer, x_s is the input data x , and φ_h is the Gaussian function of each neuron in the hidden layer.

As in the MLP neural network, the input values were normalized by equation 3, and the values in the output layer were the height or diameter. The RBF neural networks were trained using the k-means algorithm (Bishop, 1995). This algorithm attempts to select the optimal set of points that are placed on the centroid in the patterns of the training data.

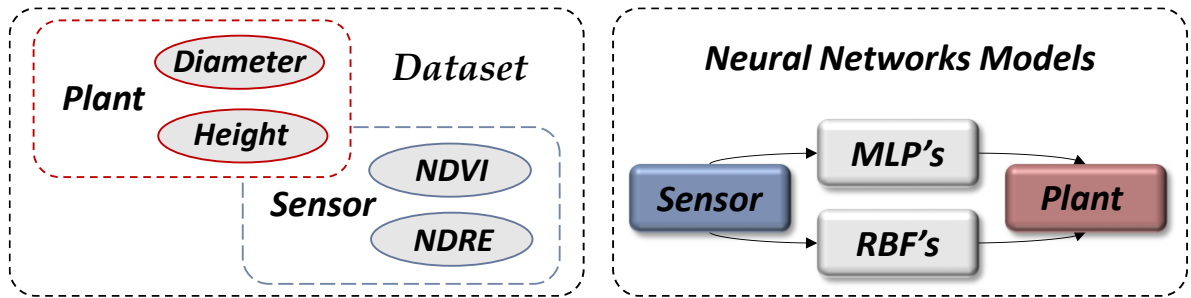
Topology optimization for estimation of plant height and diameter

In an ANN analysis, the choice of variables, neural network architecture, training algorithm, and associated parameters are critical operations and it is time-consuming. The commonly used trial-and-error approach is time-consuming and impractical to

perform manually; for this, the intelligent problem solver (IPS) tool was adopted, which involves the combination of heuristic and sophisticated optimization strategies (minima-bracketing technique and simulated annealing algorithm) (Miao et al., 2006). These procedures were used to automatically test a thousand models of different combinations of predictors, network architecture, and associated parameters, i.e., a thousand models to estimate height and a thousand for diameter, choosing the model with the best performance for MLP and RBF.

The input layer of the network was constituted by the vegetation indexes and the output by height or diameter. The number of neurons varied from 1 to 20 for each type of ANN tested, as well as the number of hidden layers (1–3) for MLP. In addition, different combinations of inputs were tested in each network (Figure 3). The ANN was trained using the k-fold method with five folds in 200 interactions. Developing a robust algorithm for predicting the physical variables of a coffee plant is beyond the research scope since the purpose was to develop a work flow to generate a variable spray volume based on the TRV principle. In order to develop a robust model to predict physical variables, information on more plant conditions is necessary, such as other varieties, plant age, and planting system. These analyses can be an objective of future research.

Figure 3. A theoretical framework showing the general data processing.



Models performance analysis

The accuracy and precision of the proposed models were evaluated by calculating the mean absolute percentage error (MAPE) and adjusted determination coefficient (adjusted R^2) according to equations 8 and 9, respectively.

$$MAPE = \frac{\sum_{i=1}^n \left(\left| \frac{Y_{est_i} - Y_{obs_i}}{Y_{obs_i}} \right| * 100 \right)}{n}, \quad (8)$$

where, n is the number of data, Y_{est_i} is the value of the variable evaluated by the network, and Y_{obs_i} is the value of the observed variable.

$$\text{adjusted } R^2 = 1 - (1 - R^2) \frac{n - 1}{1 - p - 1}, \quad (9)$$

where, n is the sample size, p is the number of variables in the model, and adjusted R^2 is the adjusted coefficient of determination.

Tree row volume (TRV) and variable rate approach

The tree row volume was calculated using equation 10, developed by Byers et al. (1971).

$$\text{TRV} = \frac{H * L * 10000}{D}, \quad (10)$$

where, TRV is the tree row volume ($\text{m}^3 \text{ ha}^{-1}$), H is the height (m), L is the diameter (m), and D is the distance between rows (m).

The best model chosen to predict the height and diameter of coffee plants was a modification of the equation developed by Byers et al. (1971), considering precision (R^2) and accuracy (RMSE). The modification consists of combining the developed neural model with the row volume equation (Equation 11).

$$\text{TRV} = \frac{(h_x = \sum j y_j w_{kj}) * (d_y = \sum j y_j w_{kj}) * 10000}{D}, \quad (11)$$

where, TRV is the tree row volume ($\text{m}^3 \text{ ha}^{-1}$), h_x is the estimated height, d_y is the estimated diameter, and w_{kj} are the synaptic weights connecting the y_j input values with each neuron k . The input layers were composed of NDRE (left and top) and NDVI (right), and the synaptic weights were obtained after training and adjustment of the resulting model.

After calculating the TRV, we used the equation calibrated to the coffee tree developed by De Sousa Júnior et al. (2017) to obtain 100 drops per cm^3 at the maturation vegetation stage to obtain the spray volume. We then compared the minimum, maximum, and average volumes of the proposed method against the volume used by the farmer, which was 200 (l ha^{-1}).

Results

Sensitivity analysis was used to rank the importance of each input variable used in the models to estimate coffee tree morphological characteristics. Among all the vegetation indexes, NDRE.Top was the most important for MLP models to estimate plant height and NDRE.Left was the most important for diameter estimation. In contrast, for RBF models, NDVI.Right should be more sensitive as a variable input to predict the biophysical parameters in a coffee tree. For plant height, the MLP model used fewer variables as input (3 variables) than RBF models (5 variables). Both the models used three variables for plant diameter, while the MLP only used the NDRE index (Table 2).

Table 2. Sensitivity analysis of the neural network's models for estimation of coffee plant height and diameter.

Physical characteristics	Neural Networks	Sensitivity Analysis					
		NDRE			NDVI		
		Right	Left	Top	Right	Left	Top
Height	MLP*	-	2	1	3	-	-
	RBF**	5	3	2	1	-	4
Diameter	MLP*	2	1	3	-	-	-
	RBF**	2	-	3	1	-	-

*MLP: Multilayer Perceptron; **RBF: Radial Basis Function, the numbers means the rank of importance in the neural network, “-” indicates that the variable was not used in the resultant model.

Among all the topologies tested, using 20 neurons in the first hidden layer was the selected quantity for both variables when using MLP models, whereas for the second hidden layer, the number of neurons was similar. For RBF networks, the best topology used 20 neurons in the hidden layer to estimate coffee height and 19 neurons to estimate plant diameter (Table 3).

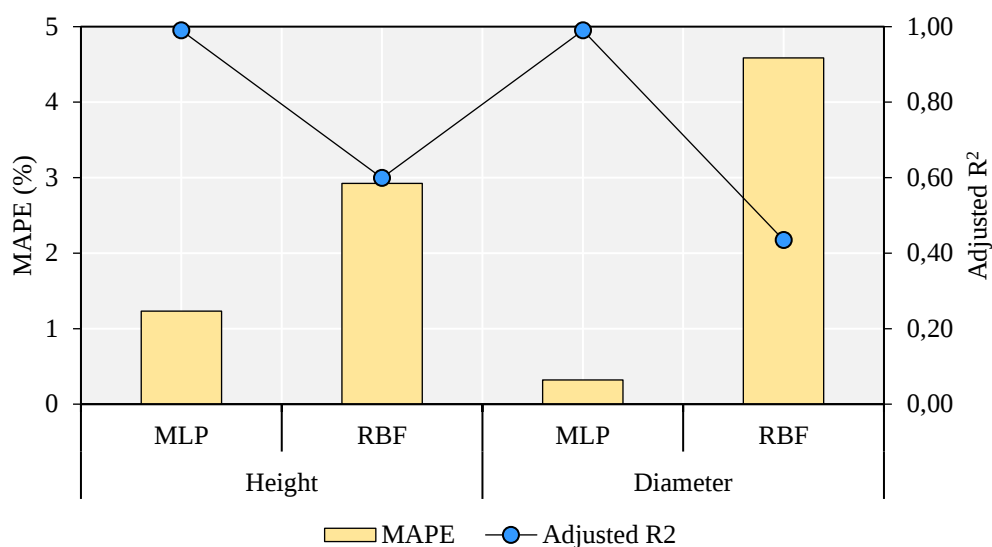
Table 3. Resultant models and their topologies.

Plant Variable	Topology*
Height	MLP 3:3-20-11-1:1
	RBF 5:5-20-1:1
Diameter	MLP 3:3-20-10-1:1
	RBF 3:3-19-1:1

* RBF: Radial Basis Function; MLP: Multilayer Perceptron; Neural network architecture is presented in the form of I:N1-N2-N3:O, where “I” is the number of input variables, “N1”, “N2”, and “N3” are the numbers of neurons in the Input, Hidden, and Output layers, respectively, and “O” is the number of output variables.

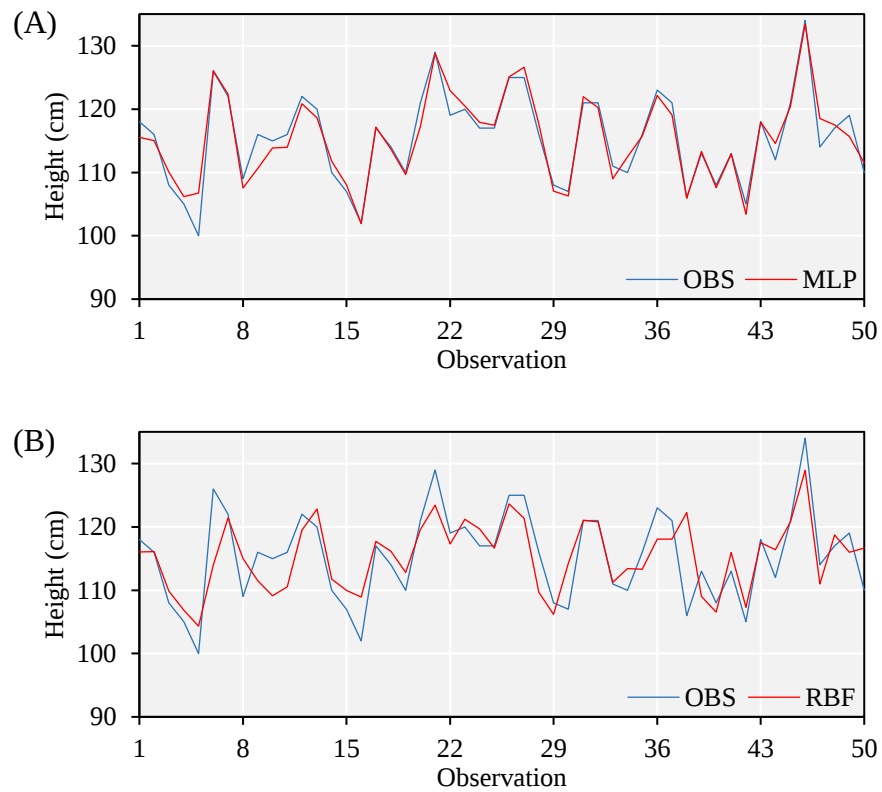
Using the metrics of accuracy (MAPE) and precision (adjusted R^2) to estimate both variables (height and diameter), the resultant models using the MLP network presented the best precision and accuracy. For estimation of coffee tree height, using the best model enhanced the accuracy (MAPE = 1.23%) and precision (adjusted R^2 = 0.99). The same pattern was found for estimation of plant diameter (MAPE = 0.32% and adjusted R^2 = 0.99). When using RBF networks, the least accuracy and precision were found for both variables: height (MAPE = 2.92 % and adjusted R^2 = 0.60) and diameter (MAPE = 4.57 % and adjusted R^2 = 0.45). Overall, the most accurate estimation can be performed using the MLP network for the diameter variable of a coffee tree (Figure 4).

Figure 4. Precision (MAPE) and accuracy (adjusted R²) for both neural networks, MLP and RBF, for estimation of height and diameter.



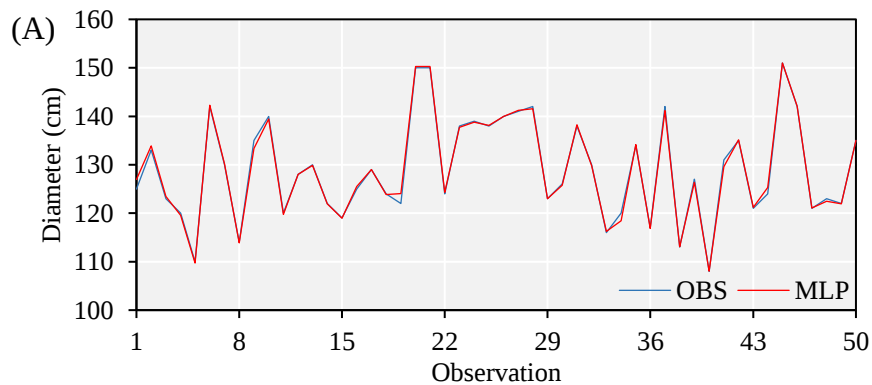
Graphical comparisons of the estimated variables (height and diameter) and observed variables using VIs as input in MLP and RBF neural models are shown in Figures 5 and 6. The MLP models estimated both variables (height and diameter) better than RBF models, based on better average prediction accuracies. It is important to mention that for RBF models, there was a pattern that when the plant height ranged between 100–108 cm, overestimation occurred, whereas when the plant height ranged between 121–134 cm, underestimation occurred. This pattern was not observed for MLP models. This pattern demonstrates that using MLP, it is possible to describe the spatial variability of the morphological variables. To be able to do this is very important in the estimation process using ANN, since a higher plant is likely to have more plagiotropic branches, thereby increasing canopy density and spray volume needed for the application.

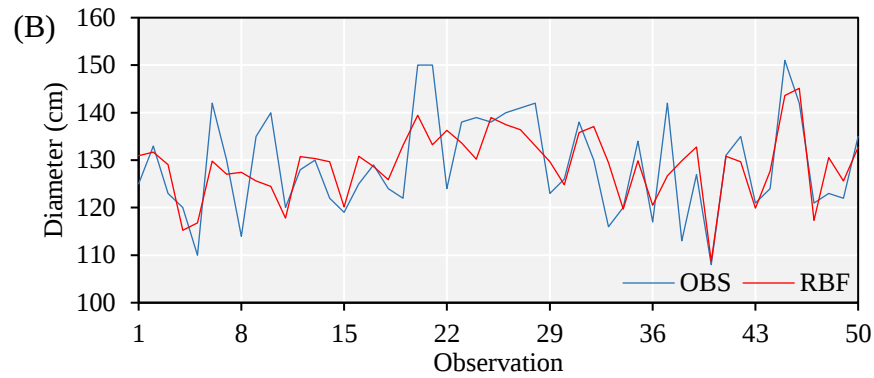
Figure 5. Comparison between measured and estimated coffee plant height using multilayer perceptron network (A) and radial basis function network (B).



For coffee tree diameter estimation using RBF models, there was a pattern that when the plant diameter ranged between 108–119 cm, overestimation occurred, whereas when the plant diameter ranged between 133–151 cm underestimation occurred. This pattern was not observed for the MLP models.

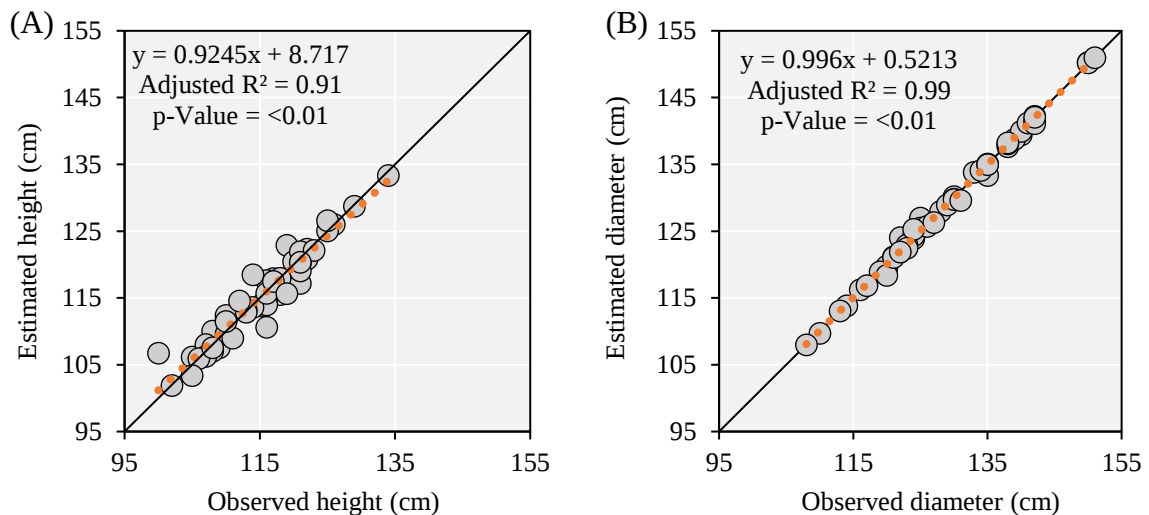
Figure 6. Comparison between measured and estimated coffee plant diameter using multilayer perceptron network (A) and radial basis function network (B).





To evaluate the performance of the selected models with the best precision and accuracy, regression analysis was performed between the in-field measured variable and the estimated variable using machine learning models with VIs as input (Figure 7). The results of the model calibration and performance indicated that the MLP models had much better ability to predict coffee tree morphological variables. The performance analysis showed significant regressions between the observed data and the estimated data with a higher adjusted R^2 for diameter estimation.

Figure 7. Performance analysis of the resultant models: (A) coffee plant height; (B) coffee plant diameter.



After predicting the morphological variables of the coffee tree using the best neural model, we used the TRV equation to calculate the volume of the tree. We found that the mean, minimum, and maximum were similar (Figure 8). We then applied the equation developed by De Sousa Júnior et al. (2017) to transform the TRV into a spraying volume and compared this value with the farmer spraying volume used in

general pesticide applications, which is a fixed volume for entire field. These results support the idea that in the field conditions, the coffee trees need different spray volumes, whereas the farmer uses a fixed volume for his entire field (Figure 9). The mean spray volume calculated using observed data and estimated by MLP models was similar at around 206 L ha⁻¹. Evaluating the mean using the traditional approach in which a fixed volume is applied, we have similar results, because the farmer was using 200 L ha⁻¹ and the mean using TRV principle was 206 L ha⁻¹. Therefore, when we evaluate the minimum and the maximum volume using the TRV principle, we see that the minimum required in the field was 160 L ha⁻¹ (in the traditional approach, approximately 25% more than the need of the plant was being applied, according to TRV method) and the maximum was 267 L ha⁻¹ (in the traditional approach, approximately 25% less than the need of the plant was being applied), according to MLP estimation. These results suggest that in the traditional approach, there was over-application in some areas and under-application in other areas. This lead us to infer that pesticide application should be done using different volumes across the field due its tree volume variability.

Figure 8. Descriptive statistics of the observed (OBS) and estimated (MLP) tree row volume.

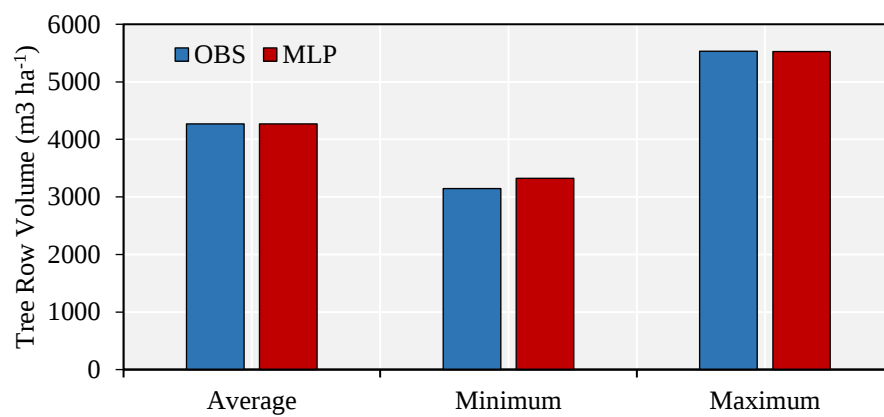
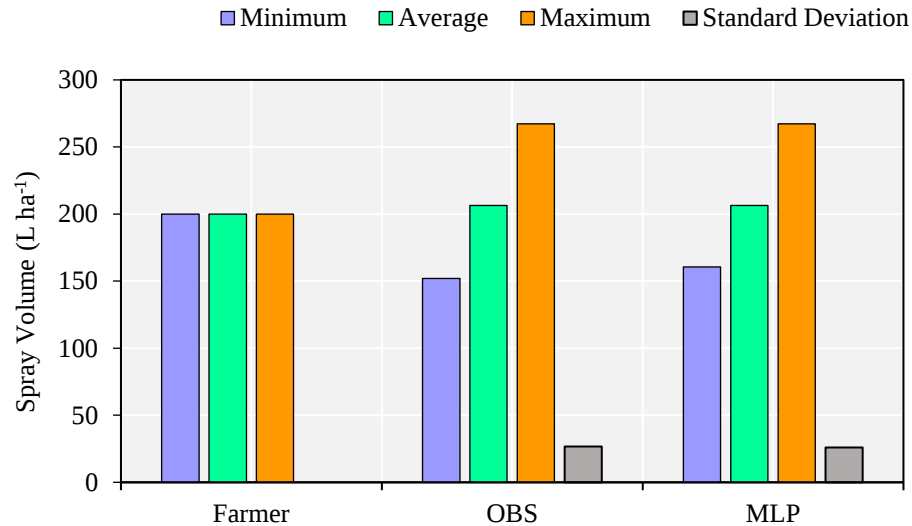


Figure 9. Descriptive statistics of the spraying volume used by the farmer, observed (OBS), and estimated spraying volume using MLP. The spraying volume was generated by De Sousa Júnior et al. (2017) equation for OBS and MLP.



In summary, to estimate coffee tree height and diameter with reasonable accuracy, it is necessary to use both the variable MLP neural network and NDRE (left and top), and NDVI (right) for plant height and for plant diameter estimation, to use all NDRE's VI.

Discussion

In this study, a method for estimating coffee tree volume was proposed, and our results present evidence that using proximal remote sensing and artificial neural networks, it is possible to estimate coffee tree volume with reasonable accuracy. This is important because with this information it is possible to maintain the chemical rate recommendation, and application should be based on crop canopy volume rather than on land area, which is key to the tree row volume concept (Llorens et al., 2010). One of the key factors for the success of variable spray rate technology is the accurate measurement of crop dimensions (Llorens et al., 2010). Other authors have achieved reasonable results to adapt to spray volume based on crop dimensions and using ultrasonic sensors in vineyards (Gil et al., 2007). This research presents an approach to use an electronic system and the TRV principle to develop variable rate spraying using an electronic system that varies the rate at the same time that the ultrasonic

sensor models the canopy of the tree. This electronic system could also be used to receive the estimate of the TRV obtained by our neural model that uses VI and then varies the spraying rate. Another approach to estimate the volume of the coffee tree is by using digital images (Cunha et al., 2019). However, using this method the farmer could miss the adequate time for pesticide application, due to the time consumed for drone mapping, ground control points acquisition, and photogrammetry processing of the images.

The first result of our analysis indicates that the two types of neural networks had different sensitivities to the inputs analyzed. The RBF networks had NDVI as the most important variable in the models for both variables (height and diameter). This fact can be attributed to the decrease in the accuracy of the models, mainly because of the saturation of the NDVI that led to a low variability in the vegetation index, whereas the plant canopy variables still changed across the field (Carneiro et al., 2019). This problem can be solved by using another wavelength of the radiometric spectrum (Van Niel and McVicar, 2004), such as the red-edge band, which allows us to calculate the NDRE, which has more resistance to saturate than NDVI (Taskos et al., 2015). Thus, the MLP models used more red-edge indexes as input, which probably increased the accuracy of the estimation. The variation in sensitivity analysis between the sides of the coffee tree can be explained by its physiological response to environmental variations, such as the level of solar radiation that implies the development of adaptation mechanisms that can interfere with plant yield and maturation (Cassia et al., 2016). These mechanisms of changes in the growth and maturation of the coffee are still unclear, however, it is known that light exposure to the afternoon sun can reduce the growth and incidence of pests and diseases and should be studied on both sides of the plant, and the planting direction influences coffee tree yield (Pinto et al., 2006).

The second and the most of the outcomes of this study showed the potential of using artificial neural networks associated with remote sensing to predict the morphological variables of coffee trees with reasonable accuracy, and with this result, the spray volume application can be varied. This result is consistent with the bases of precision agriculture, which, according to estimated variability, improves the resource use efficiency. This enables the use of site-specific crop management, which is an approach of farm management that uses information technology to ensure that the crops receive what they need to improve the yield (Virnodkar et al., 2020).

From a practical standpoint, producers and researchers could use artificial neural networks associated with remote sensing to determine the spatial within-field variability of the morphological variables to make informed decisions based on the estimated coffee tree heights and diameters. Spatial representation of predictive morphological variables can be easily achieved by using active canopy sensors that record the vegetation indexes and then apply the neural network algorithm to generate the geo-field of the estimated variables (height and diameter). The predictive morphological maps at a high resolution can resolve the variability of coffee tree height and diameter, and its patterns would be valuable for farmers in planning their sub-field scale spray application in order to achieve their production goals. Future research should focus on testing other approaches using machine learning associated with remote sensing to estimate morpho-physiological parameters of plants.

Conclusions

Artificial neural networks combined with remote sensing data can provide useful information, allowing the development of variable rate spraying volume technology on row trees. Owing to the learning of the pattern of morpho-physiological variables and vegetation indexes, it is possible to implement a new strategy of variable rate spraying volume. This can be done using a multi-layer perceptron model to estimate coffee tree height and diameter using input vegetation indexes of different parts of the plant in the left and top (NDRE) and right (NDVI). This enables the calculation of the TRV across the field variability and then using the spray volume equations to generate recommendations based on TRV. This method can be potentially applied to support agricultural decisions in terms of optimizing the spray applications.

Acknowledgments: We thank the Coordination for the Improvement of Higher Education Personnel (CAPES), Financing Code 001 for the scholarship granted to the first author, and the Laboratory of Machinery and Agricultural Mechanization (LAMMA) of the Department of Engineering and Exact Sciences – Unesp/FCAV. The authors would like to express their gratitude to Fazenda Santa Alcina for providing the experimental site for the development of this research.

References

- ALVES, M.E.B. et al. Crescimento do cafeeiro sob diferentes lâminas de irrigação e fertirrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 4, n. 2, p. 219-225, 2000.
- APARECIDO, L.E.O. et al. Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. **International journal of biometeorology**, p. 1-18, 2020.
- BYERS, R.E.; HICKEY, K.D.; HILL, C.H. Base gallonage per acre. **Virginia Fruit**, v. 60, n. 8, p. 19-23, 1971.
- CANDIAGO, S. et al. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. **Remote sensing**, v. 7, n. 4, p. 4026-4047, 2015.
- CARNEIRO, F.M. et al. Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. **Precision Agriculture**, p. 1-29, 2019.
- CASSIA, M. T. et al. Effect of planting axles, sunlight faces and rod vibration frequencies in the mechanized coffee harvesting. **African Journal of Agricultural Research**, v. 11, n. 5, p. 392-399, 2016.
- DA CUNHA, J.P.; SIRQUEIRA NETO, M.A.; HURTADO, Sandro. Estimating vegetation volume of coffee crops using images from unmanned aerial vehicles. **Engenharia Agrícola**, v. 39, n. SPE, p. 41-47, 2019.
- GIL, E. et al. Variable rate application of plant protection products in vineyard using ultrasonic sensors. **Crop Protection**, v. 26, n. 8, p. 1287-1297, 2007.
- GITELSON, A.A.; KAUFMAN, Y.J.; MERZLYAK, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289-298, 1996.
- GITELSON, A.A.; MERZLYAK, M.N. Remote estimation of chlorophyll content in higher plant leaves. **International Journal of Remote Sensing**, v. 18, n. 12, p. 2691-2697, 1997.
- HAYKIN, S. Neural networks: a comprehensive. The Knowledge Engineering Review, v. 13, n. 4, p. 409-412, 1998.
- KHAN, M.S. et al. An artificial neural network model for estimating Mentha crop biomass yield using Landsat 8 OLI. **Precision Agriculture**, v. 21, n. 1, p. 18-33, 2020.

KUBAT, M. Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7. **The Knowledge Engineering Review**, v. 13, n. 4, p. 409-412, 1999.

LLORENS, J. et al. Variable rate dosing in precision viticulture: Use of electronic devices to improve application efficiency. **Crop protection**, v. 29, n. 3, p. 239-248, 2010.

MANANDHAR, A. et al. Techno-economic impacts of using a laser-guided variable-rate spraying system to retrofit conventional constant-rate sprayers. **Precision Agriculture**, p. 1-16, 2020.

MARTÍNEZ-CASASNOVAS, J.A. et al. Mobile terrestrial laser scanner applications in precision fruticulture/horticulture and tools to extract information from canopy point clouds. **Precision Agriculture**, v. 18, n. 1, p. 111-132, 2017.

MARTINS, G.D.; GALO, M. de L.B.T.; VIEIRA, B.S. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 10, n. 12, p. 5395-5403, 2017.

MIAO, Y.; MULLA, D.J.; ROBERT, P.C. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. **Precision Agriculture**, v. 7, n. 2, p. 117-135, 2006.

MYNENI, R.B. et al. Optical remote sensing of vegetation: modeling, caveats, and algorithms. **Remote sensing of environment**, v. 51, n. 1, p. 169-188, 1995.

PINTO, F.A. et al. Orientacao da face de exposicao ao sol de parcelas de cafeeiros produzidos em regio de montanha e seu efeito na qualidade de bebida. *Revista Brasileira de Armazenamento*, Vicoso, n. 9, esp. cafe, p. 32-39, 2006.

PORTZ, G.; MOLIN, J.P.; JASPER, J. Active crop sensor to detect variability of nitrogen supply and biomass on sugarcane fields. **Precision Agriculture**, v. 13, n. 1, p. 33-44, 2012.

REZENDE, R. et al. Crescimento inicial de duas cultivares de cafeeiro em diferentes regimes hídricos e dosagens de fertirrigação. **Engenharia Agrícola**, v. 30, n. 3, p. 447-458, 2010.

ROUSE, J.W. et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA special publication**, v. 351, n. 1974, p. 309, 1974.

SHILIN, W. et al. Performances evaluation of four typical unmanned aerial vehicles used for pesticide application in China. **International Journal of Agricultural and Biological Engineering**, v. 10, n. 4, p. 22-31, 2017.

SIEGFRIED, W. et al. Dosage of plant protection products adapted to leaf area index in viticulture. **Crop Protection**, v. 26, n. 2, p. 73-82, 2007.

SOUSA JÚNIOR, J.M. de. Determinação do índice de volume de pulverização para a cultura do café: regulação e calibração de pulverizador hidropneumático. 2016.

TASKOS, D.G. et al. Using active canopy sensors and chlorophyll meters to estimate grapevine nitrogen status and productivity. **Precision agriculture**, v. 16, n. 1, p. 77-98, 2015.

TORRES-SÁNCHEZ, J. et al. High-throughput 3-D monitoring of agricultural-tree plantations with unmanned aerial vehicle (UAV) technology. **PloS one**, v. 10, n. 6, p. e0130479, 2015.

TUCKER, C.J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote sensing of Environment**, v. 8, n. 2, p. 127-150, 1979.

VAN NIEL, T.G.; MCVICAR, T.R. Current and potential uses of optical remote sensing in rice-based irrigation systems: a review. **Australian Journal of Agricultural Research**, v. 55, n. 2, p. 155-185, 2004.

VELLIDIS, G.; BEASLEY, J. Using vegetation indices to determine peanut maturity. **Report to the Georgia Agricultural Commodity Commission for Peanuts. GPC**, 2013.

VIRNODKAR, S.S. et al. Remote sensing and machine learning for crop water stress determination in various crops: a critical review. **Precision Agriculture**, v. 21, n. 5, p. 1121-1155, 2020.

XAVIER, A.C. et al. Hyperspectral field reflectance measurements to estimate wheat grain yield and plant height. **Scientia Agricola**, v. 63, n. 2, p. 130-138, 2006.

XIE, Q. et al. Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 80, p. 187-195, 2019.

YANG, C.; LIU, T.; EVERITT, J.H. Estimating cabbage physical parameters using remote sensing technology. **Crop protection**, v. 27, n. 1, p. 25-35, 2008.

YANG, Q. et al. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. **Field Crops Research**, v. 235, p. 142-153, 2019.

YANG, T. et al. An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis. **Information Sciences**, v. 418, p. 302-316, 2017.

CAPÍTULO 4 - Mapeando a maturação abaixo do solo usando UAV: estudo de caso em amendoim

Resumo

Por possuir os frutos abaixo do solo, crescimento indeterminado e desuniformidade de floração, estimar a maturação de amendoim é essencial para o manejo e tomada de decisões, principalmente para determinar o momento de colheita. Problemas complexos e não lineares como a previsão de produtividade tem sido realizada de forma acurada por meio de Redes Neurais Artificiais (RNA's) para vários cultivos. No presente estudo, bandas espectrais oriundas de sensoriamento remoto aéreo e índices de vegetação foram avaliadas para estimar o índice de maturação de amendoim. Foram utilizadas RNA's do tipo multicamada e de função de ativação de base radial. Os modelos de RNA's foram avaliados quanto a precisão e acurácia pelas métricas do coeficiente de determinação (R^2) e raiz quadrada do erro médio (RMSE). O estudo enfatiza o uso de dados das bandas espectrais (vermelho, verde, infravermelho próximo e vermelho próximo) para capturar a maturação de amendoim e obter resultados mais parcimoniosos e com alta precisão e acurácia ($R^2 = 0.91$, RMSE = 0.07). As várias bandas espectrais derivadas de imagens de drone juntas aos parâmetros da rede resultante permitiram a estimação da maturação de amendoim determinada em campos comerciais com elevada acurácia e de forma não destrutiva.

Keywords *Arachis hypogaeae* L · Modelagem de cultivo · Aprendizado de Máquinas · Maturação a campo · Imagens multiespectrais · Inteligência artificial

Introdução

A maturação das vagens de amendoim é o fator primordial para indicar o início da colheita mecanizada da cultura visando garantir máxima produtividade e redução de perdas. Determinar o tempo ideal para iniciar o arranquio é uma das decisões mais difíceis durante o cultivo de amendoim (Colvin et al. 2018), pois, a colheita antecipada ou tardia, reduzir a produtividade e a qualidade do produto (Abd-El Monsef et al. 2019).

Uma forma de determinar a maturação de vagens é a utilização do Quadro de Maturação (Peanuty Profile Board, Williams and Drexler 1981) em conjunto com o índice de maturação do amendoim (Peanuty Maturation Index, Rowland et al. 2006) em que uma amostra é considerada com maturação ótima quando o índice é de 0,7. No geral, este método é trabalhoso, subjetivo e necessita de amostragem densa, pois, o amendoim apresenta taxas de maturação em diferentes áreas de um campo (Vellidis and Beasley 2013). Diante disso, são necessárias alternativas que possam estimar a

maturação total em campos de produção, ao invés da maturação de plantas individuais, por meio de amostras não-destrutivas.

Alternativas para estimar, a maturação de amendoim de forma não destrutiva foram estudadas por Rowland et al. (2008), que demonstraram existir correlação das bandas espectrais no infravermelho com a maturação de vagens de amendoim utilizando o sensoriamento remoto proximal. Em escala de parcelas Vellidis e Beasley (2013) demonstraram que o Índice de Vegetação Não-Linear (NLI) pode estimar a maturação da cultura do amendoim. Santos (2019) demonstrou a existência de relação linear entre a maturação de amendoim e índices de vegetação modificados, em escala de campo, sendo o Modified Non-linear Index (MNLI) o índice mais promissor junto ao NLI.

Entretanto, estes trabalhos demonstraram apenas a relação linear e o ajuste de modelos, não os validando, não permitindo inferir se esses métodos são capazes de estimar a maturação com dados independentes. Khan et al. (2020) obtiveram sucesso na previsão de biomassa de menta utilizando múltiplos índices de vegetação como variáveis independentes no modelo. Entretanto, ainda não se tem relatos da utilização de métodos que tenham múltiplos índices de vegetação como variável preditora da maturação e algoritmos de identificação de padrões em amendoim.

Redes neurais artificiais (RNA's) são métodos robustos para encontrar padrões em sistemas complexos. Eles são baseadas no funcionamento do cérebro humano, pois possuem neurônios artificiais simplificados, interconectados por meio de sinapses, os quais são capazes de adquirir conhecimento por meio de um processo de aprendizagem (Haykin 1998).

Pesquisas envolvendo RNA's, sensoriamento remoto e culturas agrícolas têm se concentrado principalmente na previsão de produtividade (Yang et al. 2019), biomassa (Khan et al. 2019) e parâmetros biofísicos (Xie et al. 2019). Métodos envolvendo RNAs e sensoriamento para estimar a índice de maturação de amendoim ainda não foram propostos. Portanto, o desenvolvimento de um método como este tem o potencial de ajudar produtores e cooperativas a desenvolver estratégias adequadas, principalmente em termos de tomada de decisão para definir o melhor momento de colheita, realizando então uma colheita inteligente.

Pressupondo que, devido à capacidade de aprendizado de redes neurais, é possível estimar a maturação da cultura do amendoim por meio de índices de vegetação obtidos por sensoriamento remoto, objetivou avaliar-se RNA's e múltiplos

índices de vegetação modificados e não-modificados estimam de forma não destrutiva a maturação do amendoim.

Metodologia

Campo experimental e mensuração da refletância da cultura

Este experimento foi conduzido em dois campos comerciais de amendoim localizados nos EUA, sendo uma área irrigada sob sistema de pivô central (Coffee County, Georgia) e outra não irrigada (Douglas County, Georgia). Estas áreas foram selecionadas por representarem bem o clima da região onde se concentra a produção de amendoim no país. A cultivar utilizada foi Georgia-06G. Foram utilizadas imagens de satélite históricas (composição RGB) para seleção das parcelas experimentais. Cada área experimental foi constituída de 24 parcelas (células em grade de 1 ha) em que maturação durante 6 semanas consecutivas foi avaliada. A última mensuração da maturação foi realizada no dia da colheita. As vagens foram colocadas no Quadro de Maturação (Profile Board) e o número de vagens em cada classe de cor foi registrado pelo método hull-scrape (Williams and Drexler 1981). O índice de maturação do amendoim (PMI) foi calculado para cada amostra de acordo com Rowland et al. (2006).

Durante cada dia de coleta, uma câmera Parrot Sequoia (Paris, France) embarcada na aeronave remotamente pilotada 3DR solo (Berkley, EUA) foi utilizada para capturar imagens multiespectrais em 4 diferentes bandas: Verde (530-570 nm), Vermelho (640-680 nm), Infravermelho próximo (770-810 nm) e Vermelho próximo (Red Edge - 730-740 nm). Todos os voos da aeronave foram realizados entre 2 h do meio dia solar com altura de 90 m e sobreposições longitudinal e lateral de 70%. Em 90 m a resolução geométrica da câmera foi 0.95 m. Antes e depois de cada voo, a câmera foi usada para tomar imagens do painel de calibração radiométrica, que posteriormente foram utilizadas para calibrar as imagens capturadas sobre a área. Alvos preto e branco (1x1 m) foram colocados nos vértices de cada área (24 ha) para servir como pontos de controle do solo (PC). As coordenadas dos PCs foram gravadas com o GNSS Trimble R8 correção RTK (Sunnyvale, EUA).

O software Pix4D foi utilizado para criar os ortomosaicos de refletância em cada uma das quatro bandas, para realizar a correção radiométrica e correção geométrica.

Após criar os ortomosaicos, o software ArcGIS (Redlands, US) foi utilizado para extrair o valor médio da refletância do pixel de cada uma das 24 células em ambos os campos.

Índices de vegetação

A partir dos valores de refletância extraídos dos ortomosaicos, seis IVs não modificados, usados por outros trabalhos, foram selecionados para avaliação da maturação de amendoim como possíveis variáveis preditoras e, como alternativa, foram modificados os índices substituindo a banda do vermelho pela banda do vermelho próximo (Tabela 1).

Tabela 4 Índices de vegetação calculados

Índice de vegetação (IV)	Equação	Referência
IVs Não Modificado		
NDVI	$\frac{(NIR - RED)}{(NIR + RED)}$	(Rouse 1973)
GNDVI	$\frac{(NIR - Green)}{(NIR + Green)}$	(Gitelson et al. 1996)
NLI	$\frac{(NIR^2 - RED)}{(NIR^2 + RED)}$	(Goel and Qin 1994)
MNLI	$\frac{NIR^2 - RED + L * NIR - L * RED}{(NIR^2 + RED + L)^n}$	(Gong et al. 2003)
NDRE	$\frac{(NIR - RE)}{(NIR + RE)}$	(Fitzgerald et al. 2010)
SR	$\frac{NIR}{RED}$	(Jordan 1969)
IVs Modificados		
SR _{RE}	$\frac{NIR}{RE}$	(Santos et al. 2019)
MNLI _{RE}	$\frac{NIR^2 - RE + L * NIR - L * RE}{(NIR^2 + RE + L)^n}$	(Santos et al. 2019)
CI _{RE}	$\frac{RE - GREEN}{NIR}$	(Autor 2019)
MSR _{RE}	$\frac{(RE - 1)}{\sqrt{\frac{NIR}{RE}} + 1}$	(Wu et al. 2008)
NLI _{RE}	$\frac{(NIR^2 - RE)}{(NIR^2 + RE)}$	(Santos et al. 2019)

¹NDVI: normalized difference vegetation index; GNDVI: green normalized difference vegetation index; NLI: nonlinear index; MNLI: modified non-linear index; NDRE: normalized difference red edge index; SR: simple ration; MSR_{RE}: modified simple ration red ege index. Todos os índices com a presença da palavra _{RE} significam que a banda do vermelho foi substituída pela do vermelho próximo ²Bandas espectrais: infravermelho próximo (NIR); vermelho (RED); vermelho próximo (RE) e verde (GREEN). “ fator de ajuste (L = 0.5).

Análise estatística

O comportamento das variáveis foi avaliado por meio da estatística descritiva, utilizando diagrama de caixa (Box Plot) apresentando-os em função dos GDA. O modelo não linear de Gompertz (1825) (Equação 1) foi ajustado com o valor médio de PMI das 28 amostras por época de coleta para descrever o comportamento do PMI em função dos Graus-Dia Acumulado (GDA). Para obtenção dos graus dia acumulado foi utilizado a plataforma web peanutfarm.org onde o usuário fornece a data de semeadura e a plataforma retorna os graus dia acumulado de forma diária ao longo do ciclo da cultura. A otimização dos parâmetros foi realizada pelo Método do Gradiente Reduzido Generalizado (GRG) (Abadie and Carpentier 1965).

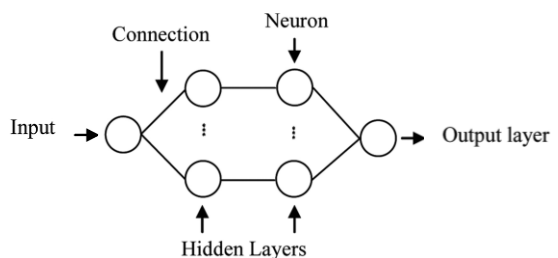
$$Y = \alpha e^{-e^{-k(X-\gamma)}} \quad (1)$$

em que, α é a capacidade de sustentação do meio que limita o crescimento, k é a taxa de crescimento média e γ é o ponto de inflexão em graus-Dia acumulados.

Modelagem da maturação da cultura por índices de vegetação

Como método de estimativa da maturação de vagens de amendoim, foram utilizadas Redes Neurais Artificiais (RNA) do tipo perceptron de multicamadas (MLP). RNA's são sistemas que consistem de neurônios artificiais, os quais desempenham funções similares aos neurônios humanos (Jiang et al. 2004), que sem pressuposições sobre a distribuição dos dados podem aprender relações entre as variáveis apresentadas como exemplo (Zhang et al. 2008). Uma vantagem da utilização de RNA's é que este método leva em consideração não linearidades na relação entre as informações de entrada e saída (Savegnago et al. 2011). A estrutura básica de rede neural do tipo multicamadas é: uma camada de entrada, duas camadas ocultas e uma camada de saída (Figura 1). O número de camadas ocultas pode variar em função do objetivo do problema que se pretende resolver e dos recursos computacionais disponíveis.

Fig. 1 Arquitetura da rede neural artificial do tipo multicamadas com 2 camadas ocultas. (Adaptado de Savegnago et al. 2011)



Em uma análise de RNA, a escolha de variáveis, arquitetura de rede neural, algoritmo de treinamento e parâmetros associados são operações críticas e demandam tempo (StatSoft 2013). A abordagem de tentativa e erro comumente adotada consome muito tempo e é impraticável para se fazer de forma manual; para isso a ferramenta Intelligent Problem Solver (IPS), uma ferramenta embutida no software Statistica 7 que envolve a combinação de estratégias heurísticas e sofisticadas de otimização (técnica de minimização de becketing e algoritmo simulado de annealing) (Miao et al. 2006). Estes procedimentos foram usados para automaticamente testar mil modelos de diferentes combinações de preditores, arquitetura de redes e parâmetros associados.

A camada de entrada da rede foi constituída pelos índices de vegetação e bandas espectrais e a saída pelo PMI. Foi variado o número de neurônios (1 a 30) para as redes MLP e de 1 a 45 para redes com Função de Ativação de Base Radial (RBF), a quantidade de camadas ocultas (1 - 3) para MLP e as entradas das redes. Foram utilizadas as seguintes combinações de entrada todas as variáveis (bandas, IV's não modificados e modificados); bandas (red, green, NIR e RE); IV's não modificados (Tabela 1); IV's modificados (Tabela 1) e cada uma das variáveis (bandas, IV's não modificados e modificados) separadamente. Para cada combinação de entrada foram treinados mil modelos. As redes do tipo multicamadas foram treinadas utilizando o algoritmo de aprendizado supervisionado de retro propagação dos erros (Bryson et al. 1979). A rede neural de retropropagação é treinada com as entradas ajustadas às variáveis de saída em duas fases (Zhang et al. 2008).

Rede MLP

As redes MLP foram interconectadas por forças de conexão representadas por valores que são chamados de pesos sinápticos, os quais são responsáveis pelo armazenamento do conhecimento adquirido. Os valores usados nas camadas de entrada foram normalizados de acordo com a equação 2.

$$y_i = \frac{x_i - x_{max}}{x_{max} - x_{min}} \quad (2)$$

em que, y_i é o valor do vetor de entrada (p. ex.: índices de vegetação modificados), x_{min} e x_{max} são o mínimo e o máximo valores observados, respectivamente.

O valor de saída de cada neurônio na camada k é expresso por $z_k = g(a_k)$ em que g é a função de ativação de a_k e a_k é a função sináptica, a qual é uma combinação linear dos valores de entrada normalizada e os pesos sinápticos (Equação 3).

$$a_k = \sum_j y_j w_{kj} \quad (3)$$

em que, w_{kj} são os pesos sinápticos ligando os y_j valores de entrada com cada k neurônio.

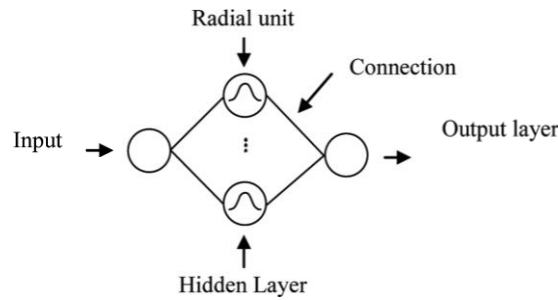
A transferência ou função de ativação nos neurônios de cada camada oculta foi a função hiperbólica, sendo 'e' o número neperiano (Equação 4).

$$g(a_k) = \frac{e^{a_k} - e^{-a_k}}{e^{a_k} + e^{-a_k}} \quad (4)$$

RBF

As redes RBF possuem apenas uma camada oculta e cada neurônio contém uma função de ativação de base radial. Um exemplo de RBF é apresentado na Figura 2.

Fig. 2 Rede neural de função de ativação de base radial



Fonte Adaptado de Savegnago et al. (2011).

Em cada neurônio a função Gaussiana ou normal foi usada como função de base radial (Bishop 1995), e o valores de distância (desvio) dessa função aumentam ou diminuem a relação para o ponto central (Haykin 1998). A função gaussiana é apresentada pela equação 5:

$$\varphi = \exp\left(-\frac{v^2}{2\sigma^2}\right) \quad (5)$$

em que, $v = \|x - \mu\|$ é a distância Euclidiana entre o vetor de entrada e o centro μ da função Gaussiana e σ é a largura. A distância Euclidiana do vetor de entrada para o centro μ é a entrada para função Gaussiana, a qual fornece o valor de ativação da unidade radial.

A saída de cada neurônio na camada de saída é definida pela equação 6:

$$y_j = \sum_{h=1}^H \omega_{jh} \varphi_h(x_s) \quad (6)$$

em que, ω_{jh} é o peso sináptico entre o neurônio h da camada oculta e o neurônio da camada de saída, x_s são os dados de entrada x , e φ_h é a função Gaussiana de cada neurônio na camada oculta.

Assim como na rede neural MLP, os valores de entrada foram normalizados pela equação 2, e os valores na camada de saída forneciam o PMI estimado.

A redes neurais RBF foram treinadas pelo algoritmo k-means (Bishop et al., 1995). Esse algoritmo tenta selecionar o conjunto ótimo de pontos que são colocados nos centroides nos padrões dos dados de treinamento.

Banco de dados para treinamento e validação

Para o treinamento e validação dos modelos, o banco de dados foi dividido em 75% para o treinamento e 25% para validação. Devido às informações terem sido coletadas em diferentes estádios fenológicos, esta percentagem de partição do banco de dados de treinamento e validação foi aplicada a cada estágio de coleta das informações.

Os procedimentos de treinamento e validação dos modelos neurais foram implementados no pacote Neural Networks do software de análise de dados Statistica (Statistica 7.0, Statsoft Inc., Tulsa, OK).

Avaliação do desempenho dos modelos

As RNA's foram avaliadas em termos de acurácia e precisão pelas métricas: Raiz do Erro Quadrático Médio (RMSE) e coeficiente de determinação (R^2) conforme as equações 7 e 8 respectivamente.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{obs} - y_{est})^2}{n}} \quad (7)$$

em que, n é o número de dados, y_{est_i} é o valor da variável estimada pela rede e y_{obs} é o valor da variável observado.

$$R^2 = \frac{SQR}{SQT} \quad (8)$$

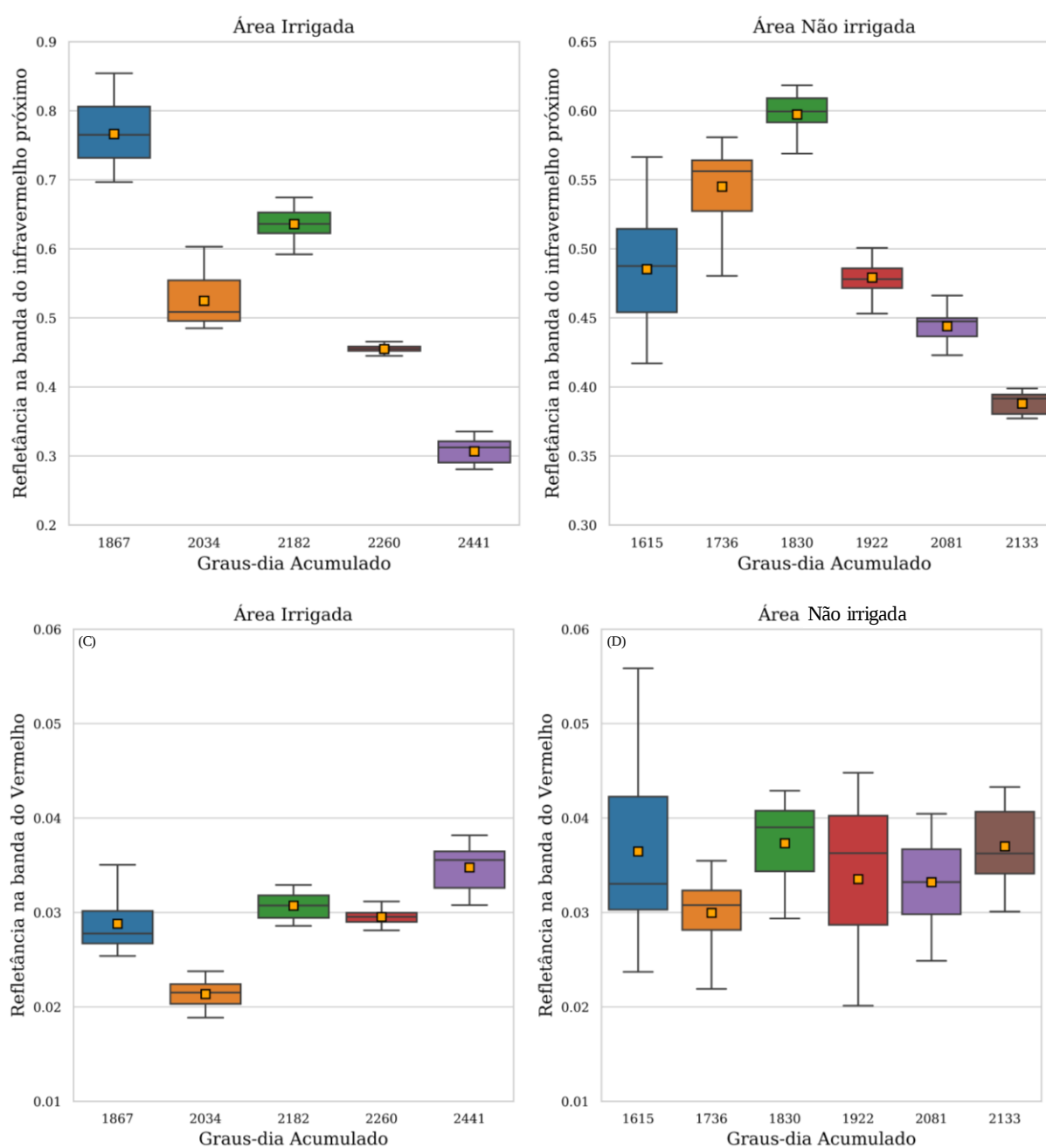
em que, SQR é a soma dos quadrados da regressão e SQT é a soma dos quadrados totais.

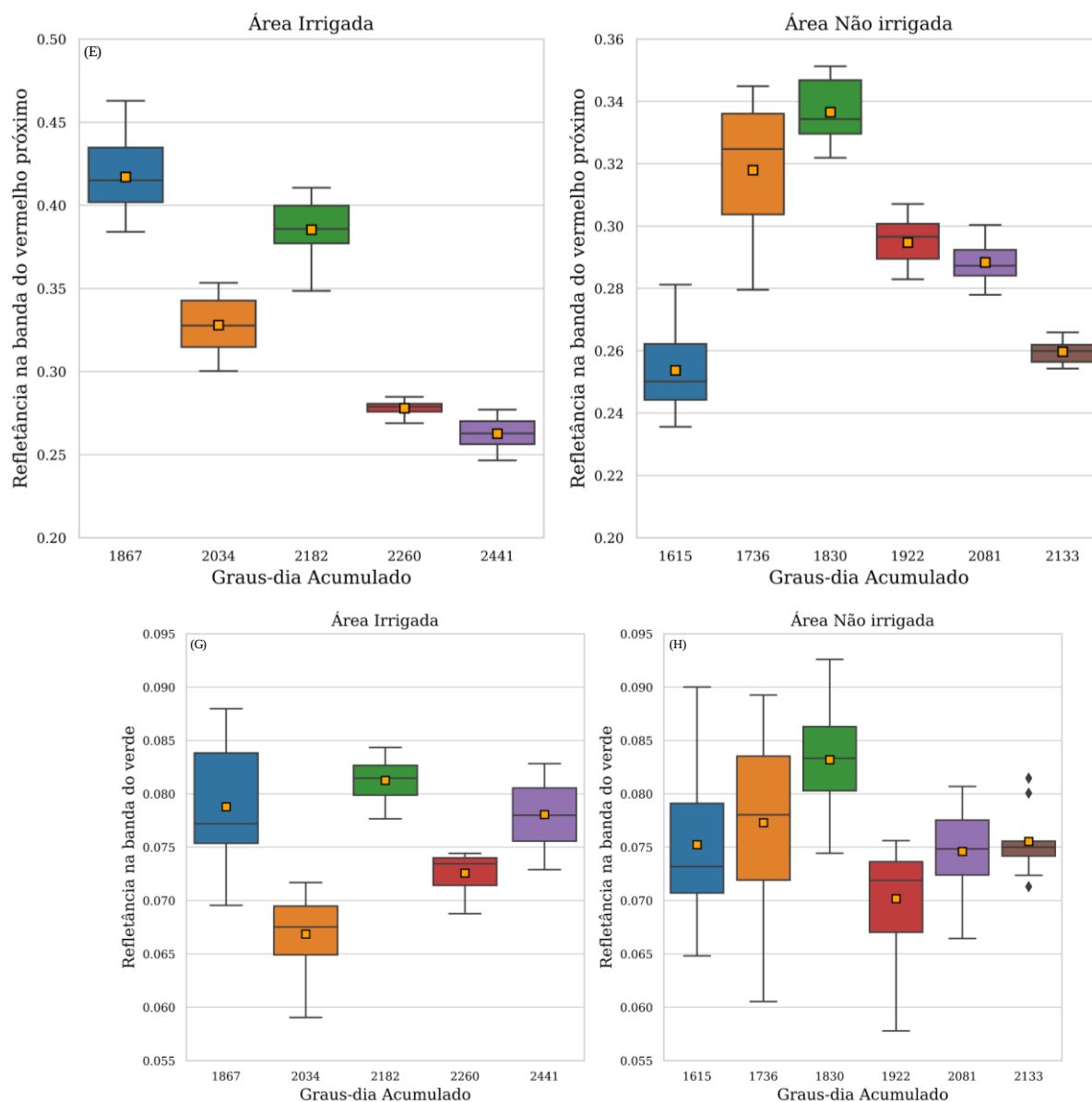
Resultados

O diagrama de caixa foi utilizado para avaliar a variabilidade da refletância das bandas espectrais em função dos Graus-Dia Acumulado (GDA, Figura 3). Nas duas primeiras amostragens independentemente da área foi onde a refletância apresentou maior variabilidade, não obstante, na última coleta a variabilidade foi menor para as áreas avaliadas. As bandas do infravermelho próximo e vermelho próximo, apresentaram tendência não linear e de forma oposta a maturação apresentando média máxima de refletância em 1830 GDA para área não irrigada, no entanto, para área irrigada esse padrão ocorreu em 1867 GDA. Para acesso a variabilidade dos IV's

modificados e não modificados em função dos GDA acessar o material suplementar (Figuras 11 e 12).

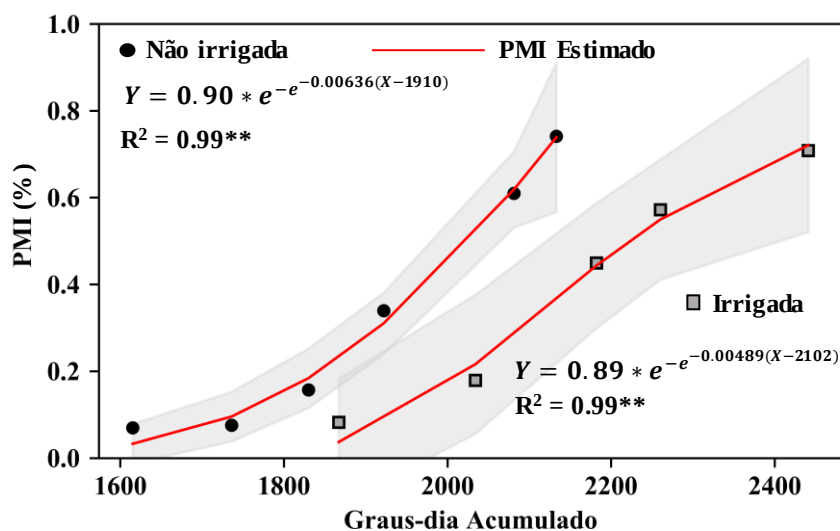
Fig. 3 Distribuição das bandas do infravermelho próximo, vermelho, vermelho próximo e verde nas áreas irrigada e não irrigada em função dos Graus-dia Acumulado. Legenda: média (quadrado central), mediana (linha horizontal interna), terceiro quartil (linha horizontal superior da caixa), primeiro quartil (linha horizontal inferior da caixa), limite superior (linha horizontal acima da caixa), limite inferior (linha horizontal abaixo da caixa)





Analisando o modelo Gompertz ajustado do PMI em função dos GDA para as duas áreas experimentais, foram obtidos modelos significativos e com parâmetros diferentes. O PMI máximo ajustado foi de 0.90 para área não irrigada e 0.89 para área irrigada. Observa-se ainda, que o valor de GDA de máximo crescimento do PMI foi, em torno de 1900 GDA para área não irrigada e 2100 para área irrigada. A taxa de crescimento máxima para as áreas foram próximas, sendo que a área irrigada apresentou taxa máxima de crescimento da maturação em torno de 0.00489 enquanto que a área não irrigada foi ao redor de 0.00636 (Figura 4).

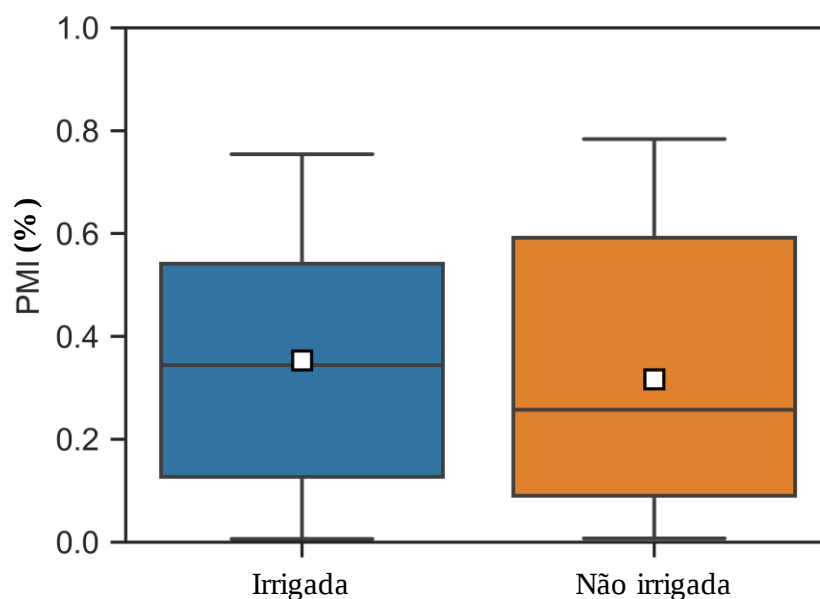
Fig. 4 Modelo Gompertz ajustado para modelar o índice de maturação do amendoim (PMI) em função dos Graus-dia Acumulado para as áreas não irrigada e irrigada, respectivamente.



** p-Valor <0.001

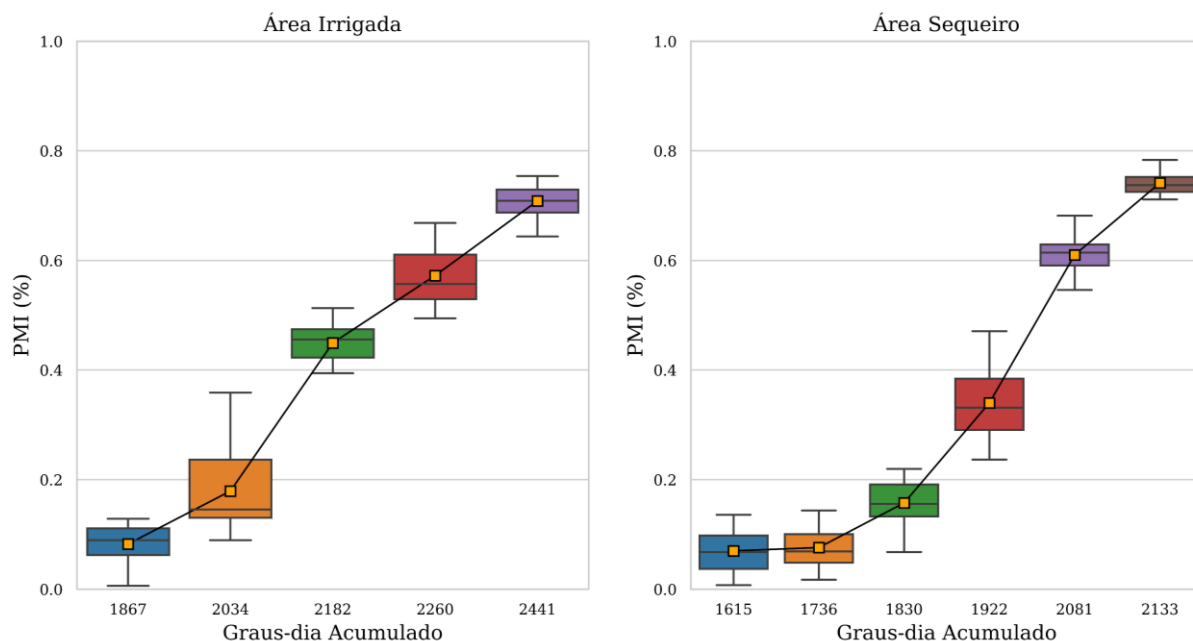
A variabilidade entre as áreas foi próxima, sendo os valores máximos e mínimos próximos. Destaca-se a mediana como a métrica mais diferente pois a área não irrigada apresentou mediana mais próxima do primeiro quartil, indicando distribuição assimétrica positiva (Figura 5).

Fig. 5 Distribuição do índice de maturação do amendoim (PMI) das áreas irrigadas e não irrigada. Legenda: média (quadrado central), mediana (linha horizontal interna), terceiro quartil (linha horizontal superior da caixa), primeiro quartil (linha horizontal inferior da caixa), limite superior (linha horizontal acima da caixa), limite inferior (linha horizontal abaixo da caixa)



Quando se analisa a variabilidade do PMI ao longo dos graus-dia acumulados, observa-se que entre 1900 e 2040 GDD ocorreram as maiores variabilidades do PMI, enquanto que após 2000 GDD a variabilidade tendeu a diminuir, estabilizando quando os valores de PMI ficaram próximos de 0.7 (Figura 6). Para área irrigada a partir dos 2034 GDD a maturação tendeu a aumentar linearmente, enquanto que para área não irrigada esse comportamento se deu por volta dos 1900 GDD (Figura 6).

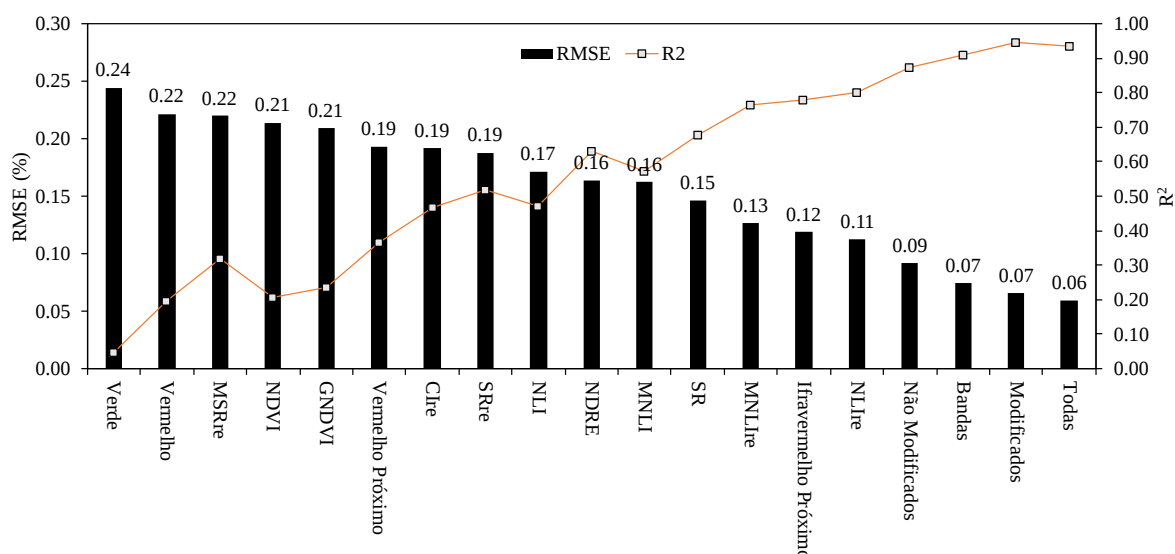
Fig. 6 Evolução temporal do PMI em função dos Graus Dia Acumulado.



Os modelos de RNA's foram utilizados com dados de 15 variáveis derivadas de dois campos comerciais para estimar PMI nas condições de irrigada e não irrigada (Tabela 1). Dentre os dados coletados, 75 % deles foram usados para o treinamento dos modelos e 25% foram usados para validação. Para validação, valores estimados de PMI obtidos dos modelos foram comparados com o PMI obtido em campo. O treinamento e a validação dos modelos derivaram a melhor combinação de parâmetros para estimar a maturação de amendoim.

Para cada dez modelos selecionados pela ferramenta IPS (em cada combinação de topologia e de entrada testada) o modelo com menor RMSE e maior R^2 foi selecionado como melhor arquitetura. De maneira geral, os melhores resultados foram obtidos com o uso de todas as informações, das bandas espectrais, dos IV's não modificados e dos modificados, obtendo modelos com acurácia (RMSE) entre 0.06 e 0.09 e precisão (R^2) acima de 0.85 (Figura 7). Para acessar a topologia resultante das melhores combinações, acessar o material suplementar (Tabela 2).

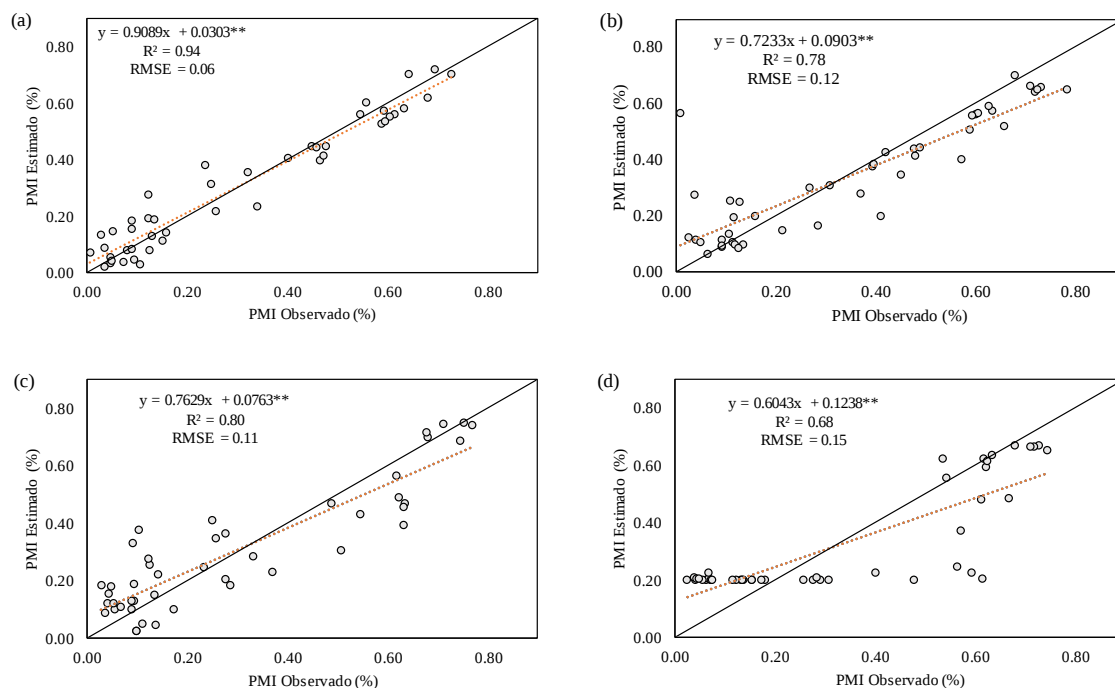
Fig. 7 Acurácia (raiz do erro quadrático médio, RMSE) e precisão (R2) da validação dos modelos.



Quando os modelos foram treinados e validados com apenas uma banda, o NIR foi a variável que estimou o PMI com maior precisão e acurácia (RMSE = 0.12 e $R^2 = 0.78$). Já quando foram utilizados somente índices modificados de forma isolada na camada de entrada, o NLI_{RE} foi o índice com maior acurácia e precisão (RMSE = 0.11 e $R^2 = 0.80$). Usando como entrada os índices não modificados, isoladamente, o SR apresentou melhor acurácia e precisão (RMSE = 0.15 e $R^2 = 0.68$) (Figura 7).

Para avaliar o desempenho dos modelos escolhidos, o PMI estimado pelo modelo resultante foi comparado com o PMI determinado em campo (Figura 8). Os resultados obtidos foram satisfatórios na fase de validação para os modelos escolhidos com a melhor acurácia e precisão, a análise de regressão linear permitiu ajustar uma equação cujos parâmetros tem significância ($p < 0.01$) entre os dados observados e estimados pelos modelos. Entre todas as combinações, quando foram utilizadas todas as variáveis como entrada no modelo foi quando se obteve melhor desempenho (Figura 8 a), pois a reta ajustada e os valores estimados encontram-se próxima da linha 1:1 e o modelo gerou resultados acurados e precisos.

Fig. 8 Desempenho do modelo utilizando todas as variáveis (a), banda infravermelho próximo (b), NLire (c) e SR (d). Legenda: ** p-Valor < 0.01

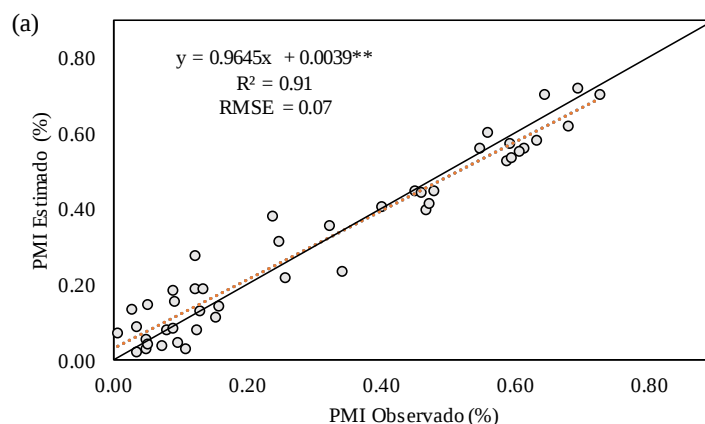


Discussão

Neste estudo, foi proposto um método para estimar o PMI e a partir dos resultados há evidências de que se pode utilizar RNA's e sensoriamento remoto aéreo para tal fim. Com base na análise de desempenho, acurácia, precisão e também pensando no ponto de vista da aplicação do método, a arquitetura mais indicada é aquela obtida utilizando somente as bandas espectrais (Figura 7).

Apesar de a combinação utilizando todas as informações ter obtido maior acurácia e precisão ($RMSE = 0.06$ e $R^2 = 0.94$), esta arquitetura necessita de muitas entradas e possui um número de operações matemáticas expressiva, pois as entradas envolvem as bandas isoladas, índices de vegetação modificados e não modificados para alimentação do modelo. Uma alternativa parcimoniosa é o uso do modelo utilizando as bandas verde, vermelho, vermelho próximo e infra-vermelho próximo como variáveis de entrada, uma vez que este modelo apresentou acurácia e precisão aceitáveis ($RMSE = 0.07$ e $R^2 = 0.91$) (Figura 9). As variáveis escolhidas e a topologia das redes neurais selecionadas deste estudo podem ser acessadas no material suplementar apresentado ao final do artigo.

Fig. 9 Análise de desempenho do modelo utilizando as bandas espectrais como entrada



O erro do modelo resultante (usando as cinco bandas) de sete por cento na estimativa do PMI foi baixo, pois considerando o ponto ótimo para a colheita que é 0.7, caso haja uma subestimação ou super estimacão esse valor irá variar de 0.65 a 0.75, o que segundo pesquisas publicadas essa faixa de variação não compromete a qualidade do grão e também não interfere nas perdas durante a colheita.

Observando a análise de sensibilidade, constata-se que a ordem de importância das bandas foi a seguinte: 1º Infravermelho próximo, 2º Vermelho, 3º Vermelho próximo e 4º Verde (Tabela 3, material suplementar). Existe uma relação direta entre o quanto de radiação fotossinteticamente ativa é absorvida e espalhada pela folha das plantas e suas propriedades biogeoquímicas (Tu et al. 2018). A estrutura da folha e do dossel estão relacionadas com o infravermelho próximo e a concentração dos pigmentos fotossintéticos se relacionam com as bandas do visível (verde, vermelho e azul) (Jensen 2015; Sellers 1987).

Para dados de sensoriamento remoto multiespectral a banda do NIR apresenta maior reflectância para vegetação verde em função do alto espalhamento interno da folha, enquanto que a banda do vermelho apresenta baixa reflectância em função da absorção de radiação pela clorofila, no entanto, uma curva de reflectância ocorre entre essas bandas chamada de Red-Edge (RE) (Horler et al. 1983).

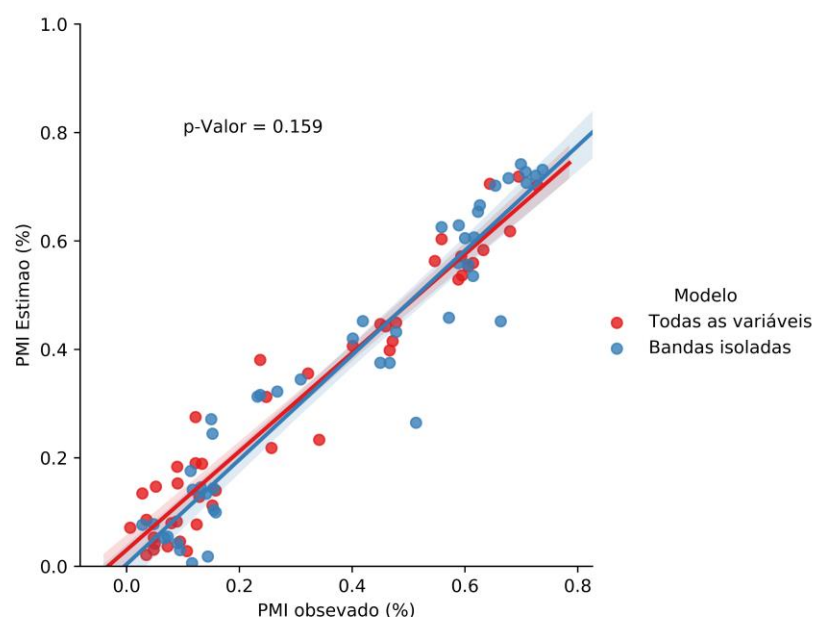
A banda RE tem alta correlação com diversos parâmetros fisiológicos da vegetação, como conteúdo de nitrogênio, conteúdo de clorofila e biomassa (Clevers and Gitelson 2013; Filella and Penuelas 1994). A ocorrência dessa mudança na reflectância da vegetação reflete as mudanças na condição biológica das plantas

(Boochs et al. 1990)). Além do mais, o ponto de inflexão da banda red edge, o qual é menos sensível a ruídos espectrais causados pelo sol e condições atmosféricas quando utilizado para estimar o conteúdo de clorofila (Cho et al. 2008; Cho and Skidmore 2006).

A banda verde apresenta menor importância de acordo com a sensibilidade, o que é explicado pela sua baixa mudança de refletância de vagens em diferentes índices de maturação, como observado por Rowland et al. (2008) onde os autores demonstraram que a refletância na banda do verde foi igual para vagens com índice de maturação 0.44 e 0.88. Já para a maior importância do NIR e do vermelho no modelo, pode ser explicado pela sensibilidade dessa região espectral às alterações da maturação das vagens, uma vez que Rowland et al. (2008) evidenciaram que essas bandas apresentaram valores de refletância significativamente diferentes para índices de maturação de 0.32, 0.91 e de 0.44 e 82. O red edge ocupar o terceiro lugar em importância do modelo, corrobora Santos et al. (2019), que encontraram forte correlação dessa banda com o índice de maturação do amendoim.

O modelo escolhido como mais adequado do ponto de vista de execução das estimativas do PMI de forma mais parcimoniosa não diferiu (p -Valor = 0.159) do modelo com maior acurácia e precisão (Figura 10).

Fig. 10 Comparação de modelos pela Análise de Covariância (ANCOVA).



Trabalhos futuros podem focar no teste e calibração desse modelo em áreas comerciais com abrangência regional e possivelmente até nacional. Modelos de estimativa para diversas culturas utilizando parâmetros espectrais como bandas e índices de vegetação podem ser desenvolvidos nas mesmas linhas.

Conclusões

O presente trabalho foca o uso de redes neurais artificiais para estimar a maturação de amendoim utilizando dados de sensoriamento remoto aéreo e informações amostradas em campo. O estudo enfatiza o uso de dados das bandas espectrais (vermelho, verde, infravermelho próximo e vermelho próximo) para capturar a maturação de amendoim e obter os resultados mais parcimoniosos. RNA's foram empregadas para identificar preditores importantes entre bandas espectrais, índices de vegetação modificados e não modificados para estimar a maturação de amendoim a qual teve boa precisão e acurácia ($R^2 = 0.91$, RMSE = 0.07) para estimar a maturação determinada em campo.

As várias bandas espectrais derivadas de imagens de drone juntas aos parâmetros da rede resultante permitem a estimação da maturação de amendoim com elevada acurácia e de forma não destrutiva.

Referências

- ABADIE, J.; CARPENTIER, J. Généralisation de la méthode du gradient réduit de Wolfe au cas de contraintes non-linéaires. In: **Proceedings IFORS Conference**. 1965. p. 1041-1053.
- ABD-EL MONSEF, H. et al. Using multispectral imagery to extract a pure spectral canopy signature for predicting peanut maturity. **Computers and Electronics in Agriculture**, v. 162, p. 561-572, 2019.
- BISHOP, C.M. et al. **Neural networks for pattern recognition**. Oxford university press, 1995.
- BOOCHS, F. et al. Shape of the red edge as vitality indicator for plants. **Remote sensing**, v. 11, n. 10, p. 1741-1753, 1990.
- BRYSON, G.M. et al.. Applied optimal control: Optimization, estimation, and control. **IEEE Transaction on Systems, Man and Cybernetics**, v. 9, p. 366-367, 1979.

CHO, M.A.; SKIDMORE, A.K.; ATZBERGER, C. Towards red-edge positions less sensitive to canopy biophysical parameters for leaf chlorophyll estimation using properties optiques spectrales des feuilles (PROSPECT) and scattering by arbitrarily inclined leaves (SAILH) simulated data. **International Journal of Remote Sensing**, v. 29, n. 8, p. 2241-2255, 2008.

CHO, M.A.; SKIDMORE, A.K. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. **Remote sensing of environment**, v. 101, n. 2, p. 181-193, 2006.

CLEVERS, J.G.P.W; GITELSON, A.A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. **International Journal of Applied Earth Observation and Geoinformation**, v. 23, p. 344-351, 2013.

COLVIN, B.C. et al. Consideration of peg strength and disease severity in the decision to harvest peanut in southeastern USA. **Journal of Crop Improvement**, v. 32, n. 3, p. 287-304, 2018.

FILELLA, I.; PENUELAS, J. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. **International journal of remote sensing**, v. 15, n. 7, p. 1459-1470, 1994.

FITZGERALD, G.; RODRIGUEZ, D.; O'LEARY, G. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). **Field Crops Research**, v. 116, n. 3, p. 318-324, 2010.

GITELSON, A.A.; KAUFMAN, Y.J.; MERZLYAK, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289-298, 1996.

GOEL, N.S.; QIN, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. **Remote Sensing Reviews**, v. 10, n. 4, p. 309-347, 1994.

GOMPERTZ, B. XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. **Philosophical transactions of the Royal Society of London**, n. 115, p. 513-583, 1825.

GONG, P. et al. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. **IEEE transactions on geoscience and remote sensing**, v. 41, n. 6, p. 1355-1362, 2003.

HORLER, D.N.H.; DOCKRAY, M.; BARBER, J. The red edge of plant leaf reflectance. **International journal of remote sensing**, v. 4, n. 2, p. 273-288, 1983.

JIANG, D. et al. An artificial neural network model for estimating crop yields using remotely sensed information. **International Journal of Remote Sensing**, v. 25, n. 9, p. 1723-1732, 2004.

JORDAN, C.F. Derivation of leaf-area index from quality of light on the forest floor. **Ecology**, v. 50, n. 4, p. 663-666, 1969.

KHAN, M.S. et al. An artificial neural network model for estimating Mentha crop biomass yield using Landsat 8 OLI. **Precision Agriculture**, v. 21, n. 1, p. 18-33, 2020.

MIAO, Y.; MULLA, D.J.; ROBERT, P.C. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. **Precision Agriculture**, v. 7, n. 2, p. 117-135, 2006.

ROUSE, J.W. et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA special publication**, v. 351, n. 1974, p. 309, 1974.

ROWLAND, D.L. et al. Canopy characteristics and their ability to predict peanut maturity. 2008.

ROWLAND, D.L. et al. Determination of maturity and degree day indices and their success in predicting peanut maturity. 2006.

SANTOS, A.F. et al. Using remote sensing to map in-field variability of peanut maturity. In: **Precision Agriculture 2019-Papers Presented at the 12th European Conference on Precision Agriculture, ECPA 2019**. 2019. p. 605-611.

SAVEGNAGO, R.P. et al. Comparison of logistic and neural network models to fit to the egg production curve of White Leghorn hens. **Poultry science**, v. 90, n. 3, p. 705-711, 2011.

SELLERS, P.J. Canopy reflectance, photosynthesis, and transpiration, II. The role of biophysics in the linearity of their interdependence. **Remote sensing of Environment**, v. 21, n. 2, p. 143-183, 1987.

SIMON, H. **Neural networks: a comprehensive foundation**. Prentice hall, 1999.

TU, Y. et al. Assessing radiometric correction approaches for multi-spectral UAS imagery for horticultural applications. **Remote Sensing**, v. 10, n. 11, p. 1684, 2018.

UNGER HOLTZ, T.S. Introductory digital image processing: A remote sensing perspective. 2007.

WILLIAMS, E.J.; DREXLER, J.S. A non-destructive method for determining peanut pod maturity. **Peanut Science**, v. 8, n. 2, p. 134-141, 1981.

WU, C. et al. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. **Agricultural and forest meteorology**, v. 148, n. 8-9, p. 1230-1241, 2008.

XIE, Q. et al. Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 80, p. 187-195, 2019.

YANG, Q. et al. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. **Field Crops Research**, v. 235, p. 142-153, 2019.

ZHANG, L et al. A new sub-pixel mapping algorithm based on a BP neural network with an observation model. **Neurocomputing**, v. 71, n. 10-12, p. 2046-2054, 2008.

Material suplementar

Tabela 2 Sumário dos modelos de rede neural selecionados

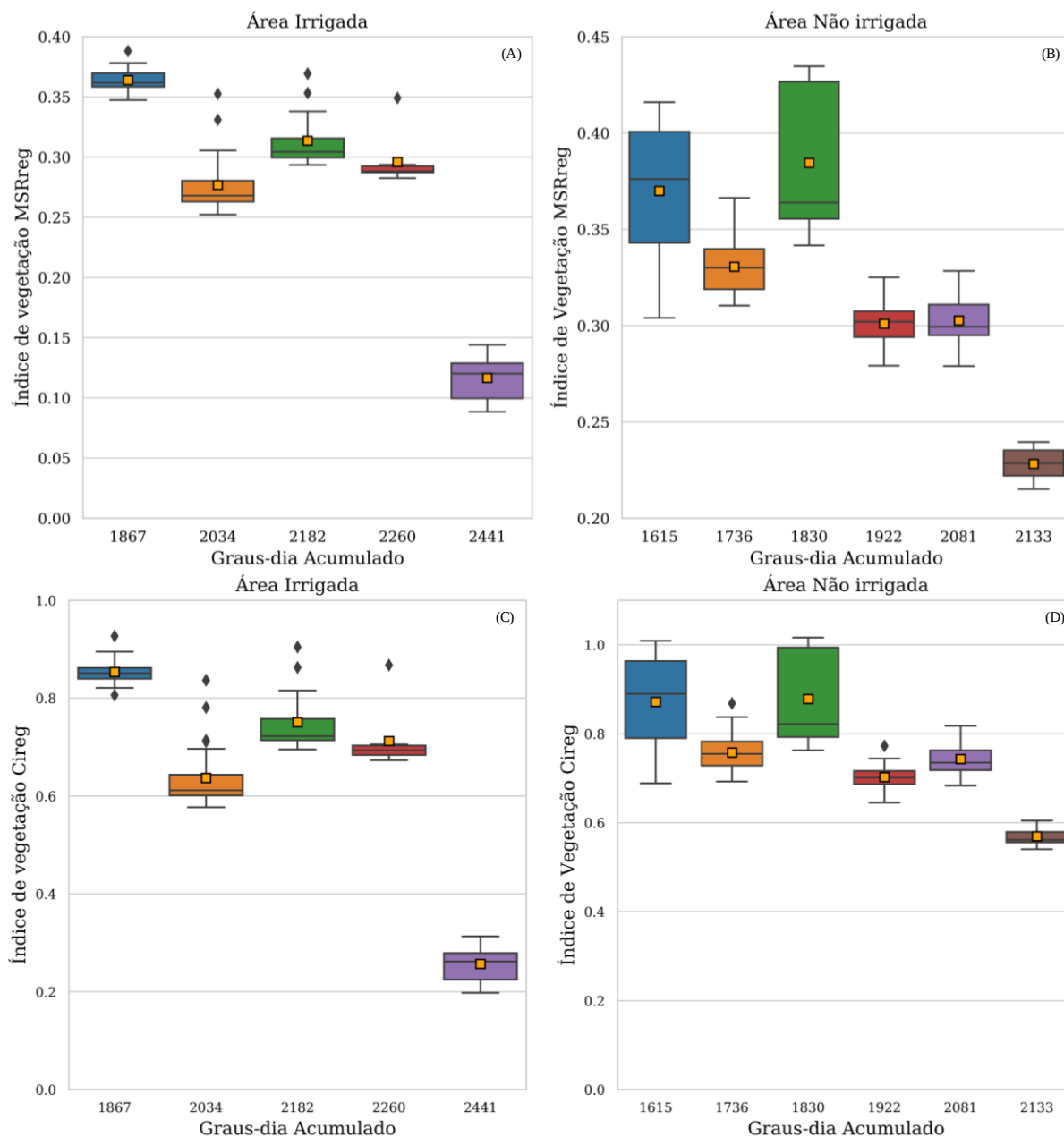
Variável Dependente	Variáveis independentes	Perfil do modelo
Peanut Maturation Index	Todas	MLP 12:12-30-19-1:1
	Bandas	MLP 4:4-30-22-1:1
	Não modificados	MLP 3:3-30-11-1:1
	Modificados	MLP 5:5-30-21-1:1
	Green	MLP 1:1-30-16-1:1
	Red-Edge	RBF 1:1-29-1:1
	Red	MLP 1:1-6-1:1
	NIR	MLP 1:1-30-21-1:1
	MNLlreg	RBF 1:1-14-1:1
	MSreg	MLP 1:1-30-21-1:1
	NLlreg	RBF 1:1-42-1:1
	SRreg	MLP 1:1-30-9-1:1
	Cireg	RBF 1:1-7-1:1
	GNDVI	MLP 1:1-30-30-1:1
	MNLI	MLP 1:1-30-14-1:1
	NDVI	MLP 1:1-18-13-1:1
	NDRE	RBF 1:1-14-1:1
	NLI	MLP 1:1-30-7-1:1
	SR	MLP 1:1-1-1:1

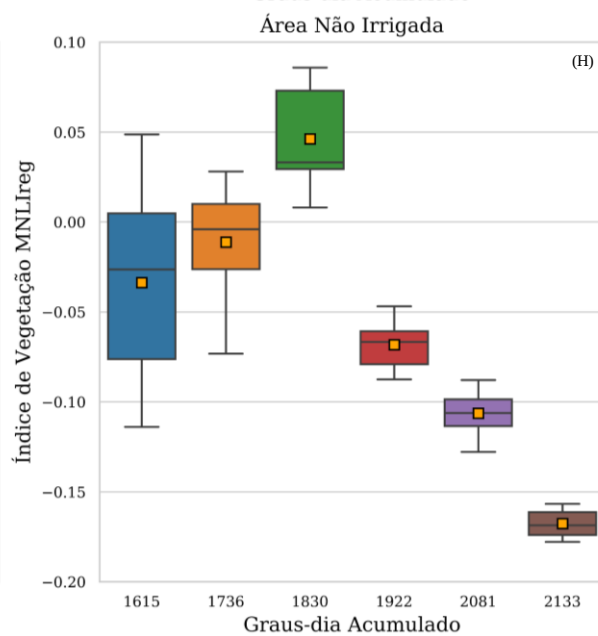
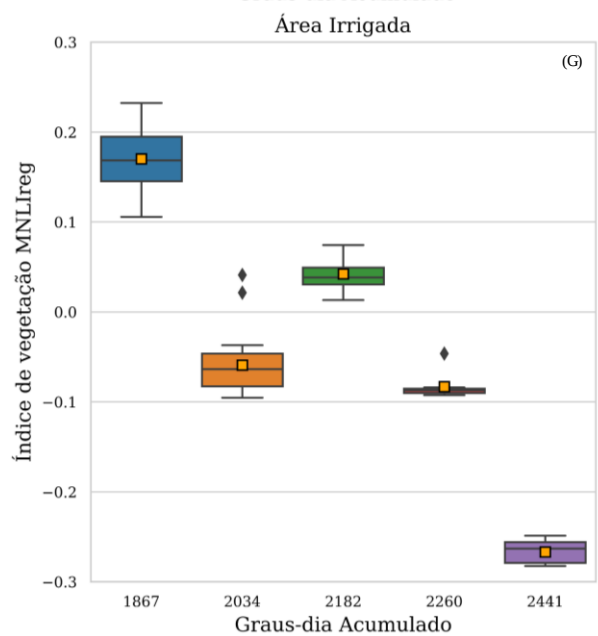
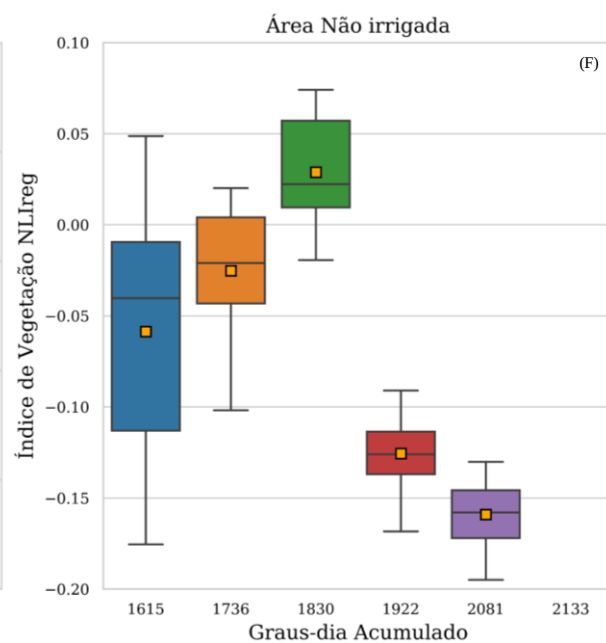
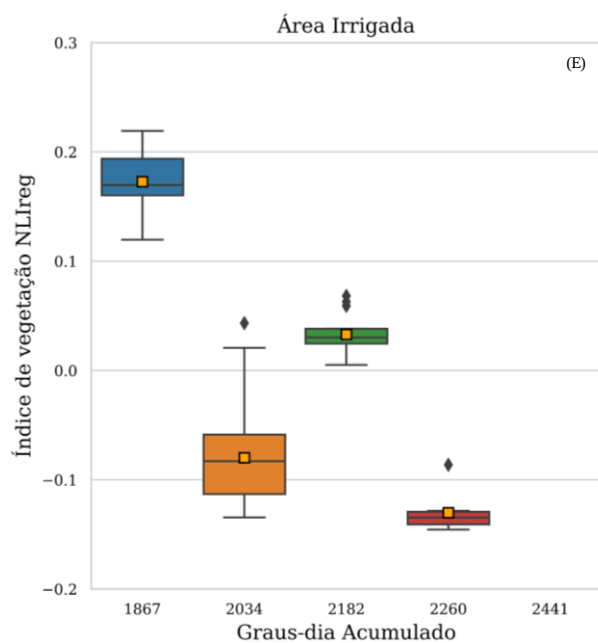
RBF: Redes de Função de Base Radial; MLP: Redes Perceptron Multicamadas; Arquitetura de rede neural é apresentada na forma de I:N1-N2-N3:O, em que “I” é o número de variáveis de entrada, “N1”, “N2” e “N3” são os números de neurônios nas camadas de Entrada, Oculta e Saída, respectivamente, e “O” é o número de variáveis de saída

Tabela 3 Análise de sensibilidade para as combinações com múltiplas entradas

Bandas e IV's	Combinações		
	Todas as variáveis	Bandas	IV's não modificados Modificados
Green	12	4	
NIR	3	1	
Red	11	2	
Red_Edge	4	3	
Cireg	2		2
MLNIreg	9		4
MSRreg	1		1
NLlreg			3
SRreg			5
GNDVI	6		
MLNI		1	
NDRE	5	2	
NDVI	8		
NLI	7	3	
SR	10		

Fig. 11 Distribuição dos índices de vegetação modificados nas áreas irrigada e não irrigada, em função dos Graus-dia Acumulado. Legenda: média (quadrado central), mediana (linha horizontal interna), terceiro quartil (linha horizontal superior da caixa), primeiro quartil (linha horizontal inferior da caixa), limite superior (linha horizontal acima da caixa), limite inferior (linha horizontal abaixo da caixa)





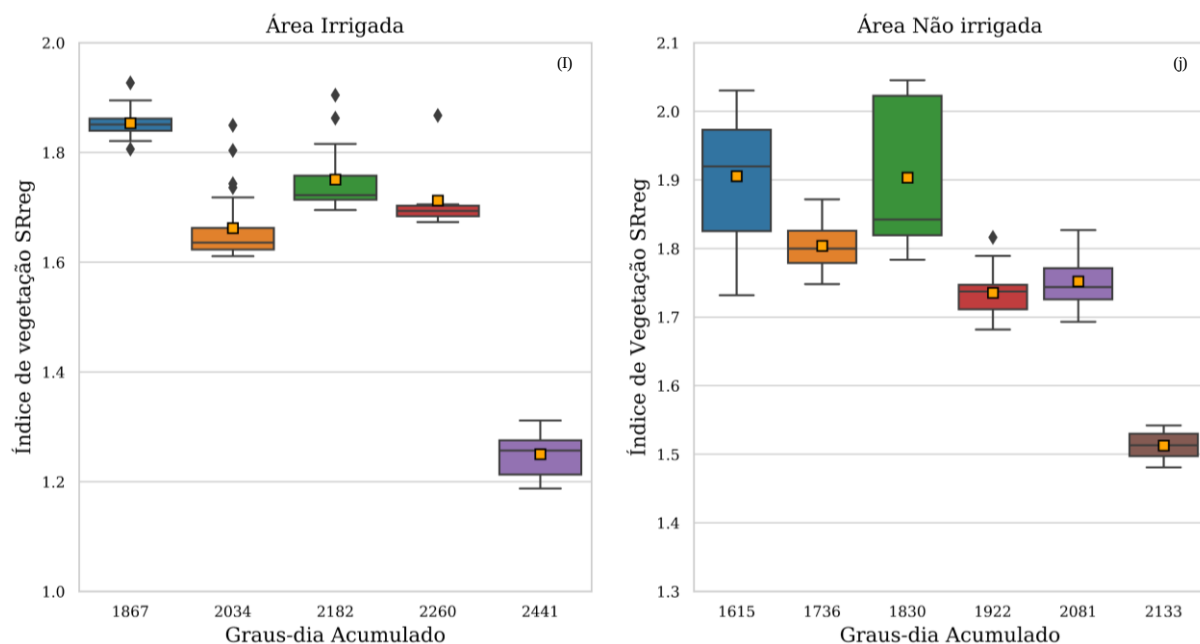
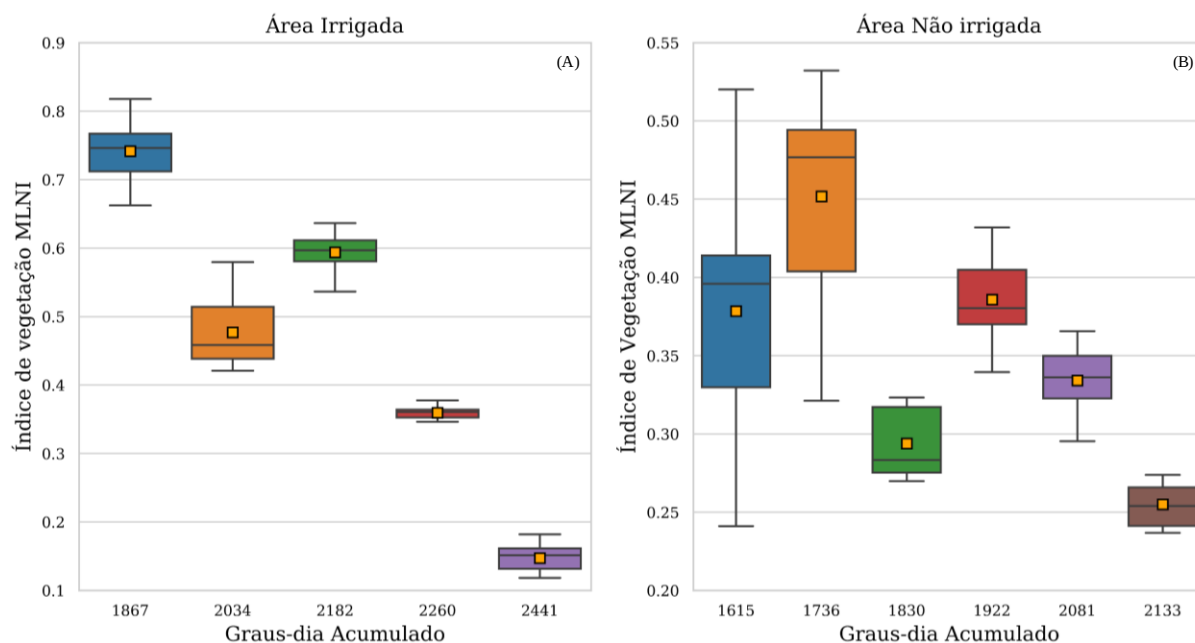
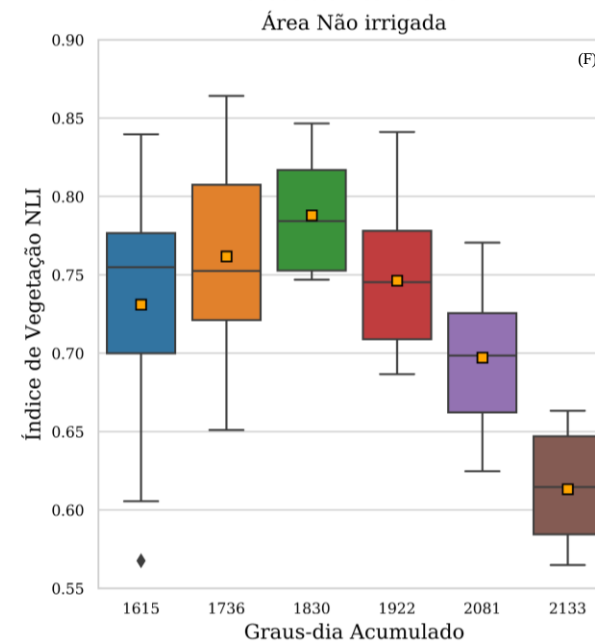
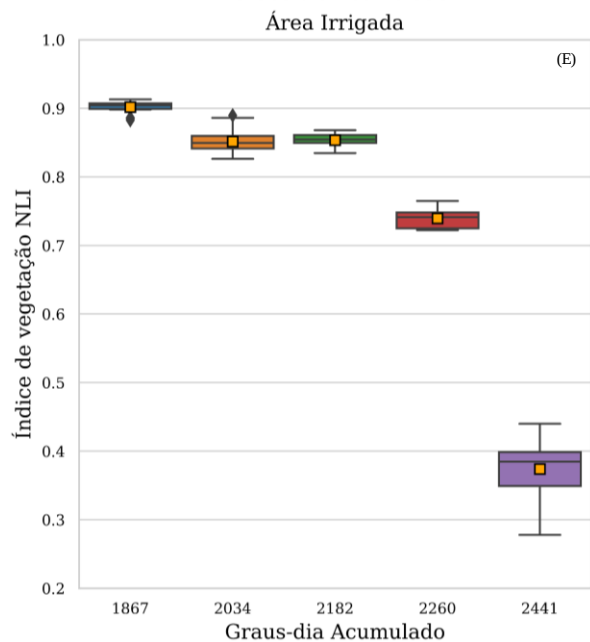
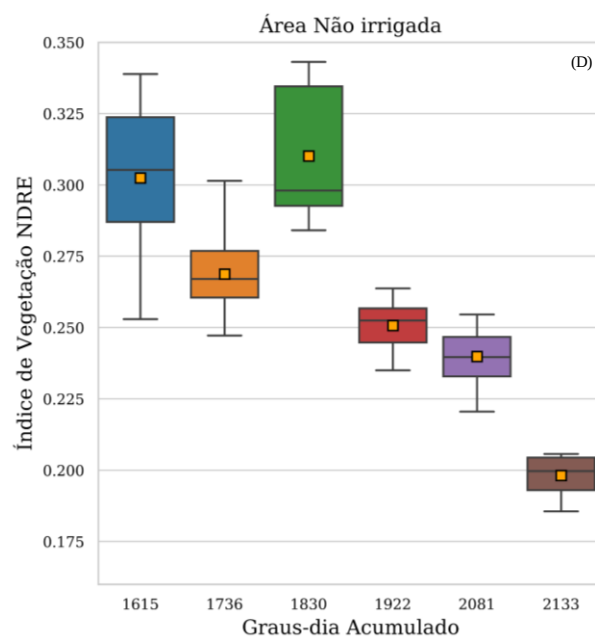
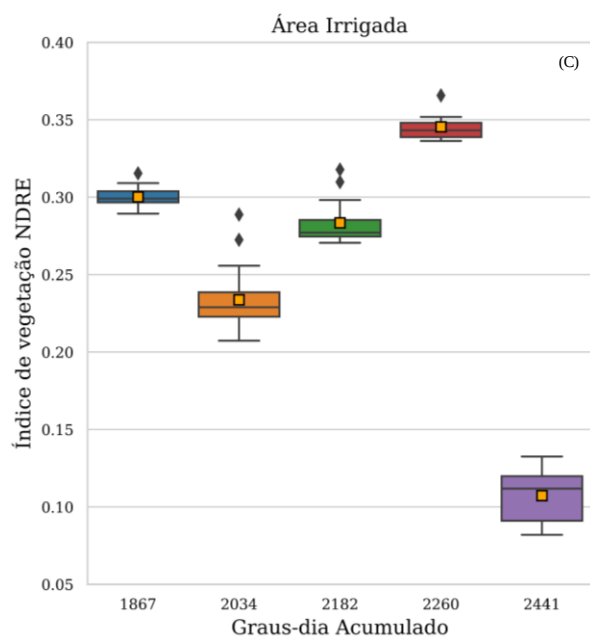
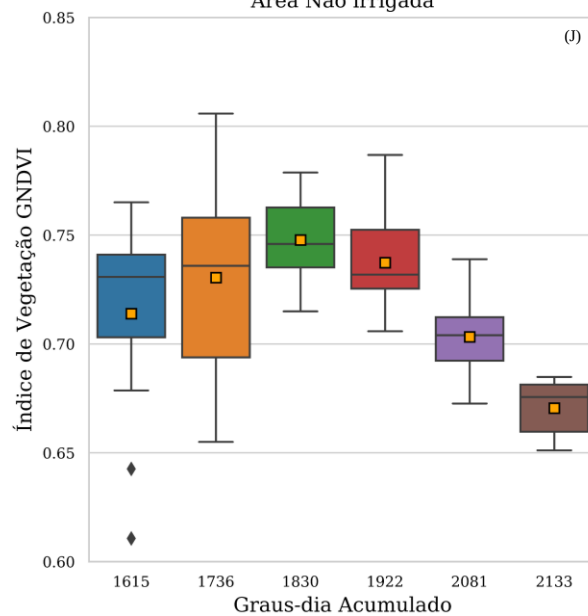
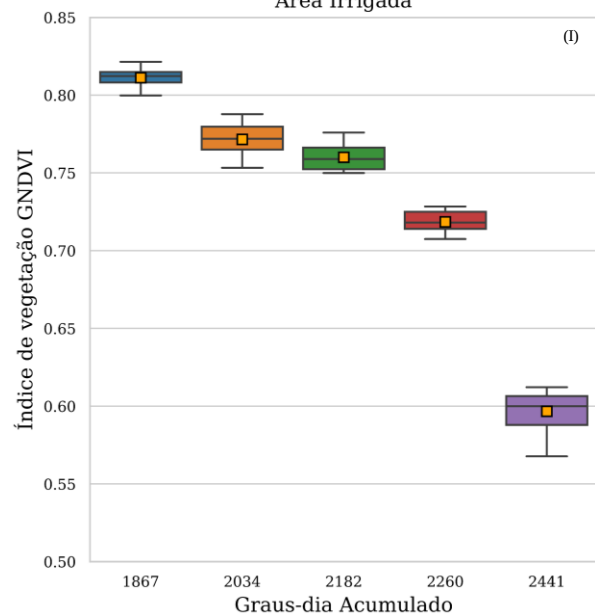
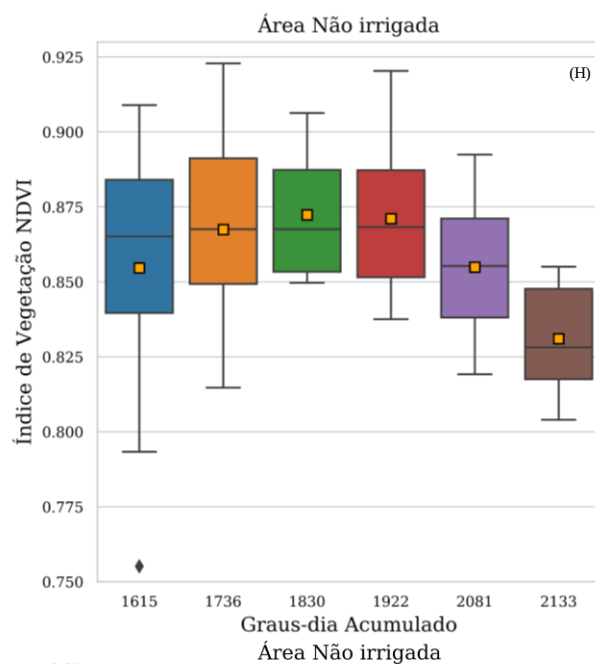
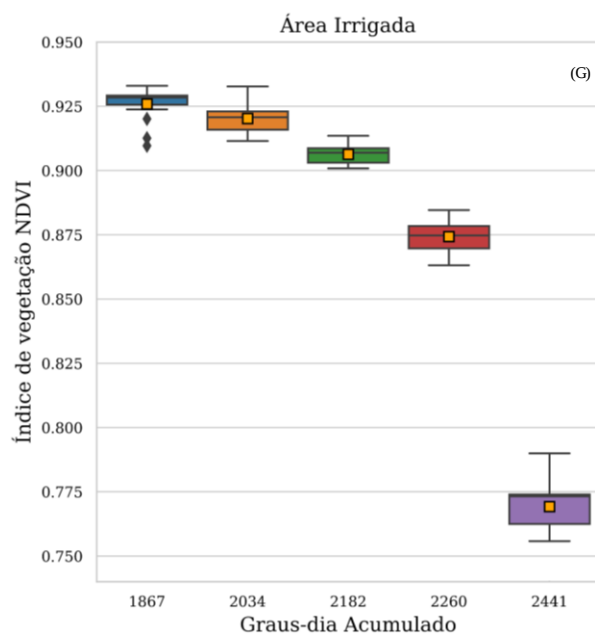
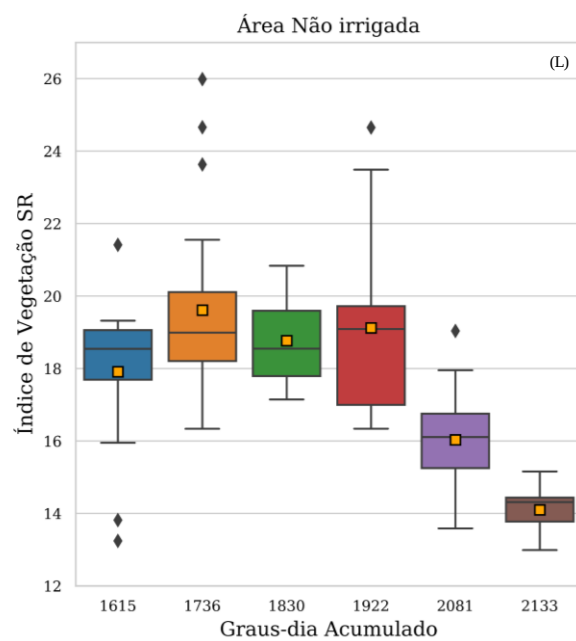
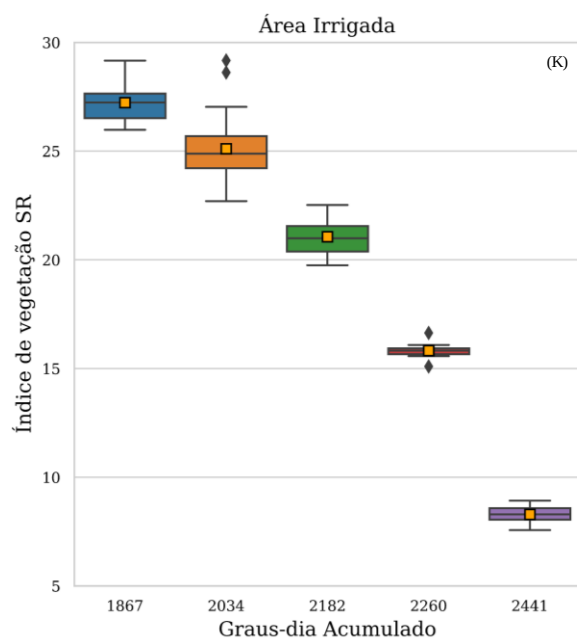


Fig. 12 Distribuição dos índices de vegetação não modificados nas áreas irrigada e não irrigada, em função dos Graus-dia Acumulado. Legenda: média (quadrado central), mediana (linha horizontal interna), terceiro quartil (linha horizontal superior da caixa), primeiro quartil (linha horizontal inferior da caixa), limite superior (linha horizontal acima da caixa), limite inferior (linha horizontal abaixo da caixa)









CAPÍTULO 5 - CONSIDERAÇÕES FINAIS

A medição direta de variáveis agronômicas é um processo complexo de execução por meios manuais, o que a torna impraticável para grandes áreas, principalmente pela grande demanda de mão de obra e por estar sujeita a erros de observação e de mensuração. Desta forma, utilizar de relações funcionais de covariáveis para previsões de variáveis agronômicas pode contribuir para o monitoramento e desenvolvimento de tecnologias aplicáveis e escaláveis para grandes áreas.

Nesta pesquisa, foram desenvolvidos métodos para previsão de produtividade da cultura do milho ao nível de zona de manejo, desenvolvimento de um algoritmo para determinação de volume aplicação variável para cultura do café e, por fim, um método para estimativa da maturação de amendoim baseado em sensoriamento remoto e redes neurais artificiais.

Verificou-se que a previsão de produtividade de milho ao nível de zona de manejo é possível, utilizando-se bandas espectrais e índices topográficos (TWI + TPI) e o aprendizado de máquinas. Foi observado também que modelos em sítio específico são necessários para o incremento de acurácia e redução da tendência das previsões.

Por meio do aprendizado do padrão de variáveis físicas da cultura do café e índices de vegetação é possível implementar uma nova estratégia de determinação do volume de calda para lavouras de café. Esta estratégia pode ser realizada usando RNA do tipo multicamadas para estimar a altura e o diâmetro das plantas, tendo como entrada os índices de vegetação NDVI direito e NDRE esquerdo e topo. Isto permite o cálculo do volume para o campo de cultivo e, por meio de equações auxiliares, gerar a variação do volume de aplicação.

Para a estimativa não destrutiva da maturação de amendoim foi descoberto que, com a utilização de bandas espectrais como entrada no modelo neural resultante, se consegue estimar a maturação de forma acurada e precisa.

Os resultados encontrados nestas pesquisas podem ser considerados promissores na previsão indireta de variáveis agronômicas, associando aprendizado de máquinas, sensoriamento remoto e índices topográficos. Baseado nisto, os próximos passos destas pesquisas são a validação a campo dos modelos e métodos desenvolvidos. Para a técnica de previsão de produtividade da cultura do milho usando índices topográficos faz-se necessário pesquisas utilizando outros índices

topográficos além do experimento em diferentes condições de relevo, para o entendimento das relações da topografia na modelagem da produtividade. Já para o método desenvolvido para o café é necessário a ampliação do método para outras variedades de café e espécies como o café conilon uma vez que esta espécie possui hábitos de crescimento distintos do café arábica no qual foi desenvolvido o método, além do desenvolvimento de modelos robustos de estimativa das variáveis físicas do café englobando diferentes condições de cultivo como o cultivo de café em montanha e lavouras em idades de cultivo diferente. Para o método de estimativa da maturação de amendoim pesquisas futuras podem expandir as condições de coleta dos dados, como em diferentes tipos de solo e métodos de semeadura, pois o amendoim pode ser cultivado em diversos métodos de semeadura como o cultivo mínimo, preparo convencional além de diferentes espaçamentos como em fileiras simples e fileiras duplas.