

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Douglas Henrique Haubert Cerutti

**COMBINAÇÃO DE IMAGENS MULTI-ESPECTRAIS E
MODELO DE BALANÇO HÍDRICO PARA PREDIÇÃO DE
RENDIMENTOS DA SOJA NO SUL DO BRASIL**

Santa Maria, RS
2019

Douglas Henrique Haubert Cerutti

**COMBINAÇÃO DE IMAGENS MULTI-ESPECTRAIS E MODELO DE BALANÇO
HÍDRICO PARA PREDIÇÃO DE RENDIMENTOS DA SOJA NO SUL DO BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Ciência da Computação**.

Orientador: Prof. Dr. João Carlos Damasceno Lima

Co-orientadora: Prof. Dr^a. Mirta Teresinha Petry

Santa Maria, RS

2019

Cerutti, Douglas Henrique Haubert

Combinação de imagens multi-espectrais e modelo de balanço hídrico para predição de rendimentos da soja no sul do Brasil / por Douglas Henrique Haubert Cerutti. – 2019.

77 f.: il.; 30 cm.

Orientador: João Carlos Damasceno Lima

Co-orientadora: Mirta Teresinha Petry

Dissertação (Mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Pós-Graduação em Ciência da Computação, RS, 2019.

1. Predição de rendimentos. 2. Sensoriamento remoto. 3. Soja. 4. Imagens multi-espectrais. 5. Classificação supervisionada. 6. Aprendizado de máquina. I. Lima, João Carlos Damasceno. II. Petry, Mirta Teresinha. III. Título.

© 2019

Todos os direitos autorais reservados a Douglas Henrique Haubert Cerutti. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

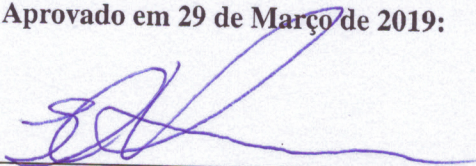
E-mail: dhaubert@inf.ufsm.br

Douglas Henrique Haubert Cerutti

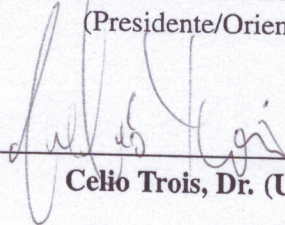
**COMBINAÇÃO DE IMAGENS MULTI-ESPECTRAIS E MODELO DE BALANÇO
HÍDRICO PARA PREDIÇÃO DE RENDIMENTOS DA SOJA NO SUL DO BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Ciência da Computação**.

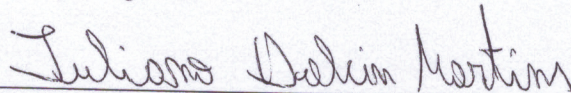
Aprovado em 29 de Março de 2019:



João Carlos Damasceno Lima, Dr. (UFSM)
(Presidente/Orientador)



Celio Trois, Dr. (UFSM)



Juliano Dalcin Martins, Dr. (IFRS)

Santa Maria, RS

2019

AGRADECIMENTOS

Agradeço inicialmente à pessoa que possibilitou que eu pudesse traçar meu próprio caminho enquanto ser humano, com amor incondicional, colocando minhas necessidades acima das suas próprias, minha mãe, Verena Haubert. Agradeço a Deus pelo dom da vida e pelos desafios que me fazem aprender todos os dias.

Agradeço à Universidade Federal de Santa Maria (UFSM) pelo ensino de qualidade e ao Programa de Pós-Graduação em Ciência da Computação pela estrutura e organização que possibilitam que diversas pessoas trilhem mais um passo em direção aos seus objetivos gratuitamente.

Agradeço aos meus orientadores João Carlos Damasceno Lima e Mirta Teresinha Petry por todos os conselhos, direcionamentos, reuniões e principalmente pela compreensão durante toda a jornada do mestrado.

Agradeço à companheira de experimento Marília Cossetin Antonello que sempre esteve disposta a ajudar e foi fundamental para que esta dissertação se tornasse realidade. Estendo o agradecimento aos seus pais Beto e Mara, à Heloísa e ao Paulo, e ao seu tio Renato, hoje todos meus amigos, por me receberem sempre contentes e de boa vontade nas Agropecuária Dois Pinheiros. Agradeço ao amigo Leonardo Cechi pelas conversas, orientações e simulações que tornaram esse trabalho possível.

Agradecimento para meus colegas do Sistema Irriga, em especial para Tobias Senger que durante todo o período foi um irmão, de uma ajuda incansável em disciplinas, trabalhos, entregas, apoios emocionais e profissionais durante toda essa jornada. Agradeço à empresa pela oportunidade de abordar um tema relacionado à agricultura, ao apoio com equipamentos e flexibilização de horários que permitiram cursar as disciplinas e demais obrigações presenciais.

Ainda, agradeço à todas as pessoas que direta ou indiretamente colaboraram com o trabalho, seja com revisões, opiniões ou vibrações positivas.

RESUMO

COMBINAÇÃO DE IMAGENS MULTI-ESPECTRAIS E MODELO DE BALANÇO HÍDRICO PARA PREDIÇÃO DE RENDIMENTOS DA SOJA NO SUL DO BRASIL

AUTOR: DOUGLAS HENRIQUE HAUBERT CERUTTI

ORIENTADOR: JOÃO CARLOS DAMASCENO LIMA

CO-ORIENTADORA: MIRTA TERESINHA PETRY

As mudanças climáticas nos últimos anos projetam um horizonte em que os recursos hídricos se tornem cada vez mais escassos e, por este motivo, irão afetar consideravelmente a produção de alimentos no mundo. Um desafio constante para a agricultura é manter o nível de produção elevado e empregar de forma inteligente e racional a água em complemento às condições climáticas para que se possa obter o melhor rendimento com precisão na aplicação de irrigação. Atualmente, o sensoriamento remoto permite o monitoramento contínuo das culturas, sobretudo em virtude das correspondências do estresse hídrico e das características biofísicas da vegetação durante seu crescimento. A combinação de índices de vegetação com as características da planta analisadas em nível de campo fornecem informações valiosas ao manejo e monitoramento de irrigação durante o desenvolvimento das culturas. Neste estudo, foram analisadas as características da cultura da soja (*Glycine Max*), na região central do Rio Grande do Sul, Brasil, para área com total disponibilidade hídrica e área de sequeiro. Dados em campo foram coletados durante todo o período de cultivo, registrando informações como estágio fenológico, fração de cobertura do solo, índice de área foliar, altura da planta e rendimento dos grãos. Estes dados foram usados tanto para simulação, quanto para a origem do modelo de predição de rendimento, utilizando coeficientes de cultura da soja gerados com índices de vegetação para classificação dos dados obtidos pelo mapa de rendimento fornecido pela colhedora ao final da safra. O experimento e coleta de dados foram estabelecidos nos meses de novembro de 2017 à abril de 2018. Os resultados dos coeficientes de cultura estimados com NDVI mostraram uma relação consistente com os dados encontrados no simulador SimDualKc para área irrigada e sequeiro, com fatores de correlação superiores 90% e coeficiente de regressão iguais a 1. Tais indicadores representam que tanto a tendência quanto as estimativas estão próximas dos valores simulados. Isso ressalta que a demanda transpirativa da soja, com o coeficiente basal K_{cb} , pode ser avaliada por meio de sensoriamento remoto em apoio ao manejo de irrigação. Estes coeficientes basais foram confrontados com o mapa de colheita e seus valores de rendimento de grãos considerando 22 imagens de satélite. Por meio de abordagens de aprendizado de máquina, encontrou-se valores de acurácia superiores a 90% considerando K_{cb} , pelo classificador de árvore de decisão. Adicionando coeficientes simulados com os coeficientes de sensoriamento remoto, o classificador *Adaboost* se mostrou mais eficiente para predição das oscilações de rendimento, com acurácia de 97,1%. O presente trabalho apresenta resultados que implicam que é possível classificar rendimentos para a cultura da soja, baseado na na transpiração da planta e dados de sensoriamento remoto.

Palavras-chave: Predição de rendimentos. sensoriamento remoto. soja. imagens multi-espectrais. classificação supervisionada. aprendizado de máquina.

ABSTRACT

COMBINING HIGH RESOLUTION MULTISPECTRAL IMAGES AND SOIL WATER BALANCE MODEL TO ESTIMATE SOYBEAN WATER REQUIREMENTS IN SOUTHERN BRAZIL

AUTHOR: DOUGLAS HENRIQUE HAUBERT CERUTTI
ADVISOR: JOÃO CARLOS DAMASCENO LIMA
COADVISOR: MIRTA TERESINHA PETRY

climatic changes in the last years project a future where water resources become more and more scarce. For that reason, it will affect considerably the world food production. A persistent challenge for agriculture is to maintain a high yield production and fit wisely and reasonably the water use with weather conditions in order to get optimal yield results. Currently, remote sensing provides constant crop scout, especially due to water stress correspondence with biophysical characteristics of vegetation during its growth. The combination of vegetation indices with plant features at the field level provides valuable information to irrigation monitoring and management during crop development. In this study, soybeans (*Glycine Max*) characteristics were analyzed, in Southern Brazil, for a field with full water demands supplied and a rainfed field. Field data were collected during all crop season, registering information as crop phenology, the fraction of ground covered, the leaf area index, crop height, and grain yield. These data were used such for the simulation phase, as a source for crop yield prediction, utilizing soybeans crop coefficients derived from vegetation indices for classification of the yield map, provided by the harvester machine at the end of the season. The experiment and data collection were established from November 2017 to April 2018. The results of the actual crop coefficients estimated with NDVi showed a consistent relation with data provided by SimDualKc, for both rainfed and irrigated fields, with correlation factors around 92% and regression coefficient equal to 1. Such numbers demonstrate how much the estimativates trend from simulated validated data. This statistics highlight how much the soybeans transpirative fluxes, with basal crop coefficient K_{cb} , might be evaluated through remote sensing to support irrigation management. These basal crop coefficients were parallel with harvest maps and yield, considering 22 satellite images. With machine learning approach, accuracy values of 90% were found using as feature K_{cb} , based on decision three classifier. Adding simulated data with the remote sensing estimates, Adaboost classifier was the most efficient, with 97,1% of accuracy. This paper show results that implies that it is possible estimate soybeans grain yield based on plant transpiration and remote sensing data.

Keywords: Yield prediction. remote sensing. soybeans. multispectral images. supervised classification. machine learning.

LISTA DE FIGURAS

Figura 1 –	Curva do coeficiente de cultura dividido por estágios de crescimento vegetativos. Autor: (ALLEN et al., 1998)	21
Figura 2 –	Diagrama exemplificando a metodologia elaborada. Fonte: autor.	36
Figura 3 –	Localização das áreas de estudo, situadas em Júlio de Castilhos - Rio Grande do Sul, Brasil. Fonte: autor.	37
Figura 4 –	a) fração de cobertura do solo; b) índice de área foliar; Ambos estimados a partir de folha de papel de tamanho A4 e régua graduada. Fonte: autor.	39
Figura 5 –	Índice de área foliar (IAF) da colheita ao plantio, para área irrigada e não irrigada. Fonte: autor.	41
Figura 6 –	Índice de vegetação por diferença normalizada (NDVI) da colheita ao plantio, para área irrigada (a) e não irrigada (b). Fonte: autor.	45
Figura 7 –	Fração de cobertura do solo (F_c)(%) e altura da planta (m) para as áreas irrigada e não irrigada. Fonte: autor	49
Figura 8 –	Representação geo-estatística dos intervalos de classificação para o mapa de rendimento seco (Ton ha^{-1}), destacando área irrigada e área de sequeiro. Fonte: autor.	52
Figura 9 –	Histograma apresentando a distribuição dos dados de rendimento seco da soja em área irrigada e não-irrigada, com separadores verticais entre as classes. a) Rendimento seco (Ton ha^{-1}) na área irrigada; b) Rendimento seco (Ton ha^{-1}) na área não-irrigada. Fonte: autor.	53
Figura 10 –	Água disponível no solo. Exibe o movimento da água no solo, em contraposição de chuvas e irrigações, para áreas a) irrigada e b) não-irrigada, considerando os dados de umidade medidos pela estação automática de coleta de umidade e os valores de umidade simulados pelo modelo. O período com falha para os valores observados ocorreu devido à falha no armazenamento de dados da estação. Fonte: autor.	58
Figura 11 –	Fração de cobertura do solo estimada, para área irrigada e de sequeiro, utilizando diferentes índices de vegetação. a) Irrigada NDVI; b) Não Irrigada NDVI; c) Irrigada SANDVI; d) Não irrigada SANDVI. Fonte: autor.	60
Figura 12 –	Coeficientes de cultura basais obtidos por sensoriamento remoto (K_{cbVI}), para área irrigada (a) e não irrigada (b). Fonte: autor.	62
Figura 13 –	Coeficientes de cultura basais obtidos por sensoriamento remoto (K_{cbVI}) em comparação aos coeficientes basais simulados pelo modelo SimDualKc, para área irrigada (a) e não irrigada (b). Fonte: autor.	62
Figura 14 –	Coeficientes de cultura obtidos por sensoriamento remoto (K_cVI), para área irrigada (a) e não irrigada (b). Fonte: autor.	63
Figura 15 –	Coeficiente de cultura (K_c) simulados, potenciais e reais (act), considerando o coeficiente de evaporação do solo (K_e) e o coeficiente basal K_{cb} , para área a) irrigada e b) não irrigada. Fonte: autor.	64

LISTA DE TABELAS

Tabela 1 –	Tabela comparativa de trabalhos relacionados, relacionando os modelos de balanço hídrico com sensoriamento remoto	34
Tabela 2 –	Variáveis padrão de evaporação do solo, percolação profunda e escoamento superficial publicados por (ALLEN et al., 1998) calibrados para a região, utilizado em ambas as áreas.....	42
Tabela 3 –	Imagens coletadas com períodos de desenvolvimento para área irrigada e de sequeiro	44
Tabela 4 –	Parâmetros utilizados para o cálculo da fração de cobertura do solo com índice de vegetação	50
Tabela 5 –	Valores em Ton ha ⁻¹ utilizados para classificação do rendimento seco.....	53
Tabela 6 –	Indicadores estatísticos a respeito da relação entre os dados de água disponível no solo (ASW) com os valores simulados pelo SimDualKc após calibração do modelo para as condições locais.	59
Tabela 7 –	Indicadores estatísticos de correlação entre os valores estimados e observados de fração de cobertura do solo.	61
Tabela 8 –	Indicadores para correlação entre os coeficientes de cultura basal simulado com SimDualKc e estimado com índice de vegetação (KcbVI)	62
Tabela 9 –	Indicadores para correlação entre os coeficientes de cultura simulado com SimDualKc e estimado com índice de vegetação (KcVI).	64
Tabela 10 –	Comparação entre classificadores supervisionados com base em derivações do coeficiente de cultura basal K_{cbVI} : coeficiente de cultura ($KcVI$), transpiração da cultura em mm (Tr), evapotranspiração da cultura em mm ($ETcVI$) como alvo a estimativa de rendimento de grãos.....	67

LISTA DE ABREVIATURAS E SIGLAS

AAE	Erro Médio Absoluto
ARE	Erro Relativo Médio
b_0	Coefficiente de regressão
EF	Modelo de eficiência
ET	Evapotranspiração
ETc	Evapotranspiração da cultura
ET_o	Evapotranspiração de referência
ETM	Enhanced Thematic Mapper
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
FAO-56	Irrigação de Drenagem número 56
Fc	Fração de cobertura do solo
Fw	Fração de solo molhada
GEE	Google Earth Engine
IAF	Índice de área foliar
IFRS	Instituto Federal do Rio Grande do Sul
kNN	k-Nearest Neighbor
Kc real	Coefficiente de cultura real
Kc VI	Coefficiente de cultura por índice de vegetação
Kc min	Coefficiente de cultura mínimo
Kcb	Coefficiente de cultura basal
Kcb VI	Coefficiente de cultura basal por índice de vegetação
Kd	Coefficiente de densidade
Kst	Coefficiente de estresse à temperatura
Ksw	Coefficiente de ajuste ao estresse hídrico
Kc	Coefficiente de cultura
Ke	Coefficiente de evaporação
Ks	Coefficiente de estresse
LST	Land surface temperature
METRIC	Mapping EvapoTranspiration at high Resolution with Internalized Calibration
NB	Naive Bayes
NDVI	Índice de vegetação por diferença normalizada
NDVIg	Índice de vegetação por diferença normalizada por verde
NDVIre	Índice de vegetação por diferença normalizada por vermelho de borda

NIR	Infra-vermelho próximo
OLS	Ordinary Least Squares
PBIAS	Porcentagem de viés
PM-FAO56	Modelo de balanço hídrico Penman-Monteith e Periódico número 56 - Food and Agriculture Organization
PPGCC	Programa de Pós-Graduação em Ciência da Computação
PPGEA	Programa de Pós-Graduação em Engenharia Agrícola
RBF	Função de Base Radial
R^2	Coefficiente de determinação
RAW	Água prontamente disponível na zona radicular
REW	Água prontamente evaporável
RED	Vermelho
RGB	Vermelho-Verde-Azul
RMSE	Erro médio quadrático
SANDVI	Combinação entre o índice de vegetação ajustado ao solo e índice de vegetação por diferença ajustada
SAVI	Índice de vegetação ajustado ao solo
SR	Sensoriamento Remoto
SVM	Support Vector Machines
TAW	Água total disponível na zona radicular
TEW	Água totalmente evaporável
Tr	Transpiração da planta
UFSM	Universidade Federal de Santa Maria
VI	Índice de vegetação
VOD	Vegetation Optical Index
Ze	Espessura da camada de solo com a água evaporável

SUMÁRIO

1	INTRODUÇÃO	11
1.1	SENSORIAMENTO REMOTO NA IRRIGAÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	SENSORIAMENTO REMOTO	16
2.1.1	Satélite Sentinel-2	17
2.1.2	Google Earth Engine	17
2.1.3	Sistemas de informação geográficos	18
2.1.4	Sensoriamento remoto de vegetação	18
2.2	IRRIGAÇÃO E DRENAGEM	19
2.2.1	Evapotranspiração de referência	20
2.2.2	Coefficientes de cultura	20
2.2.3	Evapotranspiração real da cultura com sensoriamento remoto	23
2.3	RELAÇÃO ENTRE TRANSPIRAÇÃO DA PLANTA E RENDIMENTO DE GRÃOS	23
2.4	APRENDIZADO DE MÁQUINA	24
2.4.1	Classificação supervisionada	25
2.4.2	Avaliação de modelos de predição	27
3	TRABALHOS RELACIONADOS	29
4	MATERIAIS E MÉTODOS	35
4.1	ÁREA DE ESTUDO	35
4.2	DADOS EM NÍVEL DE CAMPO	37
4.3	DADOS SIMULADOS COM MODELO SIMDUALKC DE BALANÇO HÍDRICO	39
4.4	SENSORIAMENTO REMOTO	42
4.5	PROCESSAMENTO DE ÍNDICES DE VEGETAÇÃO	43
4.6	AVALIAÇÃO DE DESEMPENHO DOS COEFICIENTES POR SENSORIAMENTO REMOTO	46
4.7	COEFICIENTES DE CULTURA COM SENSORIAMENTO REMOTO	48
4.8	APRENDIZADO DE MÁQUINA E CLASSIFICAÇÃO	51
4.8.1	Formação do conjunto de dados	51
4.8.2	Definição de classes e pré-processamento	54
4.8.3	Classificadores e avaliação de performance	55
5	RESULTADOS	57
5.1	CALIBRAÇÃO DO MODELO SIMDUALKC PARA SIMULAÇÃO DO COEFICIENTE DE CULTURA	57
5.2	ESTIMATIVA DA FRAÇÃO DE COBERTURA DO SOLO UTILIZANDO ÍNDICES DE VEGETAÇÃO	59
5.3	ESTIMATIVA DO COEFICIENTE DE CULTURA BASAL COM BASE EM REFLECTÂNCIA DE ÍNDICES DE VEGETAÇÃO (KCB VI)	61
5.4	COMBINAÇÃO DE COEFICIENTES DE CULTURA OBTIDOS POR SENSORIAMENTO REMOTO PARA PREDIÇÃO DE RENDIMENTOS	65
6	CONSIDERAÇÕES FINAIS	68
	REFERÊNCIAS	70

1 INTRODUÇÃO

A agricultura é a atividade que mais utiliza água no mundo, totalizando 68% do consumo mundial (GILBERT, 2012). A irrigação está inserida em 18% de todas as áreas cultivadas sendo responsável por 45% de toda a produção de alimentos no mundo (DÖLL; SIEBERT, 2002). Neste contexto, a água possui papel fundamental para o desenvolvimento da agricultura devido à sua capacidade de reduzir o risco de perdas de cultivo por falta de ocorrência de chuvas viabilizando o uso de regiões com clima desfavorável e, com o correto manejo de irrigação, tem a possibilidade de aumentar a produção de áreas cultivadas (PAREDES et al., 2014).

A soja (*Glycine Max L.*) representou em 2018, 44,3% da área plantada total do Brasil, produzindo mais de 117,8 toneladas de grãos. Deste montante, 16,3% são produzidos no estado do Rio Grande do Sul, cujo rendimento médio é de 3,1 Ton ha⁻¹, 0,8 Ton ha⁻¹ a menos do que o estado com maior produção, ocupando a 12^a posição dentre os estados brasileiros contabilizados (IBGE, 2019). Com isso, assume-se que a soja tem grande expressividade no Brasil pela extensão de área cultivada, porém o Rio Grande do Sul apresenta índices aquém de seus potenciais produtivos.

Um dos fatores que limita o rendimento de grãos é a disponibilidade hídrica, que na região sul é abastecida por precipitações pluviométricas com alguns períodos de estiagem (MATZENAUER; RADIN; ALMEIDA, 2011). Nesse contexto, os sistemas de irrigação, como pivô central, possibilitam que, de forma complementar, a cultura seja abastecida durante todo o ciclo, evitando quedas de rendimento por déficit hídrico.

Desta forma, estratégias para melhorar as técnicas de irrigação são importantes para que as áreas de cultivo possam alcançar rendimentos cada vez mais próximos do potencial produtivo, trazendo maior retorno econômico com a mesma área plantada.

Um dos modelos vastamente utilizados para determinação da demanda hídrica é o sugerido pela Food and Agriculture Organization (FAO) em seu periódico Irrigação de Drenagem número 56 (FAO-56) (ALLEN et al., 1998). Esta proposta utiliza a evapotranspiração de referência, representando as demandas climáticas da atmosfera para uma cultura gramínea como referência (ET_o). Este coeficiente de referência é multiplicado a um coeficiente de cultura (K_c) que representa as diferenças da estrutura da planta na dinâmica de trocas de vapor com a atmosfera e ajusta as características da cultura alvo à cultura gramínea de referência. Para determinação do K_c , o método amplamente difundido é o de regra simples, em que é calculada

uma média entre a transpiração da planta e a evaporação do solo.

Em contraposição ao K_c simples, a abordagem do K_c "dual" constitui-se de um coeficiente basal (K_{cb}) que estima a transpiração primária da planta separadamente de um coeficiente K_e , sobrepondo a evaporação do solo diária, expresso pela Equação $K_c = K_{cb} + K_e$. Este modelo permite otimizar o manejo de irrigação já que proporciona uma forma mais precisa de determinar as necessidades hídricas da planta a partir de sua transpiração. Esta abordagem se mostra eficaz tanto em plantio convencional, como em casos de plantio direto, com resíduos vegetais que tendem a reduzir a evaporação do solo (PEREIRA et al., 2015a). Ainda, a estimativa da transpiração da planta isolada de sua evaporação, possibilita associar a água transpirada pela cultura com a resposta aparente em suas folhas (ALLEN; PEREIRA, 2009). Tendo isso em vista, estratégias para monitoramento da vegetação por imagens de sensoriamento remoto podem ser usadas como forma de obter a exigência hídrica de áreas agrícolas, e, a partir desta, estimar o rendimento de grãos a partir da demanda hídrica da planta.

1.1 SENSORIAMENTO REMOTO NA IRRIGAÇÃO

A caracterização de uma só planta em nível de campo demanda tempo e recursos e em alguns casos pode ser destrutiva. Isso se amplia quando colocado em escala para a avaliação da característica completa do dossel de uma área cultivada (GONZÁLEZ-DUGO et al., 2013). O sensoriamento remoto (SR) pode atingir regiões mais extensas e colher informações com maior frequência e menor custo quando confrontado à coletas presenciais em campo (ZHANG; ANDERSON; WANG, 2015).

O uso do SR na agricultura toma vantagem principalmente do uso da reflexão da radiação solar na vegetação, utilizando índices de vegetação (VIs) para relacionar informações relevantes ao manejo, como umidade do solo, nutrientes, clorofila, entre outras (GAGO et al., 2015). Um bom manejo de irrigação prevê que o momento e quantidades certas de água sejam aplicados, realizando estimativas diárias, a partir de parâmetros do clima, solo e da planta (CARLESSO; PETRY; TROIS, 2009).

Uma das aplicações de sensoriamento remoto na agricultura é o monitoramento via imagens de satélite. Este é calculado em função da resolução espacial das imagens e do período de revisita destes equipamentos na região alvo. Os satélites de domínio público que possuem uma frequência temporal alta (diária) de imagens têm baixa resolução espacial, por volta de 1000 m, o que faz com que o conteúdo de um *pixel*, a menor parte da imagem, não tenha a represen-

tatividade esperada. Somado a isso, há o problema de cobertura das imagens por nuvens, que reduz ainda mais a abrangência das imagens orbitais. Para isso, métodos alternativos para possibilitar a estimativa frequente dos índices de vegetação são fundamentais (ADÃO et al., 2017). O uso de satélites pode inclusive ser complementado por um ou mais satélites para ampliar o número de imagens obtidas durante uma safra agrícola, aumentando a precisão das estimativas de irrigação, principalmente em períodos críticos para a planta.

A correspondência das características da cultura observadas em campo, principalmente no que se refere à cobertura vegetal e dinâmicas da planta define a ligação de diversos índices de vegetação (VI) com os valores de K_c e K_{cb} . Os índices mais utilizados para determinação do K_c são o índice de vegetação por diferença normalizada (NDVI) e o índice de vegetação de solo ajustado (SAVI), que utilizam as faixas infravermelho próximo (800 nm) e vermelho (650 nm) pelas suas altas responsabilidades reflexivas às células de clorofila, proporcionando a avaliação de estresse hídrico sobre a vegetação.

Alguns estudos citam a estimativa do K_{cb} em função da fração coberta do solo (f_c) ou ainda pelo índice de área foliar (IAF) utilizando VIs (GONZÁLEZ-DUGO et al., 2013) (ERRAKI et al., 2007) (LOPEZ-URREA et al., 2009) (PARK; BAIK; CHOI, 2017). Muitos destes, não consideram as alterações no K_{cb} relativas às condições de estresse causadas por deficiências hídricas e por esta razão não compreendem as condições reais de campo. O modelo SimDualKC provê um coeficiente de evaporação (K_e) diário, assim como um coeficiente de estresse (K_s), que garante o ajuste das condições potenciais para condições reais do K_{cb} .

O manejo de irrigação, na grande maioria dos casos, ainda é estimado utilizando medições pontuais dos dados em nível de campo rateados para toda a área. Porém, quando se trata de grandes áreas, a observação de dados em campo exige grande carga de trabalho manual ou ainda de alta distribuição de equipamentos e sensores para, assim, avaliar com precisão as particularidades de cada planta e heterogeneidade da área. O sensoriamento remoto por imagens de satélite limita-se apenas pela disponibilidade de imagens e cobertura de nuvens, possibilitando que ponderações para cada setor da área sejam estimadas. Isso ainda corrobora com os recentes avanços na resolução espacial e temporal dos equipamentos de imageamento, uma vez que, quanto maior for a representatividade de cada pixel em uma imagem e quanto mais frequentes forem as amostras obtidas, melhor será a estimativa dos atributos alvos para a cultura.

O aprendizado de máquina surgiu com as tecnologias de *big data*, a partir da evolução dos recursos computacionais, possibilitando entender grandes massas de dados provenientes

de atividades agrícolas (LIAKOS et al., 2018). Esta abordagem aplicada ao sensoriamento remoto, oportuniza a compreensão de grandes áreas agrícolas de forma não destrutiva, comparando características *in situ* com conhecimentos científicos sobre os atributos das plantas, solo e ambiente (KAMILARIS; PRENAFETA-BOLDÚ, 2018). Técnicas de aprendizado de máquina aliadas à imagens multi-espectrais podem ser usadas para acompanhamento de dinâmicas fisiológicas das culturas em resposta ao ambiente, como a troca de vapor com a atmosfera (WAHABZADA et al., 2016). A aplicação de estratégias de aprendizado de máquina possibilita uma estimativa precisa, ponto a ponto, das variações dos rendimentos, aliando elementos de tecnologia aplicados às características do ambiente (CAMPOS et al., 2016). Um dos fatores essenciais para determinação de rendimentos para as culturas é a sua demanda e disponibilidade hídrica, uma vez que influencia em diversos processos metabólicos do desenvolvimento da planta (MUELLER et al., 2017). Tendo isso em vista, técnicas para estimar a relação da aplicação de água e os rendimentos obtidos pela planta são oportunas, principalmente devido à disponibilidade de grandes volumes dados de sensoriamento remoto oferecidos por satélites de domínio público e a possibilidade de analisá-los com aprendizado de máquina.

Considerando os avanços e limitações citados, principalmente no que diz respeito à transpiração da planta e os rendimentos de grãos em áreas irrigadas e de sequeiro, propõe-se neste trabalho um modelo computacional de aprendizado de máquina com sensoriamento remoto para predição dos rendimentos para a cultura da soja, utilizando imagens satelitais multi-espectrais e medidas simplificadas em nível de campo. Tendo isso em vista, os principais objetivos deste estudo são: (i) Determinar a combinação mais representativa de índices de vegetação para correspondência das características biofísicas da planta (ii) Estimar a transpiração da cultura e ajuste de sua necessidade hídrica por meio dos índices de vegetação avaliados, determinando o coeficiente basal (K_{cb}); (iii) Relacionar a transpiração da planta durante todo o ciclo, através do K_{cb} , com o rendimento final de grãos da soja.

O presente trabalho está estruturado da seguinte maneira: no capítulo 2 é apresentada a fundamentação teórica, abordando os principais trabalhos na área de sensoriamento remoto, irrigação e aprendizado de máquina; o capítulo 3 cataloga os trabalhos relacionados, que de alguma forma tem objetivos comuns com os propostos, possibilitando uma comparação entre as abordagens existentes e a inovação deste trabalho. A seção 4 apresenta os materiais e métodos, que descrevem como cada método foi implementado, como os dados foram coletados para a área de estudo, assim como o seu processamento, remoção de ruídos e definição de constantes

para as condições de campo. Na seção 5 são apresentados e discutidos os resultados obtidos, tanto de validação dos modelos de simulação, variáveis de estimativa, como os coeficientes basais de cultura, fração de cobertura do solo e demais relacionados à transpiração da cultura, como também os classificadores que apresentaram melhor potencial de predição para o conjunto de dados em questão. A seção 5 resume o trabalho de forma a expor e clarificar os resultados encontrados, discussão das soluções encontradas e proposta de trabalhos futuros explorando as fraquezas e forças do modelo proposto.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo conceitualiza as áreas de estudo abordados neste trabalho, como sensoriamento remoto e suas potencialidades para uso na agricultura, mais especificamente, a irrigação de culturas agrícolas. São elencadas as estratégias para relação da demanda e disponibilidade hídrica com o rendimento de grãos, mostrando os principais algoritmos de aprendizado de máquina para, computacionalmente, encontrar as melhores alternativas para integrar as características observadas em campo e via sensoriamento remoto com o rendimento de grãos.

2.1 SENSORIAMENTO REMOTO

Esta seção apresenta as principais vantagens e desvantagens do sensoriamento remoto, fontes de dados utilizadas, ferramentas para obtenção e processamento de imagens assim como uma introdução do seu uso na agricultura.

O sensoriamento remoto (SR) é definido como qualquer método para obtenção de dados de um objeto observado de forma não terrestre. O sensoriamento remoto pode auxiliar de diversas formas, uma vez que pode trazer dados relevantes de uma ampla área sem necessitar, ou necessitar parcialmente, de dados coletados em nível de campo (GAGO et al., 2015).

A vantagem mais significativa dos sistemas de SR se encontra na informação obtida pelos sensores acoplados aos equipamentos, como câmeras, *scanners*, *lasers*, que proporcionam diferentes aplicações, a partir da resposta do espectro eletromagnético captada. As imagens mais comuns obtidas pelas câmeras são as RGB (Red Green Blue), que são formadas a partir de níveis dentro do vermelho, verde e azul para formar as imagens semelhantes ao que o olho humano pode ver. Além do espectro visível, com comprimentos de onda de 400 a 780 nm, são comumente utilizadas as bandas do infravermelho-próximo (NIR), com comprimento de onda de 780 a 1500 nm. Os sensores que captam estas imagens em diferentes faixas do espectro (multi-espectrais) podem ser ativos - utilizando emissão e recepção de radiação e sinais como os equipados com radares ou, ainda, passivos - como é o caso dos satélites que utilizam sensores para a captação da luz solar refletida nos objetos terrestres.

Grande parte dos sensores que obtém imagens de sensoriamento remoto são orbitais, como satélites, e aéreos, como aviões de baixa altitude, balões outros veículos aéreos. Desde 1970, a rede de satélites de domínio aberto vem sendo construída, com diferentes objetivos de

uso do mapeamento orbital, cobrindo vastas áreas do globo (BASTIAANSEN et al., 2001). Esta rede de sensores, possibilita que estudos sejam feitos com relação ao uso da terra, urbano e rural, deslocamento de geleiras, atividades nos oceanos entre outras. Sua aplicação na agricultura vem desde a determinação e mapeamento de fenologia e identificação de culturas, determinação do conteúdo de umidade do solo e saúde das plantas até a identificação de pragas e doenças (GAGO et al., 2015) (LUKAS et al., 2016) (ADÃO et al., 2017).

2.1.1 Satélite Sentinel-2

Dentre as missões mais utilizadas no sensoriamento remoto estão as missões Sentinel 2A e 2B, que foram lançadas em 2015 e 2017, respectivamente. Estas, possuem entre 10, 20 ou 60 m de resolução espacial para banda multi-espectral. A cobertura temporal é de 5 dias com um dos satélites ou, ainda, de 2 a 3 dias intercalando ambos. Cada cena possui 20 x 20 km. Os objetivos desta missão são: monitorar a vegetação; água, solo; detecção de uso de terras; monitoramento de características no clima e suporte para desastres naturais. As bandas utilizadas em todas as imagens serão do espectro eletromagnético visível, com comprimento de onda 400 a 700 nm e infravermelho próximo, com comprimento de onda de 700 a 2500 nm. Ambos os satélites serão empregados de forma complementar utilizando da intercalação das visitas para cobrir a maior frequência temporal possível, de forma a evitar cobertura de nuvens e efeitos de sombreamento destas. Como satélites de domínio público, a missão Sentinel disponibiliza suas imagens em diversas plataformas para uso da comunidade, e dentre estas plataformas está o o Google Earth Engine.

2.1.2 Google Earth Engine

A plataforma *Google Earth Engine* tem por objetivo realizar análise de dados ambientais, sejam eles em terra ou água. Possui um grande conjunto de imagens de satélites, com mais de 40 anos de imagens do globo de diferentes fornecedores, utilizando a infraestrutura do Google Cloud (GOOGLE CLOUD PLATFORM, 2018) para armazenamento e processamento das informações. Em seu ambiente integrado é possível buscar por coleções de imagens, realizando diversos tipos de filtros além da sua capacidade para executar cálculos complexos para todo o globo, a fim de desenvolver produtos como estimativas de biomassa de florestas e classificação de objetos terrestres e, também, áreas urbanas, plantações, corpos d'água, monitoramento de mudanças na vegetação, com o objetivo de quantificar áreas desmatadas ou alterações de bi-

omas (GOOGLE EARTH ENGINE, 2018). Além do ambiente integrado, acessível via *web*, o Google Earth Engine possibilita a configuração do ambiente de desenvolvimento de forma remota, utilizando as linguagens Javascript e Python para comunicação com seus centros de processamento de dados. A ferramenta por si só, já pode ser considerada um Sistema de Informação Geográfico (GIS), mas ela ainda permite integração com diversas outras ferramentas e sistemas, possibilitando a pesquisa e exploração e processamento de imagens externamente.

2.1.3 Sistemas de informação geográficos

Uma dos sistemas de informação geográficos disponíveis e mais completos é o QGIS (QUANTUM GIS GEOGRAPHIC INFORMATION SYSTEM, 2019). Ele é uma aplicação de sistemas de informação geográficos (GIS) com intuito de analisar e processar dados georreferenciados (QUANTUM GIS GEOGRAPHIC INFORMATION SYSTEM, 2019). A plataforma é de código aberto, ou seja, os códigos-fontes estão disponíveis e esta pode receber contribuições de pessoas interessadas em ampliar e colaborar com o projeto. Possui um conjunto completo de ferramentas para criar, editar, analisar e publicar informações geográficas, além de disponibilizar uma base de *plugins* - pequenos programas acoplados ao programa principal - que permitem ao usuário usar e disponibilizar aplicações na própria base do QGIS. Diversos tipos de dados são aceitos, como *shapefiles* e *geojsons*, arquivos geométricos com propriedades definidas e *rasters*, como imagens com seus *pixels* georreferenciados, possibilitando interoperabilidade entre o QGIS e outros sistemas. Por estas características, a aplicação proporciona a análise visualmente de dados georreferenciados que podem ter relação entre si, sua distribuição e frequência distribuídas espacialmente. Há diversos trabalhos considerando o uso de GIS e sensoriamento remoto na agricultura, como na avaliação de incidência de radiação solar no plantio de milho (Zhao et al., 2016), análise de mudanças climáticas em regiões agricultáveis (BRION; BALAHADIA, 2017) e performance de sistemas de irrigação com base em séries históricas para diferentes culturas (YADAV et al., 2017). Portanto, sistemas de informação geográficos tem potencial de exploração de diversos ramos da agricultura, principalmente por meio do sensoriamento remoto.

2.1.4 Sensoriamento remoto de vegetação

Uma das propriedades que pode ser explorada visando observação de áreas agricultáveis são as alterações na vegetação. As tecnologias de sensoriamento remoto são capazes de pro-

ver relações que retratam características biofísicas, estados fisiológicos e trocas de energia em comparação com as imagens em pontos de observação variados (PARK; BAIK; CHOI, 2017) (SAYAGO; OVANDO; BOCCO, 2017). Para determinar a relação destas características com as imagens é imperativo que medições em nível de campo sejam feitas, observando propriedades do desenvolvimento vegetativo da cultura, principalmente em seu quesito aerodinâmico, como índice de área foliar (IAF), fração de cobertura do solo (F_c), altura e biomassa (PÔÇAS et al., 2015).

As relações entre os parâmetros de vegetação e a demanda transpirativa da planta podem ser determinadas a partir de sua fração coberta de solo e/ou da altura da planta (ALLEN; PEREIRA, 2009). A relação entre imagens espectrais e as características do dossel das culturas foram revisadas em diversos trabalhos (PARK; BAIK; CHOI, 2017), (CAMPOS et al., 2016), (PAÇO et al., 2014), (PÔÇAS I, 2016).

Imagens comuns RGB (vermelho, verde e azul) por si só já são úteis para monitoramento do vigor de plantas (LUKAS et al., 2016). A reflectância da vegetação é mais sensível às bandas espectrais que estão a partir do vermelho, como infravermelho-próximo, vermelho de borda e infravermelho térmico (ADÃO et al., 2017). Sua alta sensibilidade às células de clorofila, permitiram que fossem desenvolvidos diversas combinações de bandas espectrais, criando os chamados índices de vegetação.

2.2 IRRIGAÇÃO E DRENAGEM

Esta seção apresenta as principais abordagens de irrigação e sua relação com sensoriamento remoto e o rendimento de grãos. A área de irrigação e drenagem está diretamente ligada com o objetivo deste trabalho ao que se refere à resposta da planta à irrigação e, por consequência, alteração na produção de grãos da soja.

A irrigação é a atividade que mais consome água no mundo, cerca de 68% do total disponível (GILBERT, 2012). Por este motivo, técnicas para determinação dos fluxos de energia que causam as trocas de água entre o sistema planta/solo e a atmosfera são constantemente avaliadas para que o uso da água possa ser otimizado. Isso ocorre de forma que as reposições de água para a planta aconteçam na quantidade e no momento certo (CARLESSO; PETRY; TROIS, 2009).

A irrigação tem objetivo de otimizar o processo de evapotranspiração das culturas, oferecendo à raiz da planta a necessidade de água até o limite que esta possa absorvê-la (BAS-

TIAANSSEN et al., 2001). O modelo PM-FAO56, proposto por ALLEN et al. (1998) é o mais utilizado para fins de experimentação, pois requer uma quantidade reduzida de informações quando comparado a outros modelos e consegue suprir a eficiência exigida para o balanço hídrico das culturas.

2.2.1 Evapotranspiração de referência

O modelo proposto pelo periódico 56 da *Food and Agriculture Organization* (FAO), especificado por (ALLEN et al., 1998) unificou as dezenas de estimativas empíricas que vinham sendo desenvolvidas desde os anos 70 (BORDNE; MCGUINNESS, 1973) (HILLEL; GURON, 1973). Com ele, novos procedimentos de cálculo foram elaborados, onde os procedimentos foram relacionados com outros métodos como os abordados por (BASTIAANSSEN et al., 2001) e (ALLEN; PEREIRA, 2009). A unificação deu-se através da padronização dos parâmetros da planta para uma cultura de referência, gramínea, que então poderia ser adaptada para as demais utilizando coeficientes de cultura.

2.2.2 Coeficientes de cultura

O coeficiente de cultura (K_c) tem por objetivo adaptar as características da cultura de referência para a cultura alvo, no que se refere aos atributos físicos da planta para a cultura de referência. Estes parâmetros levam em consideração a altura da planta, largura das folhas, cobertura do dossel, profundidade da raiz, florescimento, enchimento de grãos, cor, ou seja, qualquer vestígio que demonstre que a planta necessitará de mais água, seja pela maior absorção de radiação quanto pela retenção de água aplicada pela vegetação (ALLEN et al., 1998) (CARLESSO; PETRY; TROIS, 2009). Os valores de K_c em cada um dos estágios de crescimento são ilustrados na Figura 1, demonstrando quais deles expressam maiores valores de ajuste para a evapotranspiração, sendo eles: inicial, desenvolvimento rápido, intermediário e final. O modelo simples de K_c representa uma média entre a evaporação do solo e a transpiração da planta, o que em alguns casos pode ser o suficiente. Porém para afirmar a precisão do balanço hídrico, (ALLEN; PEREIRA, 2009), embasado por (WRIGHT, 1982), propôs uma abordagem em que a transpiração da planta é obtida de forma independente da evaporação. Isso permite que aspectos que interferem em cada uma das dinâmicas sejam avaliados individualmente por um coeficiente K_{cb} tomado como base para a transpiração e um coeficiente de evaporação K_e .

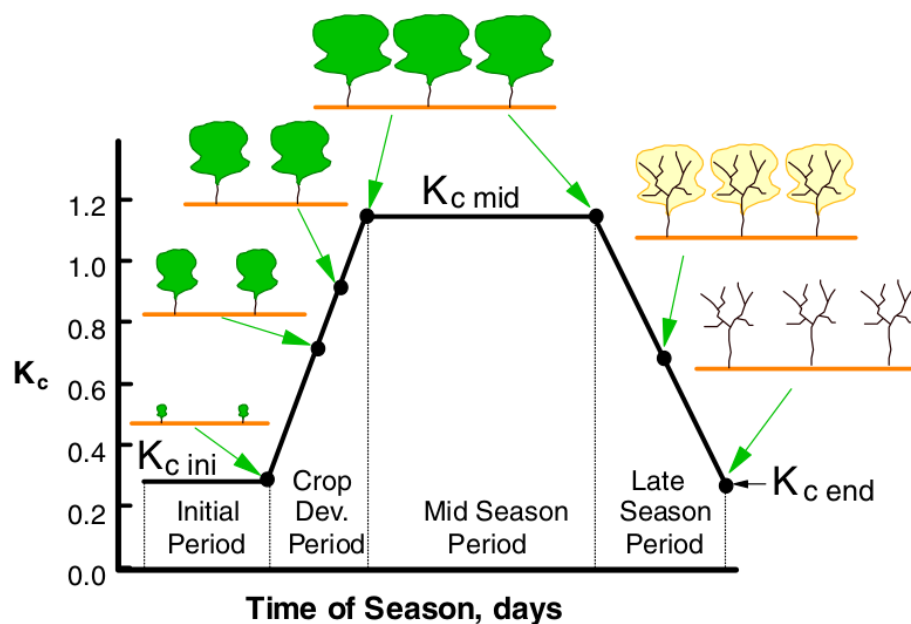
O coeficiente basal é parte da evapotranspiração da cultura, dada pela Equação 2.1,

onde este tem o papel de representar a demanda transpirativa, ou seja, mudanças na estrutura biofísica da planta que interferem no consumo da cultura. O coeficiente de evaporação K_e representa a evaporação do solo, que juntamente com o K_{cb} e K_s compõe a adaptação da planta à evapotranspiração da cultura de referência, que a ET_o descreve.

$$ET_c = ET_o(K_s K_{cb} + K_e) \quad (2.1)$$

Os valores de K_c propostos por (ALLEN et al., 1998) precisam ser ajustados a partir das características regionais como variedades da cultura, tipo de manejo, clima e tipo de solo, para que não possuam erros na estimativa da evapotranspiração da cultura (ET_c) (ODHIAMBO; IRMAK, 2012). Porém, uma calibração nos valores utilizando dados climáticos históricos se mostra mais viável, uma vez que não necessita de exigências a nível de experimentos para tal (ZHANG; ANDERSON; WANG, 2015).

Figura 1 – Curva do coeficiente de cultura dividido por estágios de crescimento vegetativos. Autor: (ALLEN et al., 1998)



Os valores de K_c e K_{cb} foram determinados para diversos tipos de cultura em (ALLEN et al., 1998) e (ALLEN; PEREIRA, 2009). Apesar disso, para melhorar a precisão dos resultados, estes valores devem ser diariamente adaptados usando um coeficiente de estresse (K_s) para adaptar os valores ideais padronizados para condições reais de possível estresse hídrico e, assim, chegar ao K_c real. Este é multiplicado pelo K_{cb} para reduzi-lo às condições em que a planta não recebe água suficientemente para suprir a demanda evapotranspirativa (PÔÇAS

et al., 2015). Assim, o coeficiente de cultura dual real é representado pela Equação 2.2.

$$K_{c_{real}} = K_s K_{cb} + K_e \quad (2.2)$$

Este coeficiente é mais adequado, uma vez que é realizada a calibração destes valores quanto aos ajustes climáticos e, também, à mudanças na geometria, densidade, altura e estrutura do dossel (ALLEN; PEREIRA, 2009).

A implementação dos modelos matemáticos até então idealizados, com objetivo de simular o balanço hídrico no solo, foram explorados por (ROSA et al., 2012), (ROSA et al., 2012) na criação do software SimDualKc. Este alia a evapotranspiração real (ET_{real}) e potencial (ET_c), dividindo o cálculo da evaporação do solo e da transpiração da cultura. Neste modelo, o valor de K_{cb} é calculado pela Equação 2.3, onde K_d é o coeficiente de densidade da cultura que representa os impactos da densidade da planta e sua área foliar e $K_{cb_{full}}$ é o coeficiente basal estimado no estágio máximo de crescimento, tendo cobertura completa do solo ou IAF maior do que 3. $K_{c_{min}}$ é o menor K_c para o solo descoberto, com ausência de vegetação. O coeficiente K_d permite contemplar aspectos de densidade da planta, que podem ser obtidos utilizando a altura média da planta e a cobertura que sua estrutura incide no solo. Este pode ser calculado pela Equação 4.11, proposta por (ALLEN; PEREIRA, 2009):

$$K_{cb} = K_{c_{min}} + K_d(K_{cb_{full}} - K_{c_{min}}) \quad (2.3)$$

$$K_d = \min\left(1, M_L f_{c_{eff}}, f_{c_{eff}}^{\left(\frac{1}{1+h}\right)}\right) \quad (2.4)$$

em que $F_{c_{eff}}$ é a fração de solo coberta ou sombreada pela vegetação próximo ao meio-dia; M_L é um multiplicador em $F_{c_{eff}}$ detalhando o efeito da densidade do dossel no sombreamento para simular os limites físicos impostos ao fluxo de evapotranspiração pela dinâmica da planta e h é a altura média da vegetação (m). Quando há falta de água para a planta o coeficiente de estresse K_s é calculado usando o balanço hídrico diário para a zona radicular ¹. No entanto, os modelos atuais de balanço hídrico e modelos baseados em dados de campo vem sendo substituídos por aqueles que consideram a distribuição espacial utilizando imagens de satélite (MUELLER et al., 2017). Os coeficientes de cultura são a representação da adaptação da planta às condições de referência da irrigação, ou seja, dinâmicas do crescimento, dossel,

¹região do solo em que a raiz da planta abrange, variável durante o desenvolvimento da cultura

raízes e trocas de água com a atmosfera. Logo, com a abordagem do Kc Dual se torna possível obter a transpiração separada da evaporação. A transpiração por sua vez, no presente trabalho, é tomada na hipótese como um dos fatores que tem correspondência com o rendimento de grãos. As evoluções do coeficiente de cultura também serão levadas em consideração, como a evapotranspiração real da cultura.

2.2.3 Evapotranspiração real da cultura com sensoriamento remoto

A evapotranspiração real da cultura, via sensoriamento remoto, vem auxiliando no desenvolvimento de modelos para mapeamento da distribuição espacial e temporal da evapotranspiração de culturas. Estas relações se dão principalmente através dos coeficientes reais da cultura, Kc e Kcb . A reflectância obtidas nas imagens processadas para índices de vegetação auxiliam na relação com atributos biofísicos da planta, como biomassa, índice de área foliar (IAF), fração de solo coberta (Fc) e processos de absorção da radiação solar (PÔÇAS et al., 2015). Muitos estudos abordam a aplicação entre Kc e os VIs, formando os chamados $KcVI$ para diversas culturas, a citar: para algodão, alho, pêssego e vinhas (GONZÁLEZ-DUGO et al., 2013); para oliveiras, cevada (PÔÇAS et al., 2015) e para milho (GEIPEL; LINK; CLAUPEIN, 2014), para vinhas (PÔÇAS et al., 2017) (KARAKIZI; OIKONOMOU; KARANTZALOS, 2015) (ER-RAKI et al., 2013); para trigo de inverno (HUNT et al., 2010) (LI et al., 2014); e para soja (CAMPOS et al., 2018). Estes trabalhos, por sua vez, apesar de realizarem a estimativa do $Kcbreal$ a partir de índices de vegetação, não consideram sua distribuição espacialmente, mas apenas médias. A distribuição espacial permite que se alcance um modelo mais próximo da realidade, com o sensoriamento remoto auxiliando na avaliação diária da necessidade da planta e do balanço hídrico, tanto em condições ideais quanto em condições de estresse, levando em consideração cada particularidade da área.

2.3 RELAÇÃO ENTRE TRANSPIRAÇÃO DA PLANTA E RENDIMENTO DE GRÃOS

Há claras relações entre a produção de biomassa, rendimento de grãos e a transpiração de plantas, variando em seus estádios fenológicos de cultura para cultura, sendo que, algumas são mais sensíveis à deficiências hídricas do que outras em períodos produtivos diferentes (P. STEDUTO T. C. HSIAO; RAES, 2012). Para compensar a falta de disponibilidade de água, a planta

reduz sua abertura estomatal² para reduzir a água consumida pela transpiração. O fechamento dos estômatos causa uma queda na capacidade da planta de equilibrar sua temperatura, desequilibrando funções metabólicas que levam à redução de fotossíntese, e, conseqüentemente, limitações na formação de biomassa (CHAVES et al., 2002). Logo, áreas não irrigadas podem ter restrições durante períodos de ausência de precipitação pluviométrica, com relação à disponibilidade da água para a planta e elevados volumes de evapotranspiração.

As abordagens baseadas na alta correlação entre as características biofísicas da vegetação obtidas por sensoriamento remoto e valores de rendimento são amplamente exploradas por modelos de regressão e classificação (CAMPOS et al., 2018). Estes modelos permitem a assimilação de fatores, como a produção de biomassa e rendimentos, com outros modelos, particionando as variáveis cujas características tem relação direta com fatores do crescimento da planta.

Segundo MUELLER et al. (2017), os fatores determinantes para a relação entre os rendimentos e a água está na duração do *déficit* hídrico e o momento do período vegetativa em que a planta se encontra. O período de maior sensibilidade da soja está nos seus períodos reprodutivos, entre os estádios fenológicos R4 e R5, de rápido desenvolvimento de grãos, em que o estresse hídrico afeta com maior intensidade o rendimento.

2.4 APRENDIZADO DE MÁQUINA

Esta seção conceitua o aprendizado de máquina, os principais modelos de aprendizado para classificação supervisionada e as situações em que cada um deles adapta-se melhor. Além disso, trabalhos utilizando estas técnicas são abordados, para diferentes culturas, regiões e propósitos.

O aprendizado de máquina é um campo científico que toma vantagem, em sua essência, da habilidade que os computadores tem de analisar um grande volume de dados sem necessitar de instruções diretas de programação sobre as entradas de dados para gerar as saídas (LIAKOS et al., 2018). Em sua maioria, estas tecnologias envolvem um processo de aprendizado, assimilado ao aprendizado humano como “experiência”, em que um conjunto de dados de treinamento, ou treino, é fornecido ao modelo para ensiná-lo a desempenhar uma tarefa específica. Este conjunto possui dados observados ou estimados, chamados de características ou variáveis, relacionadas entre si ou não, como formas de exemplo, às quais são relacionadas com um

²Estômato: estrutura da folha onde ocorrem as trocas gasosas entre a planta e atmosfera

atributo alvo. Tais características podem ser numéricas, categóricas, binárias, entre outras. No processo de predição supervisionada, novos dados são observados e colocados como entrada do modelo, chamados de dados de teste, similares aos de exemplo e a saída obtida é o atributo alvo que mais se ajuste às entradas (KAMILARIS; PRENAFETA-BOLDÚ, 2018).

2.4.1 Classificação supervisionada

Os modelos que utilizam treinamento para aprendizado são chamados de supervisionados, em contraposição aos modelos que encontram padrões e relações entre o conjunto de dados inicial, chamados de não-supervisionados. O aprendizado supervisionado permite que seja informado ao estimador uma relação conhecida entre os dados, logo permite maior acerto na predição e menor esforço computacional em comparação aos modelos não-supervisionados, uma vez que os dados são cruzados repetidamente e exaustivamente para relacioná-los (GANDGE; SANDHYA, 2018). O principal objetivo do aprendizado supervisionado é encontrar uma regra geral para mapear as entradas para as saídas.

Ainda, as estratégias de predição, principalmente em modelos supervisionados, se dividem em: 1) classificação, em que a partir das entradas é atribuído um rótulo como saída; e 2) regressão, com objetivo de quantificar as saídas, geralmente quando os alvos são numéricos. Ainda, em conjuntos numéricos, é possível separá-los em categorias pelos seus intervalos, quando sua variação não é significativa em relação aos dados de entrada, e então, permite acompanhar uma taxa de redução ou aumento dos dados numéricos de saída, baseado na relação não-linear implícita no conjunto (LIAKOS et al., 2018).

Os classificadores abordados nesse trabalho tem objetivo de classificar os intervalos de rendimento a partir da característica de transpiração da planta da soja, em uma série temporal de 22 datas durante a safra. Os algoritmos de classificação utilizados são:

- *Support Vector Machines* (SVM) é um classificador que consiste na classificação binária, criando planos para separação dos dados a serem classificados, a partir do conjunto de treino. É dinâmico quanto à incorporação de núcleos que, dependendo da natureza dos dados, pode melhorar a classificação em cada caso. O núcleo mais utilizado com SVM é o de função de base radial (RBF) e o linear, em que as separações entre os classificadores são tratados diferentemente. Para SVM com núcleo linear, a classificação divide o plano considerando os dados mais distantes do grupo classificado. Já o núcleo RBF é mais tolerante quanto à distância entre os pontos e torna o modelo menos sensível ao conjunto

de dados. É indicado principalmente onde há problemas de *overfitting*, isto é, quando um modelo se ajusta muito bem ao conjunto de treinamento, mas não a novos conjuntos de dados (LIAKOS et al., 2018). Este modelo é usado principalmente na agricultura de precisão, tratando problemas em mapas de fertilidade (CHLINGARYAN; SUKKARIEH; WHELAN, 2018), para identificação de diferentes culturas e também de plantas daninhas (ZHANG; GONG; WANG, 2018).

- O algoritmo *K-Nearest Neighbor* (kNN), está no grupo dos modelos baseados em instância, cujos exemplos novos são comparados com os atributos dentro da base de treinamento, diversas vezes, para encontrar agrupamentos. O k-NN busca as instâncias mais próximas do exemplo, geralmente a partir da distância euclidiana, para classificar o objeto na classe que tem mais vizinhos próximos. A letra k se refere ao número de elementos vizinhos mínimos necessários para a classificação (GONZALEZ-SANCHEZ; FRAUSTO-SOLIS; OJEDA-BUSTAMANTE, 2014). Aplicações com $k = 1$ geram um modelo muito sensível aos dados, porém, o aumento dos valores de k levam à diminuição da variância e aumento do viés (*bias*)³.
- Árvores de decisão são modelos baseados na arquitetura de árvores, onde o conjunto de dados é organizado em grupos homogêneos menores de dados ramificados, com diversos níveis. Cada nó desta árvore representa uma diferença entre os atributos do conjunto, e cada galho desta árvore leva a um rótulo diferente na classificação, chamado de folha, posicionado ao final de cada um destes galhos.
- O método de *Random Forests* é uma forma de combinar diversas árvores de decisão, formando uma floresta de árvores menores em cada nó. Para geração de cada nó são usadas características aleatórias, diferentemente das árvores de decisão isoladas, que buscam os melhores atributos para tal tarefa.
- O modelo *Gaussian Process* proporciona a consideração de ruído nos dados observados e intervalo de confiança entre as predições, uma vez que é baseado em probabilidades. O *Gaussian Process* é usado principalmente para integrar informações espaço-temporais, uma vez que cada exemplo de teste é mapeado para uma probabilidade de pertencer a uma classe (CHLINGARYAN; SUKKARIEH; WHELAN, 2018). É usado principalmente para encontrar similaridades entre imagens de satélite, como identificação de pixels

³bias: valores destoando de 0 indicam sub ou super estimativas entre os dados observados e estimados

que pertencem à um grupo. O modelo é utilizado para estimativa de evapotranspiração diária (HOLMAN et al., 2014) a partir de parâmetros meteorológicos e estimativa de rendimento das culturas por sensoriamento remoto (YOU et al., 2017).

- O modelo *Naive Bayes* (NB) usa fórmulas estatísticas e cálculos de probabilidades para a classificação. É indicado quando existem relações conhecidas sobre os dados para aumentar ou diminuir as probabilidades das classes. O teorema de Bayes é aplicado, conforme a Equação 2.5, em que cada iteração de $a_1, a_2 \dots a_n$ correspondo à cada objeto a ser classificado e P é a probabilidade deste pertencer a uma classe.

$$P(\text{classe}|a_1 \dots a_n) = \frac{P(a_1 \dots a_n | \text{classe}) \times P(\text{classe})}{P(a_1 \dots a_n)} \quad (2.5)$$

O trabalho abordado por PAUL; VISHWAKARMA; VERMA (2016) aplica NB para estimativas de rendimento de culturas a partir de características observadas do solo, em apoio do modelo kNN.

- O *Adaboost* é um algoritmo que combina outros modelos de aprendizado, considerados "fracos", como árvores de decisão com poucos níveis. A cada iteração um algoritmo fraco é usado e são atribuídos pesos às características. O número de estimadores é definido e determina quantas repetições serão executadas até a formulação da saída. Este processo de atribuição e modificação de pesos é chamado de intensificação (*boost*) (FREUND; SCHAPIRE, 1997). Em MISHRA; MISHRA; SANTRA (2017), o *Adaboost* é usado para regressão, para estimativa da produção de culturas usando diferentes modelos "fracos", em comparação com modelos "fortes", tendo como base as variações climáticas. No estudo, o algoritmo trouxe bons resultados quando comparado aos modelos considerados "fortes", principalmente quando a variabilidade dos dados é ampla.

2.4.2 Avaliação de modelos de predição

A avaliação de desempenho de um modelo para executar uma determinada tarefa é medida por uma métrica de performance, melhorada com experiência durante o tempo. Existe vários modelos estatísticos e matemáticos que podem ser usados. KAMILARIS; PRENAFETA-BOLDÚ (2018) cita que em sua revisão, 19 de 24 trabalhos utilizam a acurácia da classificação como métrica, em que valores superiores a 90% indicam boa performance. As métricas de performance geralmente são avaliadas sobre quatro objetos principais, quando se considera

classificação binária: VP) verdadeiros positivos, em que a predição consegue classificar um valor corretamente; FP) falso positivo, em que a predição consegue classificar um valor, mas incorretamente; VN) verdadeiro negativo, em que a predição não consegue encontrar classificação, corretamente; e por fim: FN) em que a predição não consegue encontrar uma classe que se ajuste, mas há uma.

A acurácia pode ser definida como a porcentagem de predições que foram feitas corretamente (VP e VN) sobre a soma das predições positivas e negativas para valores verdadeiros e falsos, ou seja, todas as predições realizadas, conforme a Equação 2.6:

$$Acuracia = \frac{VP + VN}{VP + FP + VN + FN} \quad (2.6)$$

A precisão do modelo é tida como a razão sobre as predições positivas, conforme a Equação 2.7. Esta demonstra a relação entre os valores que o modelo estimou como corretos, tanto os positivos quanto negativos.

$$Precisao = \frac{VP}{VP + FP} \quad (2.7)$$

Para o caso de classificação multi-classe, ou seja, o alvo não é binário, como neste trabalho, cada resultado da classificação interfere nos verdadeiros e falsos positivos e negativos dos demais, formando uma matriz. Como forma de exemplo, em uma classificação com rótulos de 1 a 5, as classificações corretas positivas do rótulo 1, contabilizam como predições corretadas negativas dos demais rótulos, e assim sucessivamente até que finalizem todas as predições. Tanto a acurácia como a precisão são reduzidas por meio de média aritmética para a avaliação final do modelo.

Acurácia e precisão são dois critérios importantes para avaliação de um classificador. Neste trabalho, estes foram utilizados para determinar quais são os melhores classificadores para a cultura da soja, para áreas irrigadas e não irrigadas, considerando o coeficiente de cultura basal por sensoriamento remoto para classificar um intervalo de produção de grãos. Para tanto, cada pixel de uma imagem de satélite é confrontada com uma coordenada geográfica fornecida pela colhedora contendo o o valor do rendimento de grãos colhido ao final da safra.

3 TRABALHOS RELACIONADOS

Neste capítulo são abordados os principais trabalhos com problemas de pesquisa e objetivos semelhantes ao presente, como forma de explorar potencialidades e deficiências destes e trazer inovação na área de sensoriamento remoto para a agricultura.

O trabalho proposto por PAÇO et al. (2014), relaciona características da planta utilizando o coeficiente de densidade Kd e um modelo de mapeamento de evapotranspiração, intitulado METRIC. O modelo considera solo e planta como entradas para gerar a evapotranspiração (ET), a partir de imagens de satélite, *pixel a pixel*, utilizando sensores de micro-ondas e térmicos, para registro das trocas de energia na superfície, subtraindo o fluxo de calor do solo da radiação efetiva na superfície. Para tanto, foram utilizadas imagens dos satélites da missão Landsat5 TM e 7 ETM+, nos anos agrícolas⁴ de 2011 e 2012. Foram comparadas duas abordagens, uma realizando estimativas com dados em nível de campo e outra, usando sensoriamento remoto, O modelo SimDualKc foi calibrado, para ajuste das condições locais à oliveiras, mostrando que os parâmetros são adequados para a comparação da evapotranspiração real observada, com a evapotranspiração simulada pelo modelo, com coeficiente de determinação $R^2 = 0,76$ e $RMSE = 0,44 \text{ mm } d^{-1}$ em 2011 e $R^2 = 0,65$; $RMSE = 0,35 \text{ mm } d^{-1}$ no ano de 2012. Levando em consideração a relevância do modelo, foi realizada a avaliação de performance do modelo Mapping EvapoTranspiration at high Resolution with Internalized Calibration (METRIC) de evapotranspiração com os valores estimados pelo SimDualKc, exibindo padrões similares quando comparados. Para os coeficientes de cultura foi encontrada forte correlação, com coeficiente de determinação R^2 DE 0,86 e coeficiente de regressão de 0,99. Ao final, os autores discutem sobre a forma complementar como os modelos usados interagem, com suas diferentes formas de determinar a evapotranspiração e, ainda, a possibilidade da estimativa usando a fração de cobertura do solo e a altura da planta como parâmetros no processamento da transpiração da planta, formando assim, um modelo eficiente para oliveiras.

Três culturas são avaliadas por PÔÇAS et al. (2015) para estimativa de coeficientes de cultura a partir de sensoriamento remoto. O trabalho propõe o uso de imagens de satélite, neste caso das missões Landsat 5 TM e Landsat 7 ETM+, combinado ao modelo SimDualKc, para estimativa da evapotranspiração. O modelo SimDualKc, similarmente ao desenvolvido por PAÇO et al. (2014), é usado para avaliar a performance do modelo, que tem como entradas a fração

⁴Ano agrícola: período onde são considerados os períodos de plantio e não somente os meses de Janeiro a Dezembro.

de cobertura do solo (F_c), observada e estimada a partir dos índices de vegetação $SAVI$ e $NDVI$. Os autores tem por objetivo adaptar o coeficiente por sensoriamento remoto, levando em consideração não apenas características potenciais, como áreas hidricamente bem abastecidas, como também avaliar os coeficientes de culturas reais, ou seja, ajustados às condições de possíveis *déficits* hídricos. Para isso, o trabalho sugere a combinação do coeficiente que representa a evaporação do solo (K_e) somado ao coeficiente de cultura basal K_{cbVI} , que representa a transpiração estimada por sensoriamento remoto, e multiplicado pelo coeficiente de estresse (K_s).

Os resultados encontrados demonstraram boa eficiência ($R^2 = 0,72$) para a avaliação do K_{cbVI} estimado e o K_{cb} simulado para as três culturas levadas em consideração, utilizando a fração de cobertura do solo a partir de índices de vegetação (F_cVI) em complemento da altura da planta, tanto para culturas perenes e densas, como as oliveiras, quanto para culturas menores e sazonais, como o milho e cevada. Estas condições foram satisfatórias tanto em ambientes com demanda hídrica suprida quanto para aqueles que apresentaram falta de água para a cultura.

MUELLER et al. (2017) disserta em seu trabalho sobre métodos para estimativa de rendimento com sensoriamento remoto, utilizando imagens de satélite em comparação com assimilação de dados e dados coletados em campo, com objetivo de acompanhar a variabilidade do rendimento da soja e milho. O uso de índices de vegetação, como $NDVI$ e Enhanced Vegetation Index (EVI) são contextualizados e desconsiderados pelos autores, devido à sua saturação em períodos com densidade elevada de vegetação e, por serem dependentes da geometria das plantas, podem não corresponder ao estresse da cultura. Em contraposição, o índice *Vegetation Optical Depth* ou Profundidade óptica da vegetação (VOD) é usado no trabalho pois é mais sensível ao conteúdo de água na superfície da vegetação. Duas variáveis são consideradas para a relação com o rendimento: evapotranspiração e umidade do solo. A evapotranspiração pode ser obtida pelo modelo ALEXI, que usa *Land Surface Temperature* (LST), para estimar a temperatura do dossel e estresse da vegetação, através de sensores térmicos. A estimativa de umidade é baseada em modelos físico-hidrológicos: Community Land Model (CLM), Catchment Land Surface Model (CLSM), NOAA, e modelo Palmer em duas camadas (PM). Estes tem diferentes parâmetros como entrada para determinar o deslocamento vertical da água no perfil do solo. O conjunto de dados usado para o rendimento foi disponibilizado pelo portal NASS, que coleta informações através de pesquisas com produtores rurais. O modelo associativo entre o rendimento e os conjuntos de dados foi a análise por ranqueamento de validação cruzada.

A avaliação foi feita no período de um ano, considerando as variáveis auxiliares em reduções mensais, com média de rendimento com distribuição espacial para os limites geográficos entre municípios, da região central e leste dos Estados Unidos. Comparações entre VOD, NDVI e EVI foram realizadas, mostrando baixa sensibilidade para VOD ao rendimento. No entanto, incorporando mapas de evapotranspiração e umidade do solo derivados de sensoriamento remoto, as estimativas obtiveram melhora considerável com os parâmetros de entrada mesclados.

Como resultados, os autores encontraram boas relações entre a evapotranspiração e umidade do solo, utilizando o índice VOD para previsão de rendimento para ambas as culturas, auxiliados da ET e umidade do solo, considerando melhores correlações do que as encontradas com índices de vegetação NDVI e EVI, devido à características captadas que não são apenas as partes que refletem a luz verde na planta, como galhos e caules. Entretanto, a baixa resolução espacial (1 km px^{-1} para imagens e $0,25 \text{ km px}^{-1}$ para atributos auxiliares) prejudicou certos aspectos, já que não é possível identificar pixels que representem apenas informações da cultura e a variabilidade espacial ficou afetada por este quesito.

O estudo feito por SCHWALBERT et al. (2018) tem por objetivo relacionar os rendimentos do milho com índices de vegetação, como índice de vegetação por diferença normalizada (NDVI), NDVI Verde (NDVIg), NDVI Vermelho de Borda (NDVIre), levando em consideração 16 campos em diferentes regiões, no Brasil e Estados Unidos, nos anos de 2016 e 2017. A metodologia utilizada no trabalho, compreende quatro etapas: aquisição e preparação de dados; seleção de treinamento e verificação; criação de modelo de previsão; e validação temporal e espacial. O período de coleta de imagens de satélite foi de 20 dias antes e depois do florescimento, dependendo da disponibilidade de imagens devido à interferência de nuvens. Para a seleção do treinamento de dados, as áreas do Brasil foram selecionadas como treinamento e as áreas americanas como validação. Para remoção de *outliers*, uma das técnicas usadas foi a de intervalo interquartil (IQR), usando o intervalo entre primeiro e terceiro quartis para selecionar os pontos que serão selecionados para o treinamento. O modelo de previsão foi construído inicialmente, levando em consideração o coeficiente de *Moran*, que mede estatisticamente a força entre a autocorrelação espacial, supondo o cruzamento entre os pontos com seus vizinhos em relação ao desvio ponderado da média. Foi realizada uma regressão linear, utilizando o método *Ordinary Least Squares* (OLS) e os três índices de vegetação do estudo.

O modelo de regressão espacial superou o modelo OLS para predição de rendimento e o índice que apresentou o melhor indicador para o modelo foi o NDVIre. Este índice se saiu

melhor visto que não sofre influência significativa para mudanças na área foliar quando comparado aos demais índices, o que faz com que este não fique saturado quando encontra valores de índice de área foliar altos. Os índices NDVI_g e NDVI, ainda assim, mostraram ser capazes de relacionar-se com a variação nos rendimentos, porém com menor potencial que o NDVI_{re}. Além disso, o trabalho mostrou que há variações na aplicabilidade de modelos de predição de rendimento quanto à sua distância espacial e temporal, porém, um fator determinante para a aplicabilidade é a distribuição dos valores de treinamento, como IQR, que direciona a resultados melhores. Por fim, conclui-se que estratégias usando imagens de alta resolução de satélites, para o período que compreende o período de desenvolvimento e intermediário da cultura do milho, são possíveis de serem utilizadas para predição de rendimentos.

Em CAMPOS et al. (2018), os autores exploram a relação entre a produção de biomassa e os coeficientes de cultura com índices de vegetação, tomando por base o modelo FAO-66 de água-produção. O índice utilizado é o Soil Adjusted Vegetation Index (SAVI), que segundo os autores, não encontrou valores de saturação em nenhum estágio de desenvolvimento da soja e milho, culturas alvo do estudo. Os autores utilizam o coeficiente de cultura basal K_{cb} , que representa isoladamente as alterações da dinâmica da planta no seu consumo de água, adaptado às condições reais com um coeficiente de ajuste ao estresse hídrico K_{sw} , e estresse à temperatura K_{st} . O K_{cb} é estimado com base no SAVI. Para tanto, estes valores foram estimados para as duas culturas do estudo, em áreas de sequeiro e irrigadas, durante o período de 11 anos, em Nebraska - EUA.

Os autores acreditam que a melhor relação entre o rendimento final, na colheita, leva em consideração tanto a biomassa quanto o rendimento, calculando a proporção do rendimento final sobre a biomassa total, para formar o coeficiente *Harvest Index* (HI). O efeito da relação $K_{cb}VI$ -biomassa foi mais evidente nas áreas de sequeiro, já que neste caso o aumento dos valores acumulados de K_{cb} reduzem a biomassa produzida. Algumas inconsistências foram encontradas no início do período quando os valores acumulados de K_{cb} não chegam a 20 e, posteriormente, nos estágios reprodutivos à maturação fisiológica. Neste momento, os valores residuais⁵ se mostraram mais elevados, revelando certa instabilidade ao que se refere ao período intermediário-final da cultura, onde inicia-se o amarelamento das folhas. Os autores ressaltam ainda que existem outros efeitos que não apenas efeito água-produção, que se demonstram quase que insignificantes com base diária, mas seu acúmulo pode reduzir as correlações com o

⁵valor residual: diferença entre os valores observados e estimados

passar do tempo. As estimativas usando HI expressaram períodos que a água mais afeta a soja e milho, a considerar da fenologia Vt à R5.

Analisando os trabalhos relacionados na Tabela 1, pode-se perceber que os índices de vegetação (VIs) - NDVI e SAVI - são aplicados em diversos estudos e mostram evidentes relações com características biofísicas das plantas. Alguns VIs foram sugeridos pelos autores como alternativa às fraquezas de ambos os índices, porém estes citados ainda são os mais populares por não necessitarem de sensores mais complexos, como de micro-onda ou térmicos. Os trabalhos tem seguido a tendência dos lançamentos de novos satélites, com melhores resoluções espectrais e maiores tempos de revisita, que tornam o monitoramento de áreas agrícolas uma tarefa progressivamente viável e precisa, a citar os avanços com o satélite Sentinel-2.

O interesse dos trabalhos relacionados em avaliar condições de estresse da planta, quanto à disponibilidade hídrica ou efeitos da temperatura, tem sido uma crescente e permite obter avaliações mais precisas do que apenas utilizando situações ideais, aproximando o tema de estudo à realidade do produtor rural. Para isso, a grande maioria dos trabalhos avalia tanto áreas irrigadas quanto áreas de sequeiro, visto que a demanda de água e possíveis condições de estresse pela planta podem ser ponderadas por sensoriamento remoto.

A relação entre a transpiração da planta e o rendimento de grãos e produção de biomassa já é explorada na última década, principalmente com a consolidação do modelo FAO-66 (P. STEDUTO T. C. HSIAO; RAES, 2012) de água-produção. Todavia, correlações entre a transpiração da cultura utilizando sensoriamento remoto e predição de rendimento ainda são considerados assuntos recentes. Os avanços atuais com o aprendizado de máquina levam à possibilidade de explorar algoritmos de classificação com as imagens de satélite e explorar a relação entre os coeficientes de cultura basais (Kcb) por sensoriamento remoto com fatores econômicos importantes para a agricultura, como o rendimento de grãos. Considerando os trabalhos relacionados, este trabalho está contextualizado quanto ao uso de índices de vegetação e traz novas descobertas de informação não-explícitas, na aplicação de classificação por aprendizado de máquina. O objetivo de prever a resposta produtiva da planta ao seu consumo de água é assunto em destaque na comunidade acadêmica, diferindo neste caso na metodologia para alcançar tal predição quando coopera entre as áreas de sensoriamento remoto e aprendizado de máquina.

Tabela 1 – Tabela comparativa de trabalhos relacionados, relacionando os modelos de balanço hídrico com sensoriamento remoto

Artigo	Índices de vegetação usados	Modelo de balanço hídrico	Imageamento	Cultura	Entradas
(PAÇO et al., 2014)	NDVI	SimDualKc e METRIC	Landsat 5 TM, Landsat 7 ETM+	Oliveira	Fração de cobertura do solo e altura da planta
(PÔÇAS et al., 2015)	NDVI, SAVI	SimDualKc e KcbVI	Landsat 5 TM, Landsat 7 ETM+	Cevada, Milho e Oliveiras	Fração de cobertura do solo e altura da planta
(MUELLER et al., 2017)	NDVI, EVI, VOD	Umidade do solo: Community Land Model (CLM), Catchment Land Surface Model (CLSM), NOAH, e modelo Palmer em duas camadas (PM). Evapotranspiração: atmosphere-land exchange inverse (ALEXI)	MODIS MOD13C2	Soja e Milho	ET, umidade do solo
(SCHWALBERT et al., 2018)	Green NDVI, NDBI, NDVI, Red-edge NDVI	-	Sentinel-2	Milho	Índices de vegetação
(CAMPOS et al., 2018)	SAVI	PM-FAO, Kc Dual	Landsat 5 TM, Landsat 7 ETM+	Soja e Milho	<i>Harvest Index</i> (HI), <i>Water Use Efficiency</i> WUE*, Kcb

4 MATERIAIS E MÉTODOS

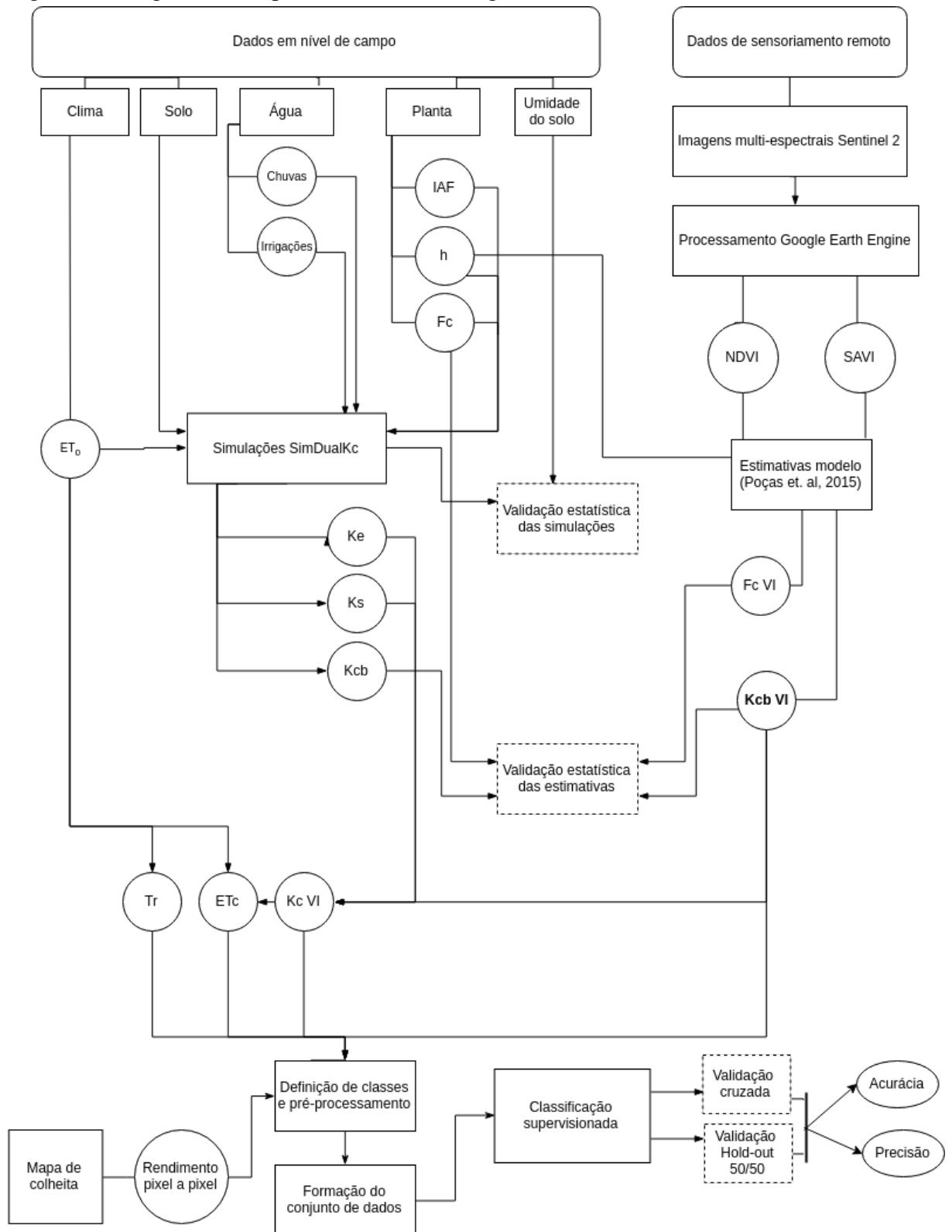
Este capítulo discute as abordagens usadas para a pesquisa, no que se refere ao experimento prático, como área de estudo, características locais, equipamentos e métodos de observação das características da planta, solo e clima. A Figura 2 retrata os dados de entrada para os modelos, o fluxo de informação desde a obtenção dos dados, as saídas simuladas e estimadas de cada modelo e suas respectivas validações. A metodologia do trabalho visa aliar o sensoriamento remoto, dados em nível de campo e modelos computacionais de aprendizado de máquina para predição dos rendimentos da soja.

4.1 ÁREA DE ESTUDO

Neste experimento foi utilizada a cultura da soja (*Glycine max*), plantada em área rural, localizada no município de Júlio de Castilhos, na região central do Rio Grande do Sul - Brasil, com latitude 29 14' 7,639"S; longitude 53 42' 49,524"O; e altitude 529 m. O solo caracteriza-se como Argissolo Vermelho-Amarelo Alumínico úmbrico (EMBRAPA, 2006). O clima da região é do tipo "Cfa" de acordo com a classificação climática de Köppen (subtropical úmido, sem estação seca definida e com verões quentes). As chuvas são bem distribuídas durante todas as estações do ano e o total anual médio é de 1678 mm. Possui evapotranspiração potencial de 869 mm ano^{-1} , com média de 105 mm durante o inverno e 351 mm durante o período de verão (MATZENAUER; RADIN; ALMEIDA, 2011). Foram analisadas duas áreas com solo homogêneo ao que se refere à textura e fertilidade. Uma das áreas, irrigada, identificada em verde (cor mais escura) na Figura 3, com cerca de 30 ha de extensão, abastecida por pivô central e a segunda área de sequeiro, em cor avermelhada (em cor mais clara), situada na periferia da primeira e outro sistema de irrigação por pivô central próximo, com aproximadamente 10 ha. Estas áreas são irregulares em seu formato, pois contém zona de preservação ambiental contidas na área que o sistema de irrigação abrange.

O material comercial utilizado foi Agroeste 3590 IPRO RR2, semeado sobre a palhada do trigo em plantio direto, com 5 toneladas de matéria orgânica por hectare. O espaçamento entre linhas foi de 0,5 metros e população de 311.000 plantas por hectare.

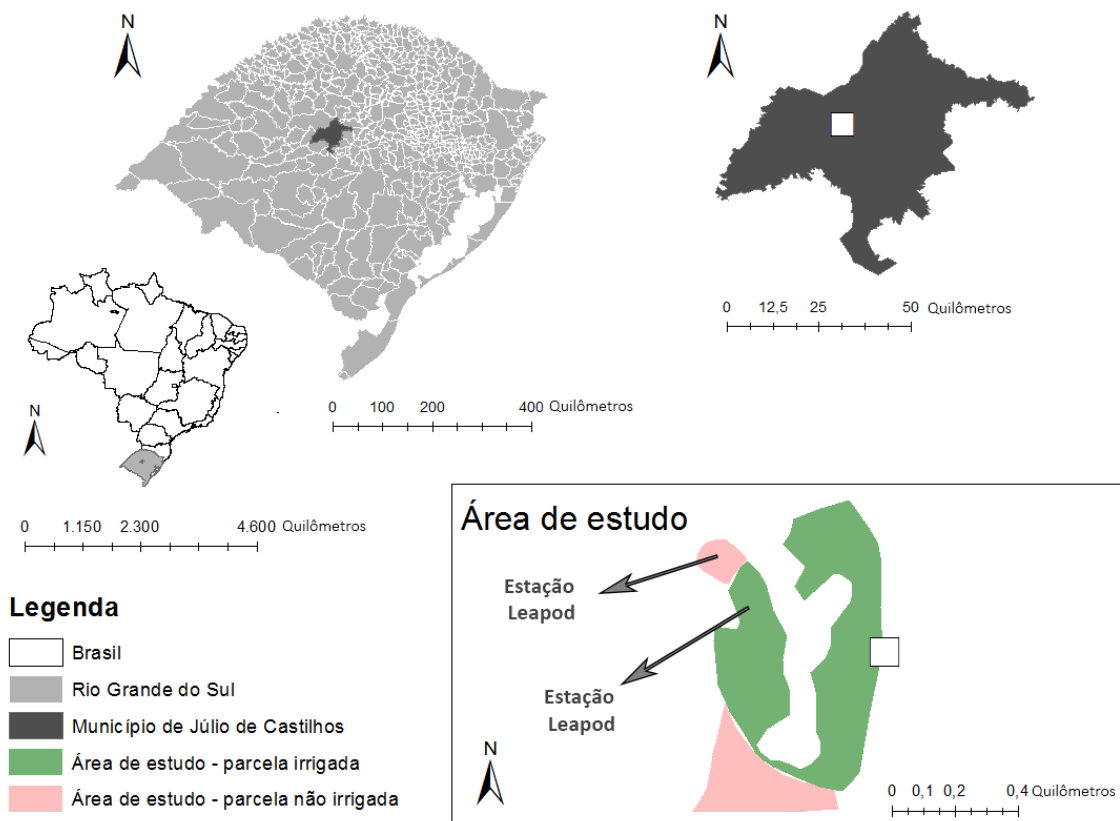
Figura 2 – Diagrama exemplificando a metodologia elaborada. Fonte: autor.



4.2 DADOS EM NÍVEL DE CAMPO

As informações climáticas foram obtidas utilizando uma estação meteorológica automática (modelo Vantage Pro 2, Davis Instruments, EB, USA), instalada na propriedade. O calendário de recomendações de irrigação será fornecido pelo Sistema Irriga (SISTEMA IRRIGA, 2019), serviço de manejo e monitoramento de irrigação há 20 anos no setor. A plataforma recomenda irrigações com base no modelo PM-FAO56 de balanço hídrico, utilizando as características do solo, da cultura e dados meteorológicos para determinar a evapotranspiração da cultura. A estimativa do momento ideal para a irrigação se dá em função dos limites de capacidade de campo e ponto de murcha permanente, mantendo a umidade do solo em até 70% da fração de água disponível para a planta.

Figura 3 – Localização das áreas de estudo, situadas em Júlio de Castilhos - Rio Grande do Sul, Brasil. Fonte: autor.



O conteúdo de água do solo e entradas de água foi monitorado em ambas as áreas por duas estações de campo automáticas (modelo Leapod 2.0, Sistema Irriga, RS, BRA), utilizando

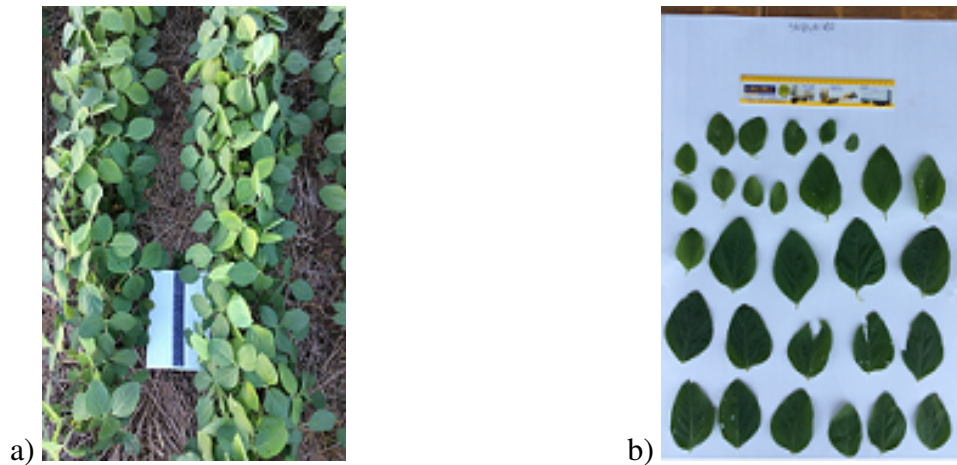
sensores *Watermark* (modelo 200SS, Irrrometer instruments, Riverside, CA, USA), que concede a média das medições a cada 15 minutos do potencial matricial do solo ($cbar$) e um pluviômetro automático acoplado à estação, instalado acima do dossel da cultura, fornecendo a medida em milímetros, com resolução de 0,2 mm. Os dados foram transmitidos através da rede General Packet Radio Services (GPRS), comumente utilizada por aparelhos celulares. Duas estações distribuídas em cada uma das áreas do experimento com três sensores instalados nas camadas de 0,1m, 0,3m e 0,5m de profundidade fizeram a análise de umidade em apoio ao balanço hídrico. O objetivo destas estações foi: 1) avaliar a disponibilidade hídrica para a área irrigada em comparação com os dados simulados pelo SimDualKc e observados; 2) identificação de períodos de estresse hídrico pela área não irrigada; e 3) calibração e validação do conteúdo de água no solo com dados observados. Para conversão dos valores de potencial ($cbar$) para porcentagem (%) foi utilizado o método de Van Genuchten (GENUCHTEN, 1980).

Foram observadas em campo, com apoio do Programa de Pós-Graduação em Engenharia Agrícola da Universidade Federal de Santa Maria (PPGEA/UFSM), características da cultura da soja, como altura da planta, índice de área foliar, fração de cobertura do solo, principais estádios fenológicos, biomassa e rendimento dos grãos. Os estágios de crescimento da cultura que foram utilizados são definidos por (ALLEN et al., 1998), a citar: 1) estágio inicial: da semente até 10% do solo coberto; 2) estágio de desenvolvimento ou crescimento rápido: de 10% do solo coberto até a cobertura do dossel máxima; 3) intermediária: da cobertura total até a maturação; e 4) final ou de senescência: acontece do amarelamento e queda das folhas até a colheita. A estimativa para os estádios de crescimento da soja, a partir da semente no dia 13 de novembro de 2017, foi de início do crescimento rápido em 27 dias, período intermediário em 46 dias, início da senescência em 120 dias e colheita em 145 dias.

A determinação da fração de cobertura vegetal da superfície do solo (F_c) foi realizada através do método da contagem dos pontos quadrados (MANNERING; MEYER, 1963), utilizando folha de papel tamanho A4 (210 x 297mm), régua graduada em contato do solo e uma fotografia é obtida acima do dossel. Foi utilizado o software ImageJ (IMAGEJ, 2019) para medida da área da folha de papel coberta pela cultura, tomando como escala a régua graduada, conforme a Figura 4-a. O mesmo software foi utilizado para contabilizar o índice de área foliar, contabilizando a média da área foliar de todas as folhas da cultura coletadas, conforme a Figura 4-b.

Foram coletadas amostras, de forma destrutiva, das plantas em cada lance do pivô cen-

Figura 4 – a) fração de cobertura do solo; b) índice de área foliar foliar; Ambos estimados a partir de folha de papel de tamanho A4 e régua graduada. Fonte: autor.



tral nos principais estádios fenológicos, com objetivo de realizar análises dos componentes de rendimento (número de plantas por área, número de vagens por planta, número de grãos por vagem e o peso de grão). Plantas foram extraídas tanto da área irrigada quanto da área de sequeiro para a definição da área foliar e biomassa. Esta foi obtida em laboratório utilizando um medidor de área foliar LI3100C (LI-COR Inc.), em que os trifólios serão destacados do caule principal e escaneados individualmente. O índice de área foliar foi obtido em função da razão entre a área foliar e a superfície ocupada pela planta. As plantas colhidas para medição de área foliar passaram pelo processo de secagem em estufa para medida do acúmulo de biomassa.

A identificação dos principais estádios fenológicos foi feita em plantas identificadas durante o crescimento da cultura, utilizando-se a escala fenológica proposta por (FEHR et al., 1971). Estas plantas identificadas foram usadas como referência para as medidas da altura da planta, medindo-se a distância vertical entre o colo da planta e nó da última folha completamente expandida (maior que 2 cm).

4.3 DADOS SIMULADOS COM MODELO SIMDUALKC DE BALANÇO HÍDRICO

Os dados foram simulados utilizando a ferramenta SimDualKc (ROSA et al., 2012), (ROSA et al., 2012) para realizar a simulação da demanda hídrica da planta e validação do abastecimento de água da área irrigada através da umidade do solo. Como entradas do modelo, são necessários parâmetros da planta e solo, coletados em campo, juntamente com dados meteorológicos e de umidade, coletados pelas estações automáticas.

Os dados inseridos do modelo precisam ser previamente calibrados, ou seja, ajustados a partir das características de clima, cultura e solo da região do experimento em comparação com os valores tabulados fixos que o software utiliza. Os métodos para calibração de clima são descritos em (ALLEN; PEREIRA, 2009) e (PEREIRA et al., 2015), em que os valores mínimos de umidade relativa do ar, velocidade do vento a 2 metros de altura e altura da planta são utilizados para equiparar os valores tabulados padrão definidos por (ALLEN et al., 1998). A calibração de clima do K_c para a região, na cultura da soja, foi determinada em função da área irrigada e replicada para a área de sequeiro. Os valores do coeficiente de cultura basal K_{cb} para o período inicial de cultivo foram de: 0,1 para o período inicial; 1,07 para o período intermediário; e 0,3 para o período final. Os coeficientes se mostraram próximos daqueles padronizados por (ALLEN et al., 1998), de 0,15; 1,10; e 0,30 para os períodos inicial, intermediário e final, respectivamente.

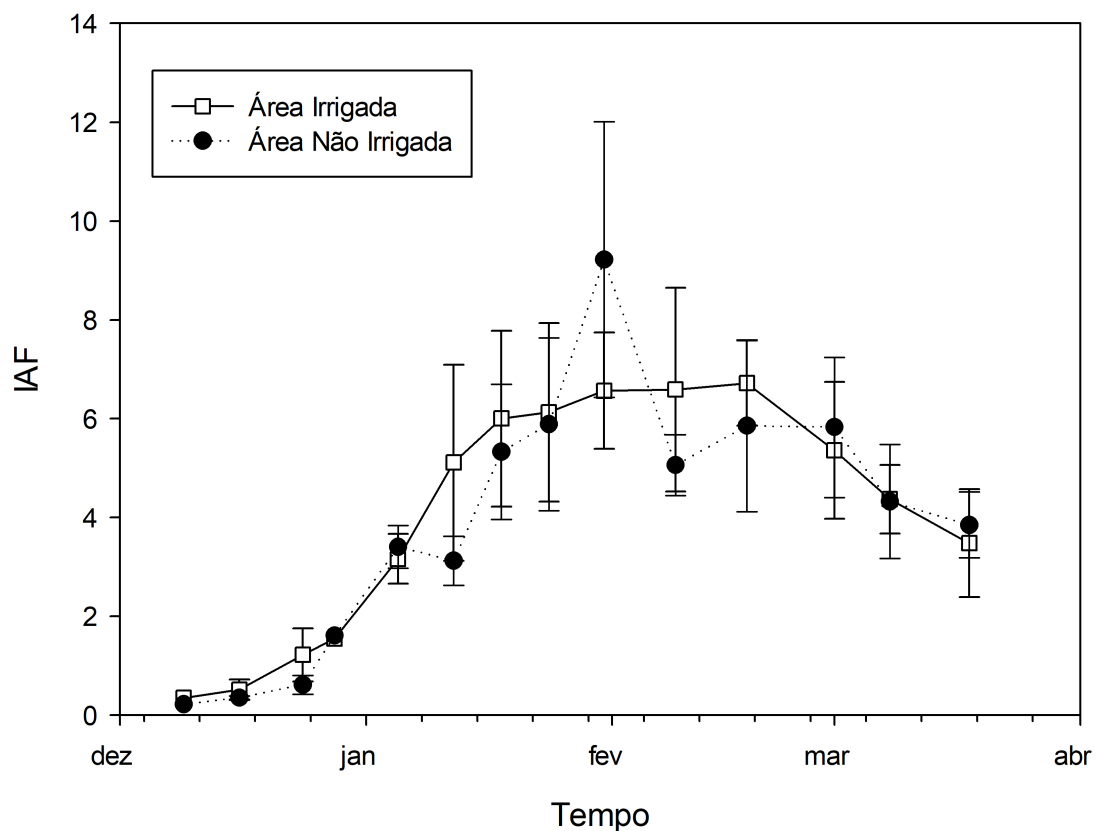
Tais modelos, como o SimDualKc, se mostram eficientes para o manejo de irrigação por facilmente processarem fatores essenciais ao balanço hídrico como clima, solo e cultura, a partir de dados que podem ser obtidos com métodos relativamente simples e estações automáticas.

Os dados de entrada de modelo são dados de solo, incluindo água total disponível na zona que a raiz da planta abrange (TAW , mm), água prontamente disponível na zona que a raiz da planta abrange (RAW , mm), valores iniciais de água prontamente evaporável (REW), espessura da camada de solo com a água evaporável (Z_e); Dados de entradas de água aplicadas, englobando lâminas (mm) aplicadas de irrigação e precipitação pluviométrica, tipo de sistema de irrigação, fração de solo molhada (f_w) e dados meteorológicos, que abrangem dados com intervalo diário a citar temperatura mínima e máxima ($^{\circ}C$), umidade relativa do ar mínima (%) e evapotranspiração de referência (mm). A evapotranspiração foi calculada externamente à ferramenta, utilizando o modelo PM-FAO56.

Além dos dados mensurados, são necessários parâmetros de entrada para o modelo no que se refere a cultura. Para tanto, os valores padrão utilizados na simulação de K_{cb} , TAW , REW e Z_e foram os publicados por (ALLEN et al., 1998), exibidos na Tabela 2 para a cultura da soja. A medida do índice de área foliar foi usada como base para a adaptação das características da planta no simulador. O Gráfico presentes na Figura 5 exibem a média do índice de área foliar, para parcela irrigada e de sequeiro, acrescidos de seu desvio padrão. A parcela irrigada demonstrou maior uniformidade em comparação com a parcela de sequeiro, cujas amostras apresentaram maior desvio padrão durante o estágio de desenvolvimento e intermediária.

rio, em cada uma das plantas analisadas. Esta variabilidade em cada amostra acontece devido ao método destrutivo de medida do *IAF*, que necessita de extração das folhas de plantas arbitrariamente escolhidas, para posteriormente contabilizar a área foliar. Logo, a cada rodada de medição, plantas diferentes são avaliadas e, em áreas de sequeiro, com maior heterogeneidade no desenvolvimento das plantas, podem apresentar maior variância.

Figura 5 – Índice de área foliar (IAF) da colheita ao plantio, para área irrigada e não irrigada. Fonte: autor.



Os valores citados na Tabela 2 representam os valores estimados para o cálculo do coeficiente de cultura, a citar: K_{cb} : coeficiente de cultura basal nos estágios inicial, intermediário e final; p : fração de depleção sem estresse; TEW : Total de água evaporável; REW : água prontamente evaporável; Z_e : camada do solo com água evaporável; CN : Curva Número⁶ (USDA, 1986); a e b são parâmetros de percolação profunda, ou seja, infiltração abaixo da zona da raiz. Os valores padrão foram calibrados para as condições da região da área de estudo, a fim de minimizar as diferenças entre os dados observados e simulados no SimDualKc, recomendados por (PEREIRA et al., 2015a).

⁶Curva Número: valores do volume de água perdido por escoamento a partir do volume de precipitação e solo.

Tabela 2 – Variáveis padrão de evaporação do solo, percolação profunda e escoamento superficial publicados por (ALLEN et al., 1998) calibrados para a região, utilizado em ambas as áreas

Variável	Valor padrão	Valor calibrado
$K_{cb\ ini}$	0,15	0,10
$K_{cb\ interm.}$	1,10	1,07
$K_{cb\ final}$	0,30	0,28
p (ini)	0,50	0,50
p (des)	0,50	0,50
p (interm.)	0,50	0,50
p (final)	0,50	0,50
TEW (mm)	49	49
REW (mm)	10	9
Ze (m)	0,15	0,10
CN	75	70
a	380	380
b	-0,005	-0,005

Para determinar o valor de água total disponível (TAW) foi utilizado o simulador e dados obtidos com análise física do solo em laboratório como a capacidade de campo e ponto de murcha permanente, que possibilitou o cálculo da água prontamente disponível (RAW), através da Equação 4.1.

$$RAW = TAWp \quad (4.1)$$

O simulador tem como saída os valores de Kc , ETc (mm), e seus subfatores, como Kcb , Ke e Ks , coeficiente basal, coeficiente de evaporação e coeficiente ajuste ao estresse hídrico, respectivamente. Estes valores são tomados como base para validação do coeficiente de cultura basal obtido por sensoriamento remoto, descritos na seção 4.7, com objetivo de avaliar a performance do modelo.

4.4 SENSORIAMENTO REMOTO

Foram utilizadas no trabalho as imagens do satélite *Sentinel 2* (European Space Agency, 2015), apresentado na seção 2.1.1, do tipo L-1C⁷, no topo da atmosfera, coletadas e processadas através da plataforma *Google Earth Engine* (GEE) (GORELICK et al., 2017). Como parte da missão do satélite *Sentinel*, está a observação de mudanças na vegetação e por isso é equipado com sensores que capturam a faixa do infravermelho próximo (NIR) assim como o espectro

⁷Imagens Level-1C: foram aplicadas correções de terreno e geométricas.

visível. Além disso, ambos os satélites *Sentinel-2a* e *Sentinel-2b*, combinados, fornecem uma cobertura de 2 a 5 dias do mesmo ponto de observação, com uma resolução espacial de 100 metros² por pixel para as bandas citadas. Por estas características é o satélite de domínio público com melhor capacidade para observação de áreas de cultivo, já que permite avaliar as mudanças no dossel durante o período de desenvolvimento rápido e resolução espacial razoável para avaliação de plantas, neste caso cobrindo aproximadamente 3000 plantas por pixel. Os índices amplamente utilizados há décadas para analisar diferenças na vegetação são o Índice de Vegetação por Diferença Normalizada (*NDVI*) (ROUSE et al., 1973) e Índice de Vegetação Ajustado ao Solo (*SAVI*) (HUETE, 1988). Tendo isso em vista, todas as imagens do *Sentinel-2*, nas bandas do vermelho (665nm, banda B4) e infravermelho próximo (NIR, 842nm, banda B8), foram processadas com os índices de vegetação, conforme descrito na seção 4.5, e analisadas do período do plantio até a colheita, tendo em vista a disponibilidade de imagens e a cobertura de nuvens ou sombreamento presente nestas. Para o diagnóstico da cobertura de nuvens, foram produzidos gráficos usando o GEE para conceber o *NDVI* médio de cada área sobre o tempo, visualizando as imagens que indicavam pontos em que o índice sofreu alterações fora do desenvolvimento esperado. A Tabela 3 apresenta todas as imagens disponíveis para o período, assinalando quais foram descartadas devido à cobertura de nuvens ou sombreamento e o período do desenvolvimento vegetativo em que cada uma foi classificada, considerando parcelas irrigada e de sequeiro. Estas diferem ligeiramente quanto à ocorrência dos estádios de desenvolvimento da cultura. De um total de 45 imagens, 22 foram selecionadas para formar o conjunto de dados, com intervalo máximo de 23 dias entre elas imagens, ocorrido durante o mês de janeiro, em que o período de chuvas foi mais intenso.

4.5 PROCESSAMENTO DE ÍNDICES DE VEGETAÇÃO

O processamento dos dados de sensoriamento remoto é feito através de cálculos sobre as faixas obtidas pelos sensores multi-espectrais. Estes índices, baseados em reflectância⁸, são amplamente empregados para determinação do volume de vegetação e do estresse hídrico de culturas. Tanto o *NDVI* quanto o *SAVI* utilizam a reflectância do vermelho e do infravermelho próximo (NIR) que proporcionam a escala da absorção da luz vermelha sobre as células de clorofila. O *SAVI* é recomendado para determinados estágios vegetativos, com IAF inferior a 3, visto que o *NDVI* mostra deficiências quando o solo não está totalmente coberto por

⁸Reflectância: proporção entre o fluxo de radiação incidente na superfície e o fluxo que é refletido.

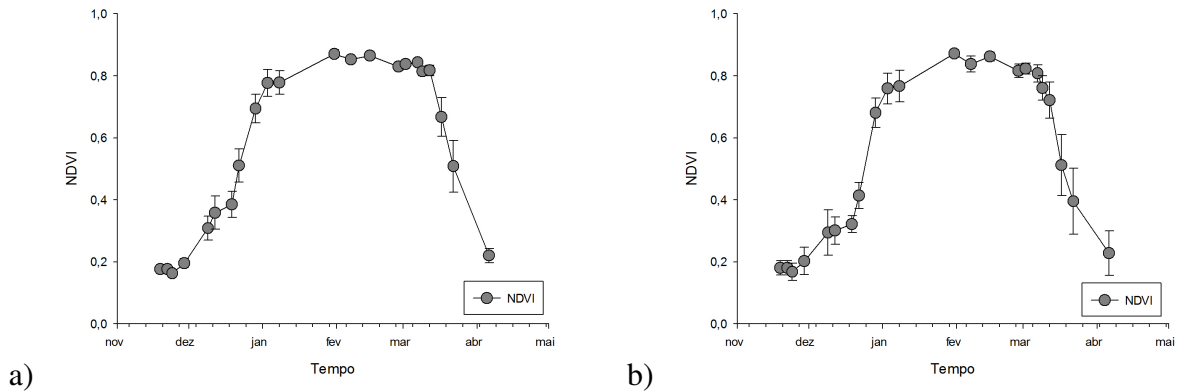
Tabela 3 – Imagens coletadas com períodos de desenvolvimento para área irrigada e de sequeiro

Data	Período ⁺		Data	Período ⁺	
	Irrig.	Seq.		Irrig.	Seq.
19/11/2017	1	1	10/02/2018 *	3	3
22/11/2017	1	1	12/02/2018 *	3	3
24/11/2017	1	1	15/02/2018	3	3
29/11/2017	1	1	17/02/2018 *	3	3
02/12/2017 *	1	1	20/02/2018 *	3	3
07/12/2017 *	1	1	22/02/2018 *	3	3
09/12/2017	2	2	25/02/2018 *	3	3
12/12/2017	2	2	27/02/2018	3	3
14/12/2017 *	2	2	02/03/2018	3	3
19/12/2017	2	2	04/03/2018 *	3	3
22/12/2017	2	2	07/03/2018	3	3
27/12/2017 *	3	2	09/03/2018	4	4
29/12/2017	3	3	12/03/2018	4	4
03/01/2018	3	3	14/03/2018 *	4	4
08/01/2018	3	3	17/03/2018	4	4
11/01/2018 *	3	3	19/03/2018 *	4	4
16/01/2018 *	3	3	22/03/2018	4	4
18/01/2018 *	3	3	24/03/2018 *	4	4
21/01/2018 *	3	3	27/03/2018 *	4	4
23/01/2018 *	3	3	29/03/2018 *	4	4
28/01/2018 *	3	3	03/04/2018 *	4	4
31/01/2018	3	3	06/04/2018	4	4
07/02/2018	3	3			

⁺ Os períodos de desenvolvimento da planta estão notados como: (1): inicial; (2): desenvolvimento rápido; (3): intermediário; (4): final.

* Imagens que foram descartadas pela presença de nuvens ou sombreamento, interferindo no cálculo dos índices de vegetação.

Figura 6 – Índice de vegetação por diferença normalizada (NDVI) da colheita ao plantio, para área irrigada (a) e não irrigada (b). Fonte: autor.



vegetação (IAF menor que 3 ou F_c menor que 70%) (ALLEN et al., 1998). O SAVI é mais apropriado em casos em que há maior exposição da superfície do solo e existem mudanças na reflectância causadas pela umidade do solo (TEIXEIRA et al., 2013) (PÔÇAS et al., 2015). O NDVI e SAVI são denotados nas Equações 4.2 e 4.3 respectivamente.

Ambos os índices serão utilizados para adequação de sua aplicação nos diferentes estágios de desenvolvimento da soja, conforme ilustrado na figura 2.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (4.2)$$

$$SAVI = \frac{NIR - RED}{NIR + RED + L} 1 + L \quad (4.3)$$

Onde NIR representa a reflectância de infravermelho próximo, RED representa a reflectância do vermelho e L representa um fator de correção para o solo usado para minimizar a radiação difusa transmitida pelo dossel, onde o valor 1 significa o maior valor para ambientes semi-áridos.

Os gráficos na Figura 6 demonstram a curva do índice de vegetação, usando a média de todos os pixels delimitados para cada área, calculado durante todo o período do experimento, indicando os períodos de troca de estágios de crescimento descritos na Tabela 3.

4.6 AVALIAÇÃO DE DESEMPENHO DOS COEFICIENTES POR SENSORIAMENTO REMOTO

A avaliação de desempenho dos dados simulados e dos coeficientes estimados com sensoriamento remoto foram realizados da seguinte forma: 1) balanço hídrico foi validado levando em consideração os dados das estações de campo convertidos para valores de porcentagem e os valores de saída do SimDualKc; 2) coeficientes de cultura por sensoriamento remoto serão comparados com o coeficiente de cultura simulado pelo SimDualKc. A análise estatística se dará através de um conjunto de indicadores descritos por (ROSA et al., 2012); (MARTINS et al., 2013) e (PAREDES et al., 2014) que também utilizaram o modelo SimDualKc. Uma regressão linear forçada à origem será gerada para comparar os dados observados, simulados e estimados, de forma que um coeficiente de regressão (b_0) próximo a 1 demonstra elevada relação entre os dados observados e simulados, e um coeficiente de determinação (R^2) próximo a 1,0 indica que a maior parte da variação dos dados observados é explicada pelo modelo, de acordo com as Equações 4.4 e 4.5:

$$b_0 = \frac{\sum_{i=1}^n O_i P_i}{\sum_{i=1}^n O_i^2} \quad (4.4)$$

$$R^2 = \left\{ \frac{\sum_{i=1}^n (O_i - O)(P_i - P)}{[\sum_{i=1}^n (O_i - O)^2]^{0,5} [\sum_{i=1}^n (P_i - P)^2]^{0,5}} \right\}^2 \quad (4.5)$$

A estimativa dos erros residuais será determinada através da raiz quadrada do erro médio (RMSE) e do erro médio absoluto (AAE), que expressam a variância dos erros e o tamanho médio dos erros estimados, respectivamente (MORIASI, 2007). Também será utilizado o erro relativo médio (ARE) para indicar o tamanho médio dos erros estimados. Estes indicadores são obtidos através da análise dos dados observados, O e P , e simulados, aos pares, respectivamente O_i e P_i ($i = 1, 2, 3, \dots, n$), e sendo calculados de acordo com as Equações 4.6, 4.7 e 4.8:

$$RMSE = \left[\frac{\sum_{i=1}^n (P_i - O_i)^2}{n} \right]^{0,5} \quad (4.6)$$

$$AAE = \frac{1}{n} \sum_{i=1}^n (O_i - P_i) \quad (4.7)$$

$$ARE = \frac{100}{n} \sum_{i=1}^n \left| \frac{O_i - P_i}{O_i} \right| \quad (4.8)$$

Estes indicadores serão calculados em cada interação da simulação para corroborar a conclusão de que os parâmetros ajustados levam à minimização dos erros de estimativa. Também foram utilizados outros indicadores, a fim de avaliar a qualidade da modelagem, como o modelo de eficiência (EF) proposto por (NASH; SUTCLIFFE, 1970), que é usado para determinar a magnitude relativa da variância residual em comparação à variância dos dados medidos (MORIASI, 2007), definida pela razão entre o erro quadrado médio e a variância nos dados observados, subtraído da unidade, conforme Equação abaixo:

$$EF = 1,0 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - O)^2} \quad (4.9)$$

Quando o valor de EF é próximo a zero ou negativo, significa que a média dos valores observados é tão boa ou melhor do que aqueles simulados pelo modelo (LEGATES; MCCABE, 1999), (MORIASI, 2007). E, por fim, será calculada a porcentagem de viés (PBIAS), que mede a tendência média dos dados simulados para ser maior ou menor do que as suas observações correspondentes, conforme a Equação 4.10:

$$PBIAS = 100 \frac{\sum_{i=1}^n (O_i - P_i)}{\sum_{i=1}^n (O_i)} \quad (4.10)$$

O valor ideal de PBIAS é 0,0; valores próximos de 0,0 indicam uma simulação do modelo exato. Valores positivos ou negativos referem-se a ocorrência de um viés de sub ou sobre estimação.

4.7 COEFICIENTES DE CULTURA COM SENSORIAMENTO REMOTO

Foram geradas geometrias para demarcar os limites da área irrigada e de sequeiro, reduzindo sua área em relação às bordas para desconsiderar valores de *NDVI* mesclados com estradas, área de preservação ambiental ou outros itens que pudessem afetar a representatividade que cada pixel cobre, conforme procedimento realizado por (PÔÇAS et al., 2015) e (CAMPOS et al., 2016). Esta medida permite que não ocorram ruídos nos dados processados com os índices de vegetação, já que apenas pixels com a cultura são avaliados. Para realizar uma estimativa com maior precisão sobre o conjunto de dados, as imagens de satélite processadas com o índice *NDVI*, foram equiparadas quanto ao seu formato, transformadas em arquivos *geojson* (GEOJSON SPECIFICATION, 2018). Este proporciona um padrão unificado para trabalhar com as imagens georreferenciadas (*rasters*⁹) em forma de matriz, em que cada pixel correspondente se torna um polígono com o respectivo valor do índice de vegetação, em uma escala de 0 a 255.

A metodologia segue a relação entre os índices de vegetação e o coeficiente de densidade *Kd* e o menor coeficiente de cultura Kc_{min} do período, apresentados por (PÔÇAS et al., 2015), aplicados na Equação 4.13.

Para determinação do coeficiente de densidade *Kd*, denotado pela Equação 4.11, se fazem necessários dois dados observados em campo: fração de cobertura efetiva (*Fc*) e altura da planta. O *Kd* pode ser estimado em função do IAF, porém, devido à variância encontrada na área de sequeiro, optou-se por utilizar a fração de cobertura efetiva. Estes podem ser observados nos gráficos da Figura 7, para a área irrigada e não irrigada, respectivamente. Os valores de altura, em centímetros, variam de 0, antes da planta emergir, à 120 cm. A fração de cobertura do solo varia entre 0 e 100%.

$$K_d = \min\left(1, M_L f_{c\text{eff}}, f_{c\text{eff}}^{\left(\frac{1}{1+h}\right)}\right) \quad (4.11)$$

A fração de cobertura do solo pode ser derivada a partir do índice de vegetação, conforme Equação 4.12.

$$F_c VI = \beta_1 \frac{VI_i - VI_{min}}{VI_{max} - VI_{min}} + \beta_2 \quad (4.12)$$

O parâmetro β_1 representa um coeficiente empírico, variando de 0 a 1, com base no valor máximo do índice de vegetação para cada estágio de crescimento da planta. β_2 é um

⁹Raster: imagem em que cada pixel contém uma informação, como por exemplo, coordenadas.

Figura 7 – Fração de cobertura do solo (F_c)(%) e altura da planta (m) para as áreas irrigada e não irrigada. Fonte: autor

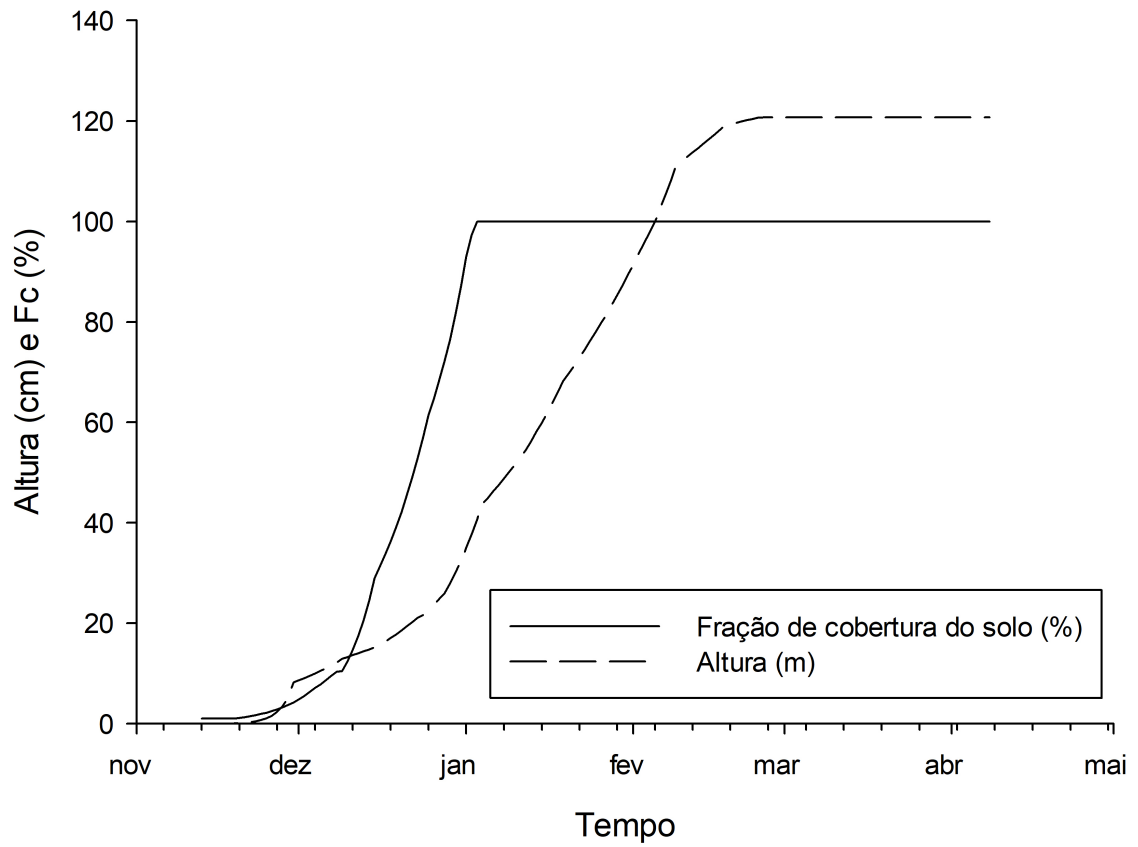


Tabela 4 – Parâmetros utilizados para o cálculo da fração de cobertura do solo com índice de vegetação

Parâmetro		Valor	
		Irrigada	Sequeiro
NDVI max	-	0.9	0.89
NDVI min	-	0.15	0.15
SAVI max	-	1	1
SAVI min	-	0.2	0.2
β_2	-	0 - 0.5	0 - 0,4
β_1	Inicial	0,3	
	Desenv.	0,6 - 0,9	0,6 - 0,9
	Interm.	1	1
	Final	1	1

fator de ajuste considerado para compensar a queda dos valores do índice de vegetação devido ao amarelamento das folhas, que é independente à F_c . Este foi calibrado de acordo com o momento da queda dos valores de VI, nas imagens de satélite. O amarelamento quantificado observando os valores de $NDVI$ processados nas imagens de satélite, a partir da primeira data com queda no valor médio total da área, indicando menor reflectância da cor verde. VI_i representa cada um dos valores de $NDVI$ para cada pixel, variando de 0 a 100. VI_{max} e VI_{min} são os valores máximo e mínimo, respectivamente, do índice de vegetação para toda a área sendo avaliada.

A Tabela 4 exibe os parâmetros que foram considerados para cada área na aplicação da Equação 4.12, levando em consideração as variações dos índices de vegetação quanto ao amarelamento das folhas e valores máximos e mínimos de $NDVI$. O F_{cVI} será utilizado para a escolha do índice de vegetação, $NDVI$ ou $SAVI$, com melhor correlação com os dados de F_c observados.

O coeficiente Kd é essencial para isolar os efeitos da densidade da vegetação da transpiração da planta e se adapta facilmente ao sensoriamento remoto já que tem boa correlação com os índices de vegetação. Tomando vantagem disso, é possível determinar - além do modelo anteriormente proposto, estimando a fração efetiva com índice de vegetação - o coeficiente de cultura basal utilizando os valores observados em campo: F_c e altura, em complemento com os índices de vegetação, conforme Equação 4.13.

$$KcbVI = Kcmin + Kd\left(\frac{VI_i - VI_{min}}{VI_{max} - VI_{min}}\right) \quad (4.13)$$

A Equação 4.13 foi utilizada para o cálculo do $KcbVI$, para as 22 imagens disponíveis,

para todos os polígonos representando os pixels da imagem de satélite, cada um com um valor de índice de vegetação calculado, normalizando os limites de 0 a 255 para valores entre 0 e 1.

O Kcb_{VI} , como coeficiente base, permite que as características da planta sejam ajustadas a partir de dados de sensoriamento remoto, para assim, de acordo com a distribuição espacial da cultura, determinar a água que a planta está perdendo por transpiração, limitando-se apenas à frequência de disponibilidade de imagens. Sendo assim, este coeficiente calculado por imagens de sensoriamento remoto auxilia na estimativa mais precisa sobre o volume de água aplicado nas irrigações, podendo vir a reduzir desperdícios de recursos hídricos.

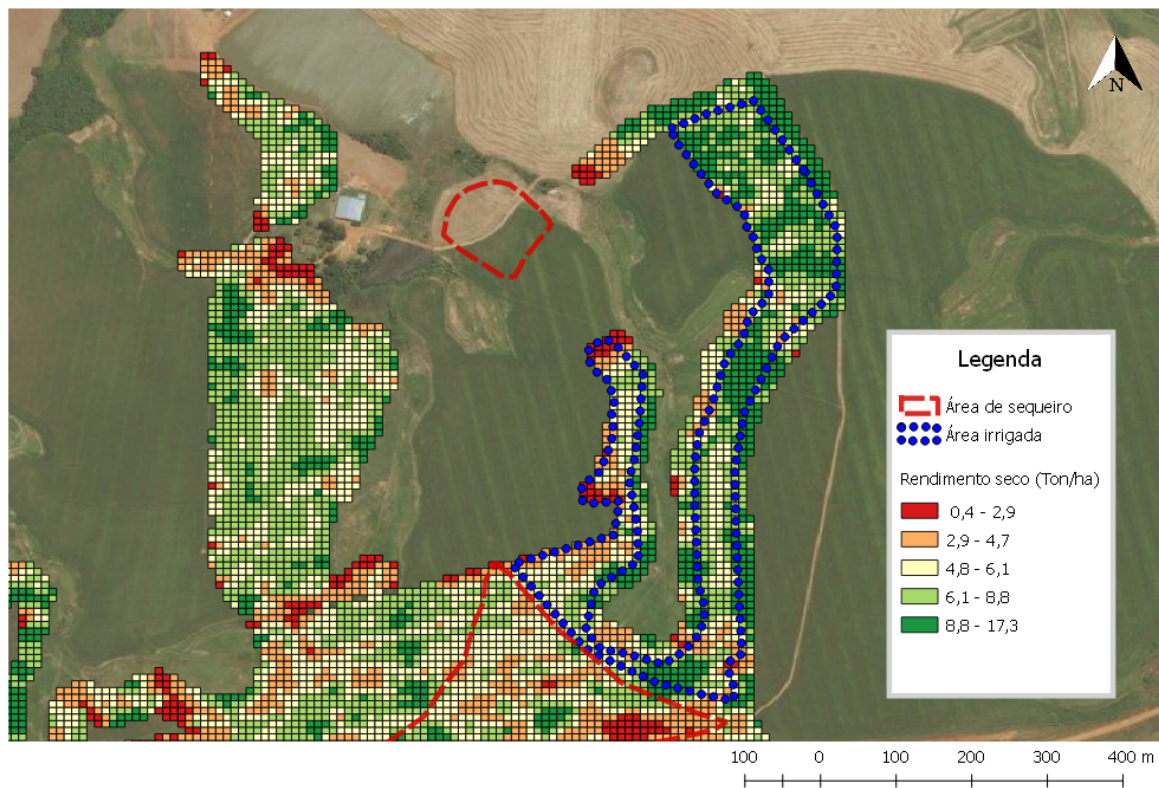
4.8 APRENDIZADO DE MÁQUINA E CLASSIFICAÇÃO

O modelo de balanço hídrico escolhido, com a abordagem do Kc Dual, fornece a diferenciação entre a água perdida na transpiração da planta da evaporação do solo. A evaporação tende a ser maior nos períodos iniciais, onde há mais solo descoberto pela vegetação e menor, quando a fração de cobertura do solo atinge valores superiores a 70%. A transpiração é um dos principais fatores relacionados ao desenvolvimento da cultura e seus rendimentos (PAÇO et al., 2014) (CAMPOS et al., 2017) (PÔÇAS et al., 2017) (P. STEDUTO T. C. HSIAO; RAES, 2012) e, por este motivo, são analisadas estratégias para correspondência da transpiração da planta com o rendimento de grãos, durante todo o período de cultivo, levando em consideração sua distribuição espacial.

4.8.1 Formação do conjunto de dados

Para as leituras dos rendimentos, foi utilizado um mapa, representado na Figura 8, contendo os pontos de colheita mecanizada georreferenciados utilizando GPS (Global Positioning System), fornecidos pela colhedora (Model CR9090, New Holland, USA). O mapa não cobre a área do experimento pois a colheita foi realizada por duas colhedoras distintas. A partir dos valores de rendimento seco, em Ton/ha, foram definidos intervalos, chamados de classes, de acordo com o algoritmo geo-estatístico de quebras naturais Jenks (JENKS, 1967), presente no QGIS (QUANTUM GIS GEOGRAPHIC INFORMATION SYSTEM, 2019), em sua versão 2.18.26. O algoritmo tem como objetivo realizar agrupamentos entre os dados de acordo com a sua distribuição, para encontrar o melhor arranjo entre as classes criadas, diminuindo a variação dentro de cada classe e aumentando a variação entre as classes.

Figura 8 – Representação geo-estatística dos intervalos de classificação para o mapa de rendimento seco (Ton ha⁻¹), destacando área irrigada e área de sequeiro. Fonte: autor.



A divisão de classes se dá a partir da ordenação do conjunto, seguido das etapas: 1. calcular a soma dos desvios quadrados entre as classes (SDBC); 2. calcular a soma dos desvios quadrados da média do conjunto (SDAM); 3. Subtrair SDAM de SDBC e igualar a soma dos desvios quadrados da média da classe (SDMC). Após inspecionar cada SDBC, é movida uma unidade da classe até a próxima SDBC; O processo é repetido e são calculados novos desvios de classe. Para avaliação do ajuste da variância é utilizada a Equação $SDAM - SDMC/SDAM$, em que valores próximos de 1 indicam um bom ajuste (JENKS, 1967). As classes definidas são apresentadas na Tabela 5 e sua distribuição em cada área de estudo pode ser visualizada na Figura 9.

Os dados de K_{cb} foram calculados na Seção 4.7, para todas as 22 datas, combinando cada ponto geográfico de intersecção entre o mapa de rendimento e o pixel da imagem de satélite. O objetivo desta intersecção é encontrar a relação que a transpiração da planta tem com os rendimentos de grãos. O $K_{cb_{VI}}$ foi calculado conforme descrição na seção 4.7 em cada ponto, de forma a distribuí-lo espacialmente. Foram formados dois conjuntos de dados, baseados na classificação de intervalos de rendimento, usando quebras naturais e a série de

Figura 9 – Histograma apresentando a distribuição dos dados de rendimento seco da soja em área irrigada e não-irrigada, com separadores verticais entre as classes. a) Rendimento seco (Ton ha⁻¹) na área irrigada; b) Rendimento seco (Ton ha⁻¹) na área não-irrigada. Fonte: autor.

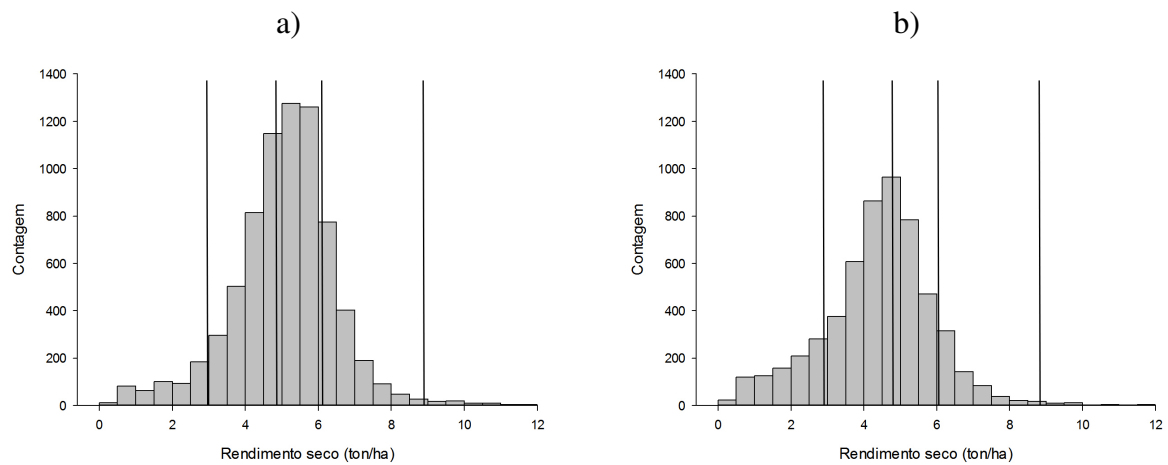


Tabela 5 – Valores em Ton ha⁻¹ utilizados para classificação do rendimento seco.

Rendimento seco
0,4 - 2,9
2,9 - 4,7
4,7 - 6,1
6,1 - 8,8
8,8 - 17,3

valores, como características, usando o Kcb_{VI} para todos os pontos, em todas as datas que compreendem o plantio à colheita, totalizando 5671 pontos para a área não irrigada (cerca de 6 ha) e 7485 pontos (cerca de 11 ha) para a área coberta pelo sistema de irrigação.

Derivações dos valores de Kcb_{VI} foram calculadas em conjunto aos dados estimados pelo SimDualKc, a citar: evapotranspiração da cultura (ET_c), conforme Equação 4.14; transpiração (Tr), em milímetros, denotado na Equação 4.15, em que não há interferência do coeficiente de evaporação; e coeficiente de cultura (Kc) dado pela Equação 4.16. Estes coeficientes, juntamente com o Kcb_{VI} foram usados como características para novos conjuntos de dados, isoladamente, para predição do rendimento seco de grãos.

$$ET_c = ET_o(K_s Kcb_{VI} + K_e) \quad (4.14)$$

$$Tr = K_s Kcb_{VI} * ET_o \quad (4.15)$$

$$Kc = K_s Kcb_{VI} + K_e \quad (4.16)$$

4.8.2 Definição de classes e pré-processamento

A classificação foi feita utilizando a biblioteca de código-aberto *Scikit Learn* (PEDREGOSA et al., 2011). Esta conta com diversos recursos para aprendizado de máquina, com possibilidade de incorporar conjuntos de dados e aplicar algoritmos de aprendizado supervisionado e não supervisionado. Para este trabalho, foram elencados alguns algoritmos de aprendizado de máquina supervisionados, também chamados de classificadores, tendo como objetivo a classificação dos rendimentos com base nos atributos de transpiração da soja.

Antes de aplicação do modelo de classificação, os dados passaram por método de filtragem para remoção de *outliers*, ou seja, pontos fora da curva, atípicos, que estão distantes dos demais valores da distribuição. A presença de *outliers*, se deve à variabilidade dos valores de rendimento contidos dentro de cada pixel da imagem de satélite, usada para obter as características a partir do Kcb_{VI} . No caso das imagens do Sentinel-2, a resolução de 10m x 10m conteve mais de 10 medições de rendimento em cada polígono, portanto, os valores de VI e, conseqüentemente, Kcb_{VI} se repetiram nas classes. Para tanto, para cada uma das datas dentro da distribuição, são analisados os valores das características (Tr , Kcb , Kc e ET_c) obtidos e delimita-se quais serão valores considerados dentro do intervalo de dispersão, e quais são

outliers. Para filtragem dos *outliers*, foi aplicado o método de intervalo interquartil (IQR), aplicado em (SCHWALBERT et al., 2018), em que a dispersão aceita apenas valores entre os quartis, a considerar que são valores do terceiro quartil subtraídos do quartil Q_3 , conforme a Equação $IQR = Q_3 - Q_1$. É mais robusta e eficaz na busca de *outliers* do que a simples abordagem usando desvio padrão e média.

Os valores então passaram pelo processo de padronização, em que a média das características é removida e é feita um dimensionamento para a variância da unidade. O processo segue a Equação $z = (x - u)/s$, em que z é o valor padronizado, x é o valor da característica e s é o desvio padrão do conjunto de treino. Esse processo é indicado quando há várias características de mesma magnitude, para ficarem mais próximos da distribuição normal. É recomendado para uso em algoritmos como o *Support Vector Machines*, onde uma característica com variância maior que as demais torna o classificador incapaz de interpretar as características restantes. A função utilizada na biblioteca *scikit-learn* foi a *StandardScaler.fit_transform* do pacote *preprocessing* (PEDREGOSA et al., 2011).

4.8.3 Classificadores e avaliação de performance

Os conjuntos de dados para área irrigada e não irrigada, considerando as variáveis de transpiração (Tr), coeficiente de cultura basal (K_{cbVI}), coeficiente de cultura (K_cVI) e evapotranspiração da cultura (ET_c), foram colocados como entrada dos modelos de aprendizado supervisionado, baseado nos algoritmos abordados por em PÔÇAS et al. (2017) e os sumarizados por (GANDGE; SANDHYA, 2018), detalhados na Seção 2.4.1.

Ao final, foram utilizadas duas abordagens para avaliação de desempenho dos classificadores: 1) *Hold-out 50/50*, é uma divisão do conjunto de treino e teste com validação estendida, com proporção 50/50, executadas 250 vezes para obter acurácia média das repetições; 2) validação cruzada utilizando *k-fold* (KOHAVI, 1995) para definição dos conjuntos de treinamento e teste. A divisão ocorre até que k experimentos de predição sejam executados, dividindo o conjunto de testes e treinamento. Neste trabalho foram utilizadas as duas validações: validação cruzada *5-fold* levando em consideração que serão cinco classes de rendimento de grãos, e cada treinamento deve ter amostras suficientes para a predição e validação estendida. A principal vantagem da validação cruzada é que o conjunto de dados é usado completamente e, da mesma forma que o *hold-out 50/50*, a média da acurácia é obtida.

O objetivo principal da classificação citada é a predição dos rendimentos obtidos durante

a colheita, a partir de atributos da transpiração da planta, comprovando a relação entre a o consumo de água para a planta e os rendimentos de grãos.

5 RESULTADOS

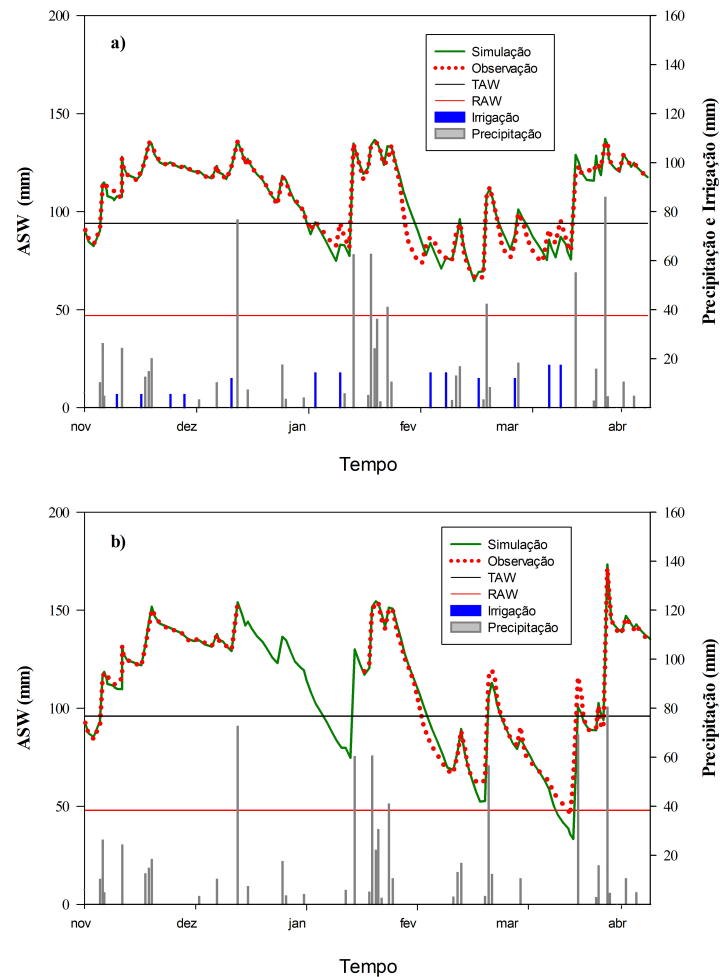
São apresentados os resultados neste capítulo, visando trazer elucidações aos problemas de pesquisa e objetivos citados no capítulo de introdução, discutindo sobre aspectos que levaram aos resultados obtidos. Como resultados estão contidos: 1. a calibração do modelo para simulação, SimDualKc; 2. a determinação do índice de vegetação que melhor se adapta às condições de campo, a partir dos resultados da fração de cobertura do solo estimada por índice de vegetação; 3. o coeficiente de cultura basal e conseqüente coeficiente de cultura, também calculados a partir de índice de vegetação, combinados com a altura medida em nível de campo; 4. avaliação de modelos de predição classificatória do rendimento de grãos a partir de derivados do coeficiente de cultura basal.

5.1 CALIBRAÇÃO DO MODELO SIMDUALKC PARA SIMULAÇÃO DO COEFICIENTE DE CULTURA

Como primeiro passo para a execução do objetivo do trabalho, a calibração do modelo às condições de campo são fundamentais, uma vez que as saídas encontradas no SimDualKc são parâmetros de comparação e validação para os demais resultados do trabalho. Conforme indicação de (PEREIRA et al., 2015b), o modelo do SimDualKc deve ser ajustado às características da cultura, do solo, do clima da área de estudo, de forma a adaptar as entradas do modelo. Este processo é denominado calibração e é de fundamental importância para que o balanço hídrico realizado se aproxime ao máximo das condições observadas em campo. A calibração do SimDualKc é aplicada aos parâmetros da cultura K_{cb} para os estágios de crescimento, assim como parâmetros de evaporação com percolação profunda e curva número de escoamento com base em tentativa e erro, até que os menores erros sejam encontrados (PÔÇAS et al., 2015).

Na Figura 10, pode-se observar a água disponível no solo para ambas as áreas do experimento, a partir dos dados medidos e convertidos da estação de campo em contraposição com as simulações da ferramenta SimDualKc. Ainda, pode-se atentar às respostas das entradas de água por precipitação pluviométrica em ambas as áreas, e, irrigação na área coberta pelo sistema de irrigação. A Tabela 6 contém a avaliação da calibração, levando em consideração a quantidade de água no solo, com coeficiente de determinação R^2 e regressão b_0 próximos a 1, erros médios por volta de 4 a 5 mm, eficiência EF próximo a 1 e PBIAS próximo de 0. Com estimativas

Figura 10 – Água disponível no solo. Exibe o movimento da água no solo, em contraposição de chuvas e irrigações, para áreas a) irrigada e b) não-irrigada, considerando os dados de umidade medidos pela estação automática de coleta de umidade e os valores de umidade simulados pelo modelo. O período com falha para os valores observados ocorreu devido à falha no armazenamento de dados da estação. Fonte: autor



pequenas de erros a calibração mostrou boa performance para o aproveitamento neste trabalho. Pode-se perceber também um período de *déficit* hídrico na área não-irrigada (b), por volta março de 2018, onde o coeficiente de estresse K_s foi diferente de 1.

A Figura 15 representa o coeficiente de cultura potencial e real, considerando o coeficiente de evaporação. O coeficiente de evaporação foi maior no período inicial uma vez que, apesar da presença de palhada, o solo estava descoberto, favorecendo tanto a incidência de radiação solar na superfície quanto a evaporação do solo sem obstrução. Ainda durante o período de crescimento rápido, elevados volumes de precipitação ocorreram, na faixa de 90 mm entre os meses de Dezembro e Janeiro, saturando o solo, favorecendo a perda de água por escoamento superficial e evaporação. No período onde houve estresse hídrico, o valor de $K_{C_{real}}$ foi menor

Tabela 6 – Indicadores estatísticos a respeito da relação entre os dados de água disponível no solo (ASW) com os valores simulados pelo SimDualKc após calibração do modelo para as condições locais.

heightÁrea	b	R ²	RMSE (mm)	ARE (%)	AAE (mm)	EF	PBIAS
Irrigada	1,00	0,96	4,16	2,79	2,59	0,95	-0,13
Não-irrigada	1,00	0,99	4,85	3,52	3,02	0,99	0,24

que o potencial pela ação do coeficiente de estresse K_s e tornou a aumentar devido à incidência de cerca de 80mm de precipitação pluviométrica.

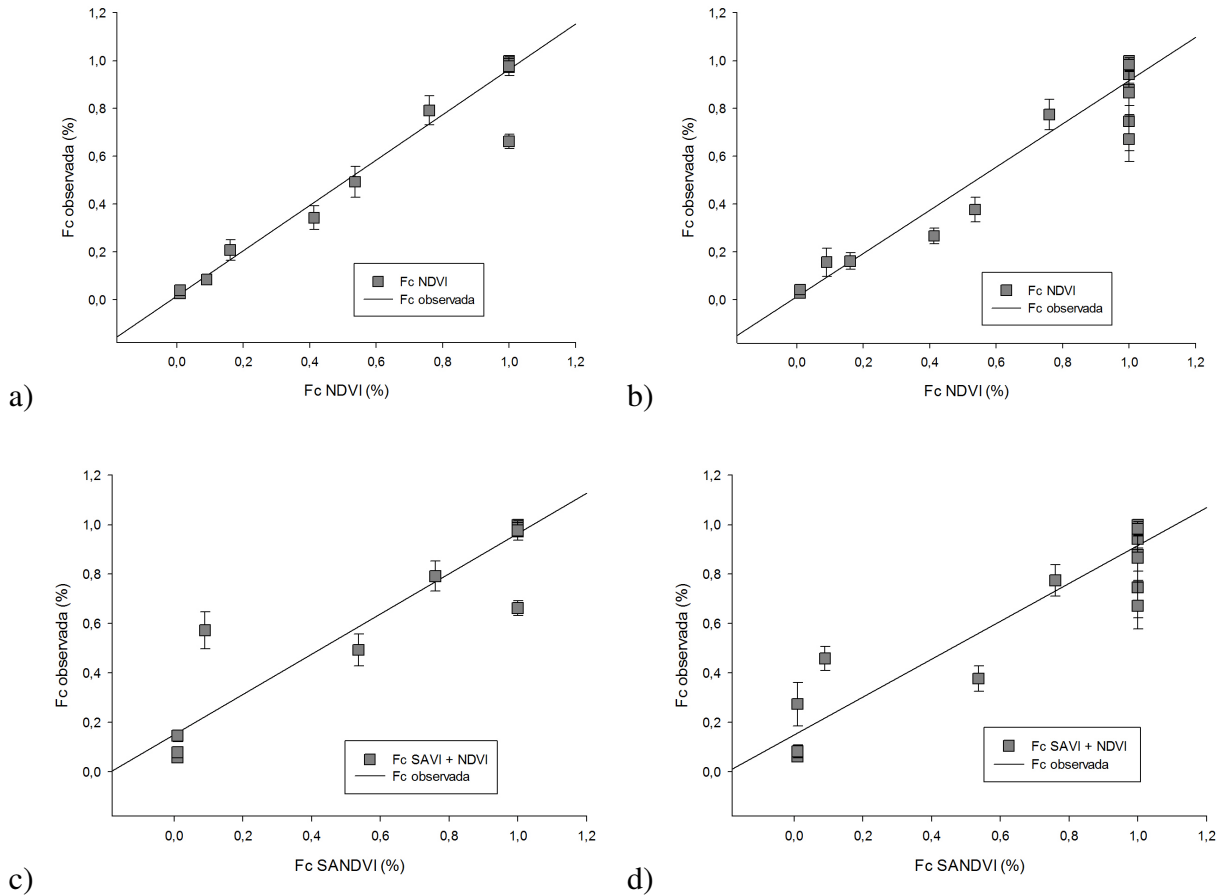
5.2 ESTIMATIVA DA FRAÇÃO DE COBERTURA DO SOLO UTILIZANDO ÍNDICES DE VEGETAÇÃO

Para determinar o índice de vegetação que melhor condiz com a situação em campo, foi utilizada a fração de cobertura do solo observada em comparação com a estimada por sensoriamento remoto. Para a estimativa desta característica foram considerados os índices de vegetação *NDVI* e *SAVI*, aplicados à Equação 4.12, reportada por (ALLEN et al., 1998). Ambos os índices utilizam as mesmas bandas¹⁰ multi-espectrais: vermelho e infra-vermelho próximo, porém o *SAVI* difere quanto ao ajuste à exposição da reflectância do solo e umidade superficial, permitindo um fator de ajuste, conforme Equação 4.3. Para o fator de ajuste, foi utilizado o valor 0,5 descrito em (PAÇO et al., 2014), (CAMPOS et al., 2018) e (PÔÇAS et al., 2015). Quando este fator de ajuste é 0, este se torna idêntico ao *NDVI*. Por este motivo, foram propostos modelos utilizando apenas o *NDVI* para todo o período de plantio e também um modelo complementar, usando *SAVI* para valores de IAF inferiores a 3.

A variabilidade espacial dos índices de vegetação faz com que a escolha e combinação destes sejam realizadas para, posteriormente, realizar estimativas de rendimento (CHLINGARYAN; SUKKARIEH; WHELAN, 2018). Para tanto, foram examinados dois conjuntos de dados: 1) utilizando apenas valores de *NDVI*; 2) utilizando valores de *SAVI* do início até 70% da fração cobertura do solo e *NDVI* para o ciclo restante, intitulado por este trabalho de *SANDVI*. Para avaliação inicial sobre qual abordagem se adequa melhor ao problema, foi realizada a estimativa da F_c utilizando os índices de vegetação *NDVI* e *SAVI*, proposta por (PÔÇAS et al., 2015). O fator M_L , que multiplica à F_c para simular os limites físicos da planta da raiz, caule e folhas, foi determinado a partir dos dados em (GRATTAN et al., 1998), que

¹⁰banda: uma determinado intervalo dentro do espectro da luz

Figura 11 – Fração de cobertura do solo estimada, para área irrigada e de sequeiro, utilizando diferentes índices de vegetação. a) Irrigada NDVI; b) Não Irrigada NDVI; c) Irrigada SANDVI; d) Não irrigada SANDVI. Fonte: autor.



aponta $M_L = 2,0$ para leguminosas.

Os valores observados em campo para F_c e altura foram linearizados utilizando progressões geométricas entre os intervalos das medições para intersecção destas com as datas de captura das imagens pelo satélite. O conjunto de dados foi reduzido, já que é calculado para cada pixel, utilizando a média aritmética para ambas as áreas. Não foram levadas em consideração as três últimas imagens, relativas aos dias 17/03/2018, 22/03/2018 e 06/04/2018, visto que marcam o início da queda de F_c devido ao amarelamento e senescência, e, por este motivo, não foram considerados nas observações em campo.

A Figura 11 exibe um gráfico de dispersão, apontando as quatro avaliações confrontadas com a fração de cobertura do solo observada em campo, apresentando em (a) a F_c para a área irrigada considerando apenas $NDVI$; em (b) a F_c para área não irrigada também considerando $NDVI$; em (c) a F_c para a área irrigada considerando a combinação $SANDVI$; e (d)

Tabela 7 – Indicadores estatísticos de correlação entre os valores estimados e observados de fração de cobertura do solo.

	Área	VI	b_0	R^2	RMSE	AAE	ARE	EF	PBIAS
a)	Irrigada	NDVI	0,99	1	0,06	0,02	49,83	1	0,33
b)	Não-irrigada	NDVI	0,96	0,98	0,06	0,04	59,21	0,98	3,66
c)	Irrigada	SANDVI	0,97	0,93	0,13	0,06	213,31	0,92	-6,06
d)	Não-irrigada	SANDVI	0,96	0,98	0,13	0,1	295,42	0,92	-6,64

representa F_c para a área não irrigada utilizando *SANDVI*.

Os indicadores estatísticos nesta comparação entre a F_c efetiva observada e F_cVI são apresentados na Tabela 7. A área irrigada apresentou os melhores indicadores, principalmente para o uso de *NDVI*, com erro quadrado médio *RMSE* de 8%, fator de determinação $R^2 = 0,97$ e fator de correlação b_0 próximo a 1. Ambas as áreas, mostraram melhores ou equivalentes fatores de correlação e determinação para o índice *NDVI*, com os indicadores de erros *RMSE*, *AAE* e *ARE* indicando menores variações nos erros encontrados para o índice aplicativo individualmente. O *NDVI* se mostrou mais eficiente segundo o modelo de eficiência residual¹¹ *EF* com uma sobre-estimativa de 0,33 e 3,66, para as áreas irrigada e não irrigada, respectivamente, segundo indicador *PBIAS*.

Para os fatores seguintes do cálculo do *KcbVI* serão utilizadas as estimativas de F_c com sensoriamento remoto para o índice *NDVI*, uma vez que foi apontado com melhor correlação entre os dados da vegetação encontrados em superfície.

5.3 ESTIMATIVA DO COEFICIENTE DE CULTURA BASAL COM BASE EM REFLEC-TÂNCIA DE ÍNDICES DE VEGETAÇÃO (KCB VI)

Os coeficientes basais com sensoriamento remoto *KcbVI* foram estimados com base em cada pixel da imagem de satélite, utilizando a Equação 4.13, em que um coeficiente de densidade K_d , proposto por (ALLEN; PEREIRA, 2009), é calculado em função de valores máximos, mínimos e para cada pixel do Índice de vegetação por diferença normalizada (*NDVI*). Estes são calculados, geograficamente distribuídos, com objetivo de determinar as diferenças transpirativas da soja espacialmente, para posterior relação com os rendimentos de grãos encontrados.

A Figura 12 retrata a dispersão dos valores de *KcbVI*, considerando como referência os valores de *Kcb* simulados pelo SimDualKc. A Figura 13 demonstra a distribuição e o com-

¹¹Resíduo: diferença entre os valores observados e estimados.

Figura 12 – Coeficientes de cultura basais obtidos por sensoriamento remoto (KcbVI), para área irrigada (a) e não irrigada (b). Fonte: autor.

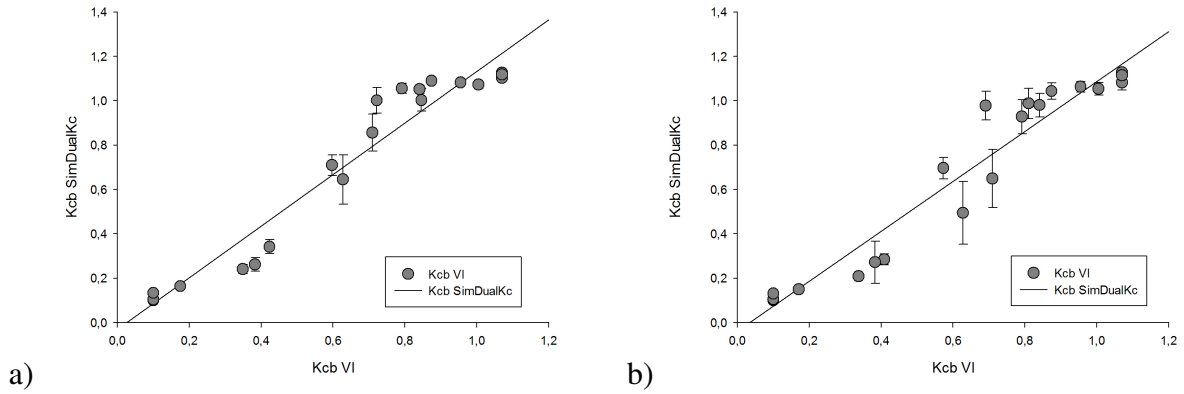


Figura 13 – Coeficientes de cultura basais obtidos por sensoriamento remoto (KcbVI) em comparação aos coeficientes basais simulados pelo modelo SimDualKc, para área irrigada (a) e não irrigada (b). Fonte: autor.

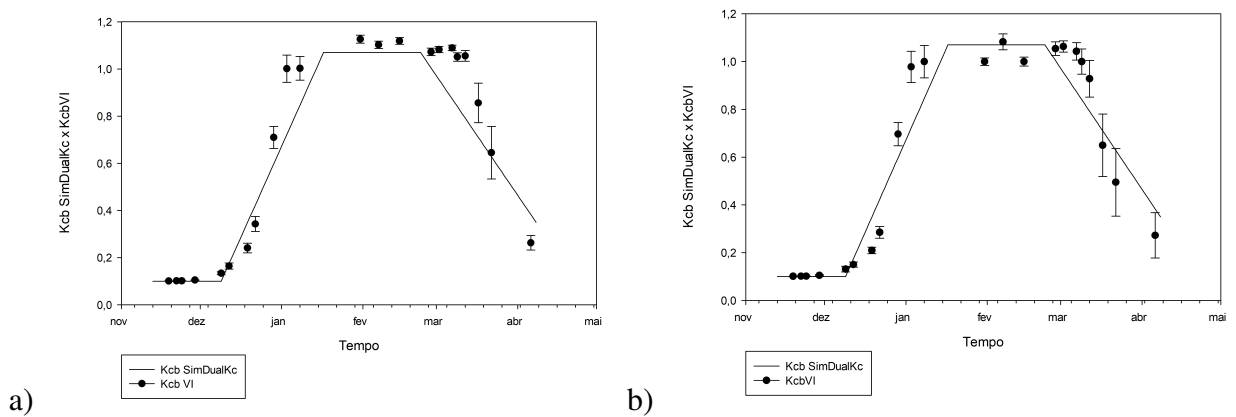


Tabela 8 – Indicadores para correlação entre os coeficientes de cultura basal simulado com SimDualKc e estimado com índice de vegetação (KcbVI)

Área	b_0	R^2	RMSE	AAE	ARE	EF	PBIAS
Irrigada	1,13	0,96	0,06	0,08	12,01	0,89	-11,18
Não-irrigada	1,05	0,93	0,06	0,08	13,35	0,91	-6,71

portamento da curva, validado pelo coeficiente de determinação R^2 , do início do plantio até a colheita. A partir dos indicadores na Tabela 8, a estimativa teve coeficiente de determinação $R^2 = 0,96$ e $0,93$, para área irrigada e não-irrigada, respectivamente, demonstrando a proporção entre os valores estimados por sensoriamento remoto possuem com os mesmos simulados, com coeficientes de regressão superiores a 1. A variância dos erros foi mínima, com RMSE = $0,06$, identificando tamanho dos erros maior com ARE por volta de 13. PBIAS indica uma subestimativa do modelo, para ambas as áreas.

Para a estimativa do coeficiente final, o coeficiente de cultura real, foram combinados os valores de K_e e K_s , utilizando a Equação $K_c = K_s K_{cb_{VI}} + K_e$, em que ambos são originados da saída do *software* SimDualKc. Na Figura 14 é possível analisar a dispersão dos valores encontrados de K_c por sensoriamento remoto e os valores simulados. Na Figura 15 é possível ambientar os valores de $K_{cb_{VI}}$ estimados sobre o panorama do balanço hídrico durante todo o período de cultivo da soja. Com este gráfico, constata-se que, por mais que os valores simulados sejam base para validação, eles são calculados sobre uma média das observações de campo. Já os indicadores de $K_{cb_{VI}}$ consideram a distribuição espacial da área, logo, existem mais amostras observadas sobre toda a área. Os resultados dos indicadores foram bastante semelhantes aos mostrados para o $K_{cb_{VI}}$, com $R^2 = 0,93$ e $0,92$, para área irrigada e de sequeiro, respectivamente, conforme Tabela 9.

Figura 14 – Coeficientes de cultura obtidos por sensoriamento remoto ($K_c VI$), para área irrigada (a) e não irrigada (b). Fonte: autor.

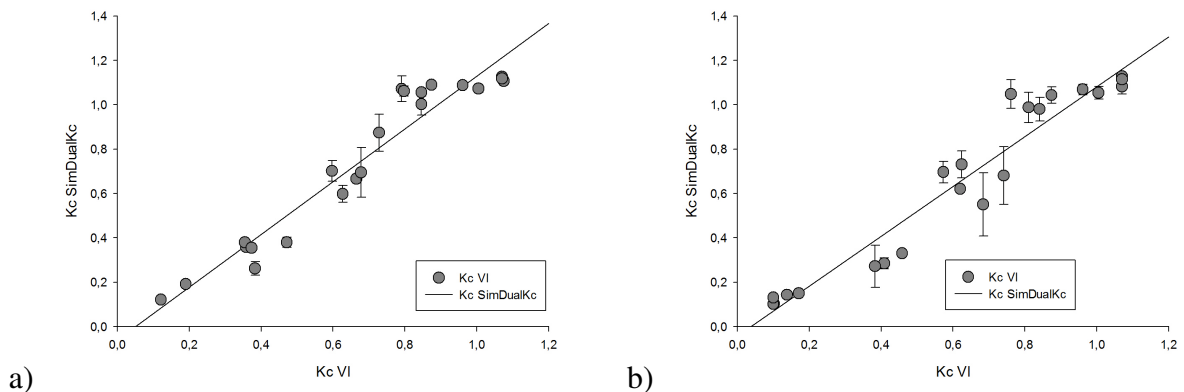
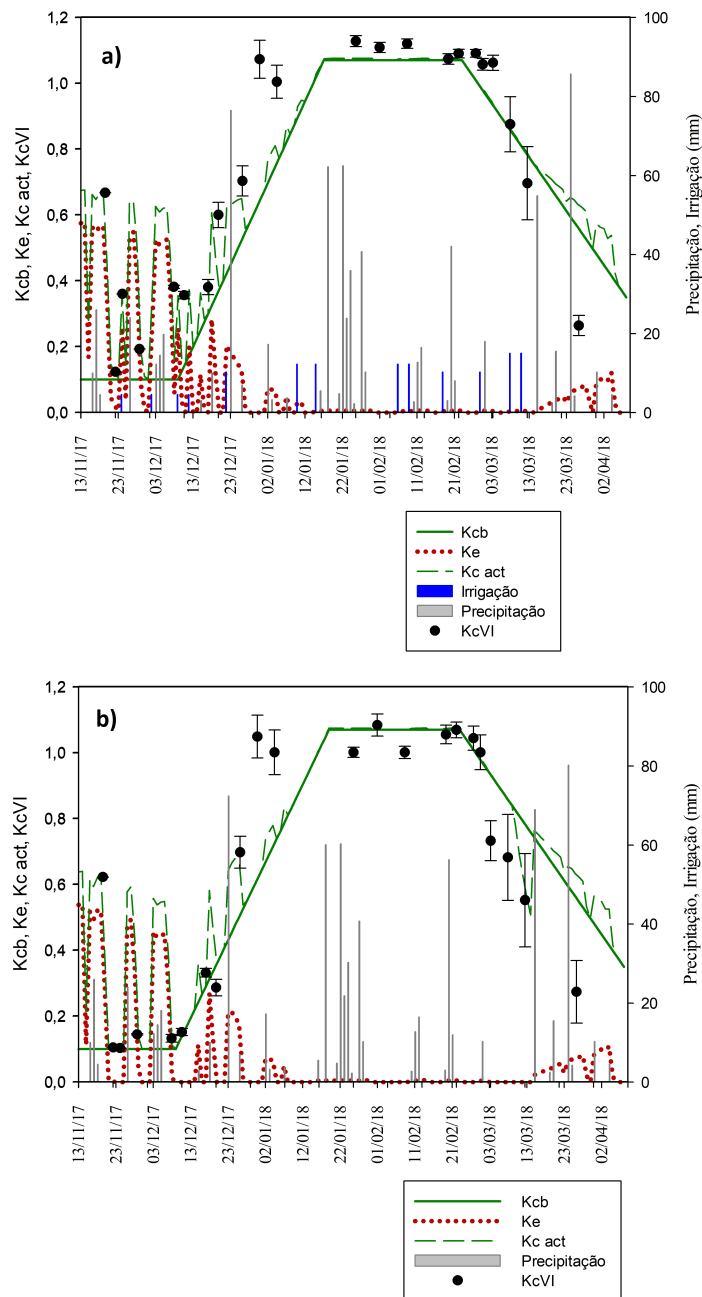


Tabela 9 – Indicadores para correlação entre os coeficientes de cultura simulado com SimDualKc e estimado com índice de vegetação (KcVI).

Área	b_0	R^2	RMSE	AAE	ARE	EF	PBIAS
Irrigada	1,11	0,93	0,06	0,08	10,23	0,82	-9,78
Não-irrigada	1,05	0,92	0,06	0,08	12,6	0,89	-6,15

Figura 15 – Coeficiente de cultura (Kc) simulados, potenciais e reais (act), considerando o coeficiente de evaporação do solo (Ke) e o coeficiente basal Kcb , para área a) irrigada e b) não irrigada. Fonte: autor.



Tanto o K_{cb} quanto o K_c , estimados com o índice de vegetação NDVI, para a cultura da soja, se mostraram confiáveis e consistentes, com coeficientes de determinação superiores a 0,9. Mesmo para a área não-irrigada, onde ocorreu período de estresse hídrico, a representatividade do K_c foi semelhante à área irrigada. Os valores médios encontrados de K_{cbVI} foram similares aos apontados por (ALLEN; PEREIRA, 2009), anteriormente citados na Tabela 2: 1) K_{cb} inicial, 0,13 (0,15); 2) K_{cb} intermediário: 1,13 (1,10); 3) K_{cb} final 0,2 (0,3);

O resultado obtido mostra que é possível obter o coeficiente de cultura por meio de índices de vegetação para a soja, realizando assim o ajuste da condição da planta à evapotranspiração de referência, limitando-se apenas à disponibilidade de imagens livre de nuvens no período. Neste trabalho, com intervalo de imagens mínimo de 2 dias e máximo de 23 dias, é possível afirmar que o modelo é eficiente quanto às estimativas e proporciona uma forma relativamente simples de precisar a demanda transpirativa da planta, necessitando apenas do dado de altura da mesma. Ainda, com parâmetros de solo, dados meteorológicos e entradas de água, a partir de uma calibração prévia, o modelo SimDualKc permite combinar suas saídas com o K_{cb} para apontar o K_c apropriadamente para a cultura da soja.

5.4 COMBINAÇÃO DE COEFICIENTES DE CULTURA OBTIDOS POR SENSORIAMENTO REMOTO PARA PREDIÇÃO DE RENDIMENTOS

Para a comparação entre a característica utilizada para classificação, foram comparados sete (7) algoritmos de classificação supervisionada, para as duas áreas do estudo, a fim de determinar a performance de cada modelo de acordo com sua acurácia, em dois métodos de validação. A Tabela 10 exhibe as avaliações realizadas, considerando como característica para a classificação o coeficiente de cultura basal obtido por sensoriamento remoto (K_{cbVI}), o coeficiente de cultura dual obtido por sensoriamento remoto (K_cVI), a transpiração da planta, calculada em função da Equação $Tr = ET_o * K_{cb} * K_s$ onde o coeficiente de evaporação não é considerado no cálculo da ET_c , isolando as demandas transpirativas da cultura com o K_{cb} , com o fator de estresse K_s e a evapotranspiração e referência ET_o . Por fim, a evapotranspiração da cultura (ET_c), representando a quantidade de água perdida pelo solo e planta por evaporação e transpiração, respectivamente. O coeficiente de cultura basal K_{cb} , estimado por sensoriamento remoto, para o cálculo de todos os coeficientes, e, agregado a ele, os demais fatores como K_e , K_s e ET_o , cujas estimativas foram realizadas pelo SimDualKc como média para toda a área.

Para determinação do melhor classificador, foram estimadas a acurácia considerando

duas validações, conforme a Tabela 10: 1) acurácia da predição considerando divisão de 50% do conjunto de dados de teste e 50% para treinamento, com repetição de 250 vezes obtendo a média da acurácia entre todas as iterações; 2) acurácia da predição utilizando validação cruzada pela a estratégia 5-fold; e 3) Precisão da validação cruzada pela a estratégia 5-fold.

O algoritmo de *Decision Tree* ou Árvore de decisão, com profundidade de 10 níveis, foi o classificador que encontrou a maior acurácia do conjunto de dados para a área de sequeiro, considerando *K_{cb}* como atributo de classificação. Este obteve acurácia de 95% e precisão de 76% e acurácia de hold-out 50/50 de 100%. A árvore de decisão foi especialmente útil ao passo que define ramificações a partir de dados com maior variação, como o caso observado nas observações de índice de área foliar das áreas de sequeiro.

Para a área irrigada, o classificador que melhor se ajustou à transpiração da planta foi o *Adaboost* com 50 estimadores para potencialização de Árvore de Decisão, com acurácia de 97,1%, precisão de 97% e acurácia de hold-out 50/50 de 92,3%, tanto para *K_{cb}* quanto para *ET_c*. Em segundo lugar, o classificador *Gaussian Naive Bayes* foi o segundo melhor em desempenho para avaliar *K_{cb}* e *K_c*, com 94,3%, 86% e 92,3%, para acurácia de validação cruzada, precisão de validação cruzada e acurácia de validação estendida.

O classificador de árvore de decisão utilizando apenas *K_{cb}* como parâmetro é o principal resultado encontrado neste trabalho, uma vez que pode ser obtido de maneira simplificada, apenas com medições ou estimativas da altura da planta e imagens multi-espectrais de satélite.

A classificação com árvore de decisão foi abordada por GANDGE; SANDHYA (2018), com acurácia de 87% para a soja. Tal algoritmo divide cada característica em um conjunto binário de informações, distribuindo-as em novos níveis, sucessivamente, até que o nível estipulado (10), seja alcançado. Este modelo se torna especialmente útil nesta análise, já que a mesma característica de *K_{cb}* foi distribuída temporalmente no conjunto de dados e a árvore de decisão distribuiu todos os valores em nós, tendo como folhas os rendimentos. O classificador *Adaboost* também utiliza a mesma abordagem, porém realiza mais iterações com árvores de decisão potencializadas, alterando pesos distribuídos em cada data que o *K_{cb}* foi estimado. Considerando que alguns períodos de demanda transpirativa são mais críticos para o rendimento de grãos e aumento da biomassa da cultura, o modelo soube explorar estas potencialidades características do conjunto.

Tabela 10 – Comparação entre classificadores supervisionados com base em derivações do coeficiente de cultura basal $KcbVI$: coeficiente de cultura ($KcVI$), transpiração da cultura em mm (Tr), evapotranspiração da cultura em mm ($ETcVI$) como alvo a estimativa de rendimento de grãos.

Coef ⁺	Classificador [*]	Irigada			Não-irrigada		
		a	b	c	a	b	c
Kcb	SVM RBF ¹	53,8%	57,5%	28,9%	75,0%	73,0%	44,0%
Kc	SVM RBF ¹	53,8%	57,5%	28,9%	75,0%	73,0%	44,0%
T	SVM RBF ¹	53,8%	57,5%	28,0%	75,0%	73,0%	44,0%
ETc	SVM RBF ¹	53,8%	57,5%	29,0%	75,0%	73,3%	44,0%
Kcb	kNN ²	53,8%	70,0%	57,0%	75,0%	73,3%	44,0%
Kc	kNN ²	53,8%	70,0%	57,0%	75,0%	73,3%	44,0%
Tr	kNN ²	53,8%	70,0%	57,0%	75,0%	73,0%	44,0%
ETc	kNN ²	53,8%	70,0%	57,0%	75,0%	73,0%	44,0%
Kcb	Random Forest ³	74,6%	71,0%	52,0%	73,0%	53,0%	54,0%
Kc	Random Forest ³	74,5%	77,9%	56,0%	75,0%	56,7%	64,0%
Tr	Random Forest ³	72,6%	72,2%	61,0%	73,0%	30,0%	51,0%
ETc	Random Forest ³	73,7%	70,0%	54,0%	73,0%	51,7%	44,0%
Kcb	Árvore de decisão (10) ⁴	81,5%	94,3%	96,0%	100,0%	95,0%	76,0%
Kc	Árvore de decisão (10) ⁴	82,6%	91,4%	97,0%	100,0%	72,0%	73,0%
Tr	Árvore de decisão (5) ⁵	81,7%	91,4%	100,0%	100,0%	81,7%	77,0%
ETc	Árvore de decisão (10) ⁴	82,8%	94,3%	92,0%	100,0%	70,0%	54,0%
Kcb	Gaussian NB ⁶	92,3%	94,3%	86,0%	75,0%	85,0%	71,0%
Kc	Gaussian NB ⁶	92,3%	94,3%	86,0%	75,0%	85,0%	71,0%
Tr	Gaussian NB ⁶	92,0%	94,3%	86,0%	75,0%	85,0%	71,0%
ETc	Gaussian NB ⁶	92,0%	94,0%	86,0%	75,0%	85,0%	71,0%
Kcb	Gaussian Process ⁷	84,6%	75,0%	60,0%	75,0%	46,7%	47,0%
Kc	Gaussian Process ⁷	84,6%	75,0%	60,0%	75,0%	46,7%	47,0%
Tr	Gaussian Process ⁷	84,6%	75,0%	60,0%	75,0%	46,7%	47,0%
ETc	Gaussian Process ⁷	84,6%	75,0%	60,0%	75,0%	46,7%	47,0%
Kcb	AdaBoost ⁸	92,3%	97,1%	97,0%	100,0%	85,0%	58,0%
Kc	AdaBoost ⁸	91,7%	97,1%	97,0%	100,0%	85,0%	71,0%
Tr	AdaBoost ⁸	91,7%	94,3%	97,0%	100,0%	85,0%	71,0%
ETc	AdaBoost ⁸	92,3%	97,1%	97,0%	100,0%	85,0%	71,0%

¹ Support Vector Machines, utilizando núcleo RBF (SVM RBF);

² *k-Nearest Neighbor* (kNN); *Random Forest*;

³ Árvores de decisão, com 10 níveis;

⁴ Árvores de decisão, com 5 níveis;

⁵ *Gaussian Naive Bayes* (Gaussian NB);

⁶ *Gaussian Process*;

⁷ *AdaBoost*.

⁺ Coeficientes: Kcb - coeficiente de cultura basal; Kc - coeficiente de cultura real; Tr - transpiração, em milímetros; e ETc - evapotranspiração da cultura.

^a Acurácia pela divisão do conjunto de treino e teste 50/50 repetido 250 vezes;

^b Acurácia pela validação cruzada 5-fold;

^c Precisão sobre acertos do classificador.

6 CONSIDERAÇÕES FINAIS

Este estudo considerou imagens multi-espectrais de satélite para determinação da transpiração da cultura, com mapeamento espacial para predição dos rendimentos de grãos encontrados durante a colheita. As abordagens usadas neste trabalho, considerando todo o período de cultivo, se mostraram relevantes para a predição de rendimento para a soja. Foram demonstradas diferentes abordagens, baseados em estudos previamente validados, para determinação de características da vegetação como fator principal para o rendimento.

A metodologia proposta para estimativa dos coeficientes de cultura se mostrou eficiente, tanto quanto à sua precisão, quanto a aplicabilidade do experimento, visto que necessita basicamente de dados de fácil acesso, como altura da planta, coletada em campo, imagens de satélite de domínio público e o coeficiente de evaporação fornecido pelo SimDualKc. O modelo foi validado pelas simulações da ferramenta que utiliza a abordagem de Kc Dual, para área irrigada e não irrigada, que é amplamente explorada para modelos de balanço hídrico do solo. Com isso, com este trabalho, obtém-se um ajuste da estimativa hídrica da área, dependente apenas da disponibilidade de imagens de satélite, que considera a distribuição da vegetação espacialmente, estimulando o uso racional da água e energia elétrica dos sistemas de irrigação.

Somado a isso, a estimativa preditiva do rendimento ao final da colheita, possibilita que sejam corrigidas estratégias de irrigação e permite ao produtor o planejamento da estimativa de produção de grãos, tanto em áreas irrigadas quanto em áreas não irrigadas (sequeiro) e consequentemente, sua produtividade¹². A aproximação das alterações de rendimentos, de mesmo modo, contribui para a identificação de áreas que fornecem melhor retorno econômico, dado que a distribuição espacial do rendimento pode ser representada em forma de mapas para posterior análise e correção de deficiências. A abordagem utilizando intervalos de produção se mostrou relevante, uma vez que é possível determinar aumento e quedas de rendimento com relação à mesma área em diferentes pontos a partir da demanda transpirativa da planta. A relação entre transpiração da planta e rendimento de grãos é pouco explorada na comunidade acadêmica, sendo mais evidente apenas diretamente dos índices de vegetação ou de dados diretamente mensurados em campo, como clima ou solo.

Além dos avanços citados, encontrando valores de acurácia superiores a 90%, é possível afirmar que o rendimento de grãos é principalmente impactado pela demanda hídrica, represen-

¹²considera o rendimento e aspectos econômicos da produção

tandos pelos coeficientes (ET_c , Tr , K_{cb} e K_c). Sugere-se em trabalhos futuros adicionar à este conjunto de dados fatores como solo, clima, fertilidade, presença de pragas, doenças, entre outras dentro deste processo biofísico complexo.

A metodologia proposta propiciou uma estimativa da condição da planta em apoio ao manejo de irrigação, contribuindo com o uso racional da água, durante todo o período de cultivo. A frequência temporal e espacial de obtenção de imagens, com satélites privados ou ainda com o uso de veículos aéreos não-tripulados pode vir a melhorar a precisão dos resultados, visto que a representatividade dos valores de rendimento de grãos por planta será maior. Somado a isso, a implicação de mais fatores, como a fertilidade, contribui para complementar os impactos que o manejo agrônomo da área tem sobre a produção de grãos.

Sugere-se ainda, como trabalho futuro, a replicação do experimento em outro ano agrícola, para validação independente, onde um conjunto de dados não tem interferência no outro, para comparação entre mapas de rendimento estimados e os obtidos ao final da safra. As implicações de áreas com menor rendimento podem ser exploradas tomando viés econômico, com a avaliação da produtividade de áreas irrigadas e de sequeiro a partir das predições. Fatores econômicos podem ser abordados também considerando a previsão de produção para cooperativas, estimativas de capacidade de armazenagem, possíveis perdas de produção por estiagens prolongadas ou períodos específicos do desenvolvimento da soja.

A partir de conjuntos com maior amostra de dados, a nível de região, estado ou país, abordagens não apenas com classificação supervisionada, mas também com regressão, podem ser exploradas para determinar os valores de produção, uma vez que o sensoriamento remoto e o aprendizado de máquina se mostram mais vantajosos em grande escala.

REFERÊNCIAS

- ADÃO, T. et al. Hyperspectral Imaging: a review on uav-based sensors, data processing and applications for agriculture and forestry. **Remote Sensing**, [S.l.], v.9, n.11, p.1110, 2017.
- ALLEN, R. G. et al. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. **Irrigation and Drainage**, [S.l.], p.1–15, 1998.
- ALLEN, R. G.; PEREIRA, L. S. Estimating crop coefficients from fraction of ground cover and height. **Irrigation Science**, [S.l.], v.28, n.1, p.17–34, 2009.
- BASTIAANSEN, W. G. et al. Low cost satellite data for monthly irrigation performance monitoring: benchmarks from nilo coelho, brazil. **Irrigation and Drainage Systems**, [S.l.], v.15, n.1, p.53–79, 2001.
- BORDNE, E. F.; MCGUINNESS, J. L. Some procedures for calculating potential evapotranspiration. **Professional Geographer**, [S.l.], v.25, n.1, p.22–28, 1973.
- BRION, J. D.; BALAHADIA, F. F. Application of Remote Sensing and GIS for Climate Change and Agriculture in Philippines. , [S.l.], 2017.
- CAMPOS, I. et al. Combining a water balance model with evapotranspiration measurements to estimate total available soil water in irrigated and rainfed vineyards. **Agricultural Water Management**, [S.l.], v.165, p.141–152, 2016.
- CAMPOS, I. et al. Reflectance-based crop coefficients REDUX: for operational evapotranspiration estimates in the age of high producing hybrid varieties. **Agricultural Water Management**, [S.l.], v.187, p.140–153, 2017.
- CAMPOS, I. et al. Water productivity and crop yield: a simplified remote sensing driven operational approach. **Agricultural and Forest Meteorology**, [S.l.], v.249, n.December 2016, p.501–511, 2018.
- CARLESSO, R.; PETRY, M. T.; TROIS, C. The use of a meteorological station network to provide crop water requirement information for irrigation management. In: IFIP INTERNATIONAL FEDERATION FOR INFORMATION PROCESSING. **Anais...** [S.l.: s.n.], 2009. v.293, p.19–27.

- CHAVES, M. M. et al. How plants cope with water stress in the field. Photosynthesis and growth. **Annals of Botany**, [S.l.], 2002.
- 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.May, p.61–69, 2018.
- DÖLL, P.; SIEBERT, S. Global modeling of irrigation water requirements. **Water Resources Research**, [S.l.], v.38, n.4, p.8–1–8–10, apr 2002.
- EMBRAPA. **Empresa Brasileira de Pesquisa Agropecuária**. Rio de Janeiro: Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos., 2006. 306p.
- ER-RAKI, S. et al. Combining FAO-56 model and ground-based remote sensing to estimate water consumptions of wheat crops in a semi-arid region. **Agricultural Water Management**, [S.l.], v.87, n.1, p.41 – 54, 2007.
- ER-RAKI, S. et al. Determination of crop evapotranspiration of table grapes in a semi-arid region of Northwest Mexico using multi-spectral vegetation index. **Agricultural Water Management**, [S.l.], v.122, p.12–19, 2013.
- European Space Agency. **SENTINEL-2 User Handbook**. [S.l.: s.n.], 2015. (1.2).
- FEHR, W. R. et al. Stage of Development Descriptions for Soybeans, Glycine Max (L.) Merrill. **Crop Science**, [S.l.], v.11, n.6, p.929, 1971.
- FREUND, Y.; SCHAPIRE, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. **Journal of Computer and System Sciences**, [S.l.], 1997.
- GAGO, J. et al. UAVs challenge to assess water stress for sustainable agriculture. **Agricultural Water Management**, [S.l.], v.153, p.9–19, 2015.
- GANDGE, Y.; SANDHYA. A study on various data mining techniques for crop yield prediction. **International Conference on Electrical, Electronics, Communication Computer Technologies and Optimization Techniques, ICEECCOT 2017**, [S.l.], v.2018-Janua, p.420–423, 2018.
- GEIPEL, J.; LINK, J.; CLAUPEIN, W. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. **Remote Sensing**, [S.l.], v.6, n.11, p.10335–10355, 2014.

GENUCHTEN, M. T. **A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils**. 1980.

GEOJSON Specification. Acessado em: 2018-11-11, <https://tools.ietf.org/html/rfc7946>.

GILBERT, N. Water under pressure. **Nature**, [S.l.], v.483, n.7389, p.256–257, 2012.

GONZÁLEZ-DUGO, M. P. et al. Monitoring evapotranspiration of irrigated crops using crop coefficients derived from time series of satellite images. II. Application on basin scale. **Agricultural Water Management**, [S.l.], v.125, p.92–104, 2013.

GONZALEZ-SANCHEZ, A.; FRAUSTO-SOLIS, J.; OJEDA-BUSTAMANTE, W. Predictive ability of machine learning methods for massive crop yield prediction. **Spanish Journal of Agricultural Research**, [S.l.], v.12, n.2, p.313–328, 2014.

GOOGLE Cloud Platform. Acessado em: 2018-10-30, <https://cloud.google.com/>.

GOOGLE Earth Engine. Acessado em: 2018-09-23, https://earthengine.google.com/case_studies/.

GORELICK, N. et al. Remote Sensing of Environment Google Earth Engine : planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, [S.l.], v.202, p.18–27, 2017.

GRATTAN, S. et al. New crop coefficients estimate water use of vegetables, row crops. **California Agriculture**, [S.l.], 1998.

HILLEL, D.; GURON, Y. Relation between evapotranspiration rate and maize yield. **Water Resources Research**, [S.l.], v.9, n.3, p.743–748, 1973.

HOLMAN, D. et al. Gaussian process models for reference ET estimation from alternative meteorological data sources. **Journal of Hydrology**, [S.l.], 2014.

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

HUNT, E. R. et al. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. **Remote Sensing**, [S.l.], v.2, n.1, p.290–305, 2010.

IBGE. Série histórica da estimativa anual da área plantada, área colhida, produção e rendimento médio dos produtos das lavouras. **Levantamento Sistemático da Produção Agrícola, LSPA**, [S.l.], v.Tabela 658, 2019.

IMAGEJ. Acessado em: 2019-09-15, <https://imagej.nih.gov/ij/index.html>.

JENKS, G. The data model concept in statistical mapping. **International Yearbook of Cartography**, [S.l.], 1967.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: a survey. **Computers and Electronics in Agriculture**, [S.l.], v.147, n.July 2017, p.70–90, 2018.

KARAKIZI, C.; OIKONOMOU, M.; KARANTZALOS, K. Spectral discrimination and reflectance properties of various vine varieties from satellite, UAV and proximate sensors. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives**, [S.l.], v.40, n.7W3, p.31–37, 2015.

KOHAVI, R. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE - VOLUME 2, 14., San Francisco, CA, USA. **Proceedings...** Morgan Kaufmann Publishers Inc., 1995. p.1137–1143. (IJCAI'95).

LEGATES, D. R.; MCCABE, G. J. Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. **Water Resources Research**, [S.l.], 1999.

LI, F. et al. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. **European Journal of Agronomy**, [S.l.], v.52, p.198–209, 2014.

LIAKOS, K. G. et al. Machine learning in agriculture: a review. **Sensors (Switzerland)**, [S.l.], v.18, n.8, p.1–29, 2018.

LOPEZ-URREA, R. et al. Water use of spring wheat to raise water productivity. **Agricultural Water Management**, [S.l.], v.96, n.9, p.1305 – 1310, 2009.

LUKAS, V. et al. The combination of UAV survey and Landsat imagery for monitoring of crop vigor in precision agriculture. **International Archives of the Photogrammetry, Remote**

Sensing and Spatial Information Sciences - ISPRS Archives, [S.l.], v.41, n.July, p.953–957, 2016.

MANNERING, J. V.; MEYER, L. D. The Effects of Various Rates of Surface Mulch on Infiltration and Erosion1. **Soil Science Society of America Journal**, [S.l.], v.27, n.1, p.84, 1963.

MARTINS, J. D. et al. Dual crop coefficients for maize in southern Brazil: model testing for sprinkler and drip irrigation and mulched soil. **Biosystems Engineering**, [S.l.], v.115, n.3, p.291–310, 2013.

MATZENAUER, R.; RADIN, B.; ALMEIDA, I. R. de. **Atlas Climático: rio grande do sul**. Porto Alegre: Secretaria da Agricultura Pecuária e Agronegócio; Fundação de Pesquisa Agropecuária (FEPAGRO), 2011.

MISHRA, S.; MISHRA, D.; SANTRA, G. H. **Adaptive boosting of weak regressors for forecasting of crop production considering climatic variability: an empirical assessment**. 2017.

MORIASI, D. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. , [S.l.], 2007.

MUELLER, R. et al. Intercomparison of Soil Moisture, Evaporative Stress, and Vegetation Indices for Estimating Corn and Soybean Yields Over the U.S. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, [S.l.], v.10, n.4, p.1328–1343, 2017.

NASH, J.; SUTCLIFFE, J. River flow forecasting through conceptual models Part 1 - A discussion of principles. **Journal of Hydrology**, [S.l.], 1970.

ODHIAMBO, L. O.; IRMAK, S. Evaluation of the impact of surface residue cover on single and dual crop coefficient for estimating soybean actual evapotranspiration. **Agricultural Water Management**, [S.l.], v.104, p.221–234, 2012.

P. STEDUTO T. C. HSIAO, E. F.; RAES, D. **Crop Yield Response to Water. FAO Irrigation and Drainage Paper 66**. Rome, Italy: [s.n.], 2012.

PAÇO, T. A. et al. Evapotranspiration and crop coefficients for a super intensive olive orchard. An application of SIMDualKc and METRIC models using ground and satellite observations. **Journal of Hydrology**, [S.l.], v.519, n.PB, p.2067–2080, 2014.

PAREDES, P. et al. Partitioning evapotranspiration, yield prediction and economic returns of maize under various irrigation management strategies. **Agricultural Water Management**, [S.l.], v.135, p.27–39, 2014.

PARK, J.; BAIK, J.; CHOI, M. Satellite-based crop coefficient and evapotranspiration using surface soil moisture and vegetation indices in Northeast Asia. **Catena**, [S.l.], v.156, n.July 2016, p.305–314, 2017.

PAUL, M.; VISHWAKARMA, S. K.; VERMA, A. Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND COMMUNICATION NETWORKS, CICN 2015, 2015. **Proceedings...** [S.l.: s.n.], 2016.

PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, [S.l.], v.12, p.2825–2830, 2011.

PEREIRA, L. S. et al. **Crop evapotranspiration estimation with FAO56: past and future**. 2015. 4–20p. v.147.

PEREIRA, L. S. et al. Modeling malt barley water use and evapotranspiration partitioning in two contrasting rainfall years. Assessing AquaCrop and SIMDualKc models. **Agricultural Water Management**, [S.l.], v.159, 2015.

PEREIRA, L. S. et al. **Crop evapotranspiration estimation with FAO56: past and future**. 2015. 4–20p. v.147.

PÔÇAS, I. et al. Estimation of actual crop coefficients using remotely sensed vegetation indices and soil water balance modelled data. **Remote Sensing**, [S.l.], v.7, n.3, p.2373–2400, 2015.

PÔÇAS, I. et al. Hyperspectral-based predictive modelling of grapevine water status in the Portuguese Douro wine region. **International Journal of Applied Earth Observation and Geoinformation**, [S.l.], 2017.

PÔÇAS I, R. A. G. S. C. P. P. L. ASSESSING GRAPEVINE WATER STATUS IN DOURO DEMARCATED REGION USING VEGETATION SPECTRAL MEASUREMENTS. , [S.l.], 2016.

QUANTUM GIS Geographic Information System. Acessado em: 2019-01-22, <https://qgis.org/>.

- ROSA, R. D. et al. Implementing the dual crop coefficient approach in interactive software. 1. Background and computational strategy. **Agricultural Water Management**, [S.l.], v.103, p.8–24, 2012.
- ROSA, R. D. et al. Implementing the dual crop coefficient approach in interactive software: 2. model testing. **Agricultural Water Management**, [S.l.], v.103, p.62–77, 2012.
- ROUSE, J. W. et al. Monitoring Vegetation Systems in the Great Okains with ERTS. **Third Earth Resources Technology Satellite-1 Symposium**, [S.l.], v.1, p.325–333, 1973.
- SAYAGO, S.; OVANDO, G.; BOCCO, M. Landsat images and crop model for evaluating water stress of rainfed soybean. **Remote Sensing of Environment**, [S.l.], v.198, p.30–39, 2017.
- SCHWALBERT, R. A. et al. Forecasting maize yield at field scale based on high-resolution satellite imagery. **Biosystems Engineering**, [S.l.], v.171, p.179–192, 2018.
- SISTEMA Irriga. Acessado em: 2019-04-22, <https://www.sistemairriga.com.br/>.
- TEIXEIRA, A. H. d. C. et al. **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE**. 2013. 6917–6922p. n.1986.
- USDA. **Urban Hydrology for Small Watersheds TR-55**. [S.l.: s.n.], 1986.
- WAHABZADA, M. et al. Plant Phenotyping using Probabilistic Topic Models: uncovering the hyperspectral language of plants. **Scientific Reports**, [S.l.], 2016.
- WRIGHT, J. L. New Evapotranspiration Crop Coefficients. **Proceedings of the American Society of Civil Engineers, Journal of the Irrigation and Drainage Division**, [S.l.], v.108, n.IR2, p.57–74, 1982.
- YADAV, M. K. et al. Impact of canal irrigation on agriculture in Halvad block of Gujarat using remote sensing and GIS technology. , [S.l.], p.28–35, April 2017.
- YOU, J. et al. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In: THIRTY-FIRST AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI-17). **Proceedings...** [S.l.: s.n.], 2017.

ZHANG, H.; ANDERSON, R. G.; WANG, D. Satellite-based crop coefficient and regional water use estimates for Hawaiian sugarcane. **Field Crops Research**, [S.l.], v.180, p.143–154, 2015.

ZHANG, Z.; GONG, Y.; WANG, Z. Accessible remote sensing data based reference evapotranspiration estimation modelling. **Agricultural Water Management**, [S.l.], v.210, n.November 2017, p.59–69, 2018.

Zhao, K. et al. The influence of planting angle on maize based on GIS. , [S.l.], p.1–5, July 2016.