



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA
DOUTORADO EM ENGENHARIA DE TELEINFORMÁTICA

NÍDIA GLÓRIA DA SILVA CAMPOS

SMART&GREEN: UM FRAMEWORK DE INTERNET DAS COISAS PARA
AGRICULTURA INTELIGENTE

FORTALEZA

2020

NÍDIA GLÓRIA DA SILVA CAMPOS

SMART&GREEN: UM FRAMEWORK DE INTERNET DAS COISAS PARA
AGRICULTURA INTELIGENTE

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. Danielo Gonçalves Gomes

Coorientadora: Prof^a. Dr^a Atslands Rego da Rocha

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C215s Campos, Nídia Glória da Silva.
Smart&Green: Um Framework de Internet das Coisas para Agricultura Inteligente / Nídia Glória da Silva Campos. – 2020.
118 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2020.

Orientação: Prof. Dr. Danielo Gonçalves Gomes.

Coorientação: Prof. Dr. Atslands Rego da Rocha.

1. Agricultura de Precisão. 2. Internet das Coisas. 3. Framework. 4. Predição de Umidade do Solo. I. Título.

CDD 621.38

NÍDIA GLÓRIA DA SILVA CAMPOS

SMART&GREEN: UM FRAMEWORK DE INTERNET DAS COISAS PARA
AGRICULTURA INTELIGENTE

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 28 de janeiro de 2020

BANCA EXAMINADORA

Prof. Dr. Danielo Gonçalves Gomes
(Orientador)
Universidade Federal do Ceará (UFC)

Prof^a. Dr^a Atslands Rego da Rocha
(Coorientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Paulo Molin
Universidade de São Paulo (USP)

Prof. Dr. Walmes Marques Zeviani
Universidade Federal do Paraná (UFPR)

Dr. Rubens Sonsol Gondim
Embrapa Agroindústria Tropical

Prof^a. Dr^a Ticiania L. Coelho da Silva
Universidade Federal do Ceará (UFC)

A Deus e aos meus pais.

AGRADECIMENTOS

Agradeço a Deus, Fonte de Sabedoria Eterna, por instruir-me no caminho da verdade. E, também, agradeço à Santa Mãe de Deus, pela Sua intercessão junto ao seu Filho que me concedeu paz, resiliência e alegria nesses anos de trabalho.

Agradeço aos meus pais, Neide e Antônio, pelo amor demonstrado nos mais diversos cuidados e incentivos durante toda minha vida: tudo que sou devo a vocês. Agradeço à Comunidade Salve Rainha pela orações para boa conclusão deste trabalho.

Agradeço aos meus orientadores, professores Danielo e Atslands, por nunca desistirem de mim. MUITÍSSIMO obrigada por terem dividido o seu valioso tempo e conhecimento comigo em favor do meu sólido aperfeiçoamento acadêmico.

Agradeço à Embrapa Agroindústria Tropical pelo estágio e dados do campo experimental. Muito obrigada Rubens pela sua orientação em agricultura de precisão, pelas excelentes sugestões, críticas e esmero nas revisões deste trabalho.

Agradeço à professora Ticiano pela a orientação em modelagem computacional, por toda atenção, paciência e inúmeras revisões deste trabalho.

Agradeço aos professores José Molin e Walmes Zeviani por terem participado da banca examinadora mesmo estando de férias. Muito obrigada pelas sugestões e críticas construtivas.

Agradeço ao CNPq pela aprovação do projeto Smart&Green (nº 432820/2016-7) dentro do qual pude desenvolver minha tese. Agradeço à equipe do projeto Smart&Green pelo suporte técnico neste trabalho. Muito obrigada Airton Filho, Alexandre Costa, Andrei Torres, Guilherme Vieira, Gustavo Fachine, Leandro Machado, Magno Felipe e Sandy Costa.

Agradeço a todos que trabalham no GReat, principalmente, o pessoal da recepção e suporte técnico. Agradeço aos meus colegas de doutorado Daniel, Manoel, Thiago e Rafael pela torcida virtual e presencial, especialmente, nas vésperas da defesa.

Agradeço a todos meus colegas do departamento de Telemática do IFCE, campus Fortaleza, pelo apoio nesses anos de trabalho, em especial, Janaína, Fernando e Régis, que no papel de chefes, sempre me ajudaram na medida do possível. Agradeço também ao professor Paulo Régis pelas preciosas dicas sobre escrita científica.

"Quem semeia entre lágrimas colherá com alegria. Quando vai, vai chorando, levando a semente para plantar; mas quando volta, volta alegre, trazendo seus feixes".

(Salmo 126, 5s)

RESUMO

A irrigação é uma das atividades da agricultura que mais consomem água no mundo e que tem aumentado ao longo do tempo. A escolha do manejo de irrigação depende da disponibilidade dos dados de um campo monitorado. Um sistema da agricultura inteligente coleta dados de várias fontes; no entanto, os dados não estão garantidos de serem livres de valores discrepantes (i.e. *outliers*), que podem prejudicar a precisão da irrigação. Além disso, dados de diferentes fontes devem ser adequados no mesmo intervalo de tempo requerido pelo manejo da irrigação. Nesta tese, o framework Smart&Green é proposto para oferecer serviços de monitoramento, pré-processamento, fusão, sincronização e armazenamento de dados usados nos manejos de irrigação do Turno de Rega, Balanço Hídrico e Potencial Mátrico. Técnicas de remoção de *outliers* permitem uma maior precisão aos manejos da irrigação. Para campos sem sensores de umidade do solo (total ou parcial), um modelo de predição estima o potencial mátrico a partir de informações climáticas, do cultivo e da irrigação. O potencial mátrico é aplicado ao modelo de Van Genuchten para determinar a umidade do solo usada nos manejos do Balanço Hídrico e Potencial Mátrico. Assim, obtém-se uma economia média de 9% a 90% da água da irrigação total necessária quando aplicamos as técnicas de remoção de *outliers* Zscore, MZscore e Chauvenet e as funções Média e Máximo como técnicas de fusão redundante aos dados estimados pelos modelos de predição.

Palavras-chave: Agricultura de Precisão; Internet das Coisas; Framework; Predição de Umidade do Solo

ABSTRACT

Irrigation is one of the most water-intensive agricultural activities in the world, which has been increasing over time. Choosing an optimal irrigation management plan depends on having available data in the monitoring field. A smart agriculture system gathers data from several sources; however, the data are not guaranteed to be free of discrepant values (i.e., outliers), which can damage the precision of irrigation management. Furthermore, data from different sources must fit into the same time interval required for irrigation management and the data preprocessing must be dynamic and automatic to benefit users of the irrigation management plan. In this paper, we propose the Smart&Green framework to offer services for smart agriculture, such as data monitoring, preprocessing, fusion, synchronization, storage, and irrigation management of 'Turno de Rega', Water Balance, and Matric Potential. Outlier removal techniques allow for more precise irrigation management. For fields without soil moisture sensors, the prediction model estimates the matric potential using weather, crop, and irrigation information. We apply the predicted matric potential approach to the Van Genuchten model to determine the moisture used in an irrigation management scheme of water balance or matric potential. We can save, on average, between 9% and 90% of the irrigation water needed by applying to the predicted data the Zscore, MZscore, and Chauvenet outlier removal techniques and the functions Mean and Maximum as redundant fusion techniques.

Keywords: Smart agriculture; Internet of Things; framework; Soil Moisture Prediction

LISTA DE FIGURAS

Figura 1 – Propriedade rural com diferentes tipos de campos. Fonte: Autora.	20
Figura 2 – Visão Geral do Smart&Green. Fonte: Autora	28
Figura 3 – Exemplos de integração de fontes de dados usados pelos manejos de irrigação. Fonte: Autora.	45
Figura 4 – Seleção de manejos de irrigação. Fonte: Autora.	47
Figura 5 – Smart&Green - modelo de previsão de umidade do solo como um sistema de suporte a decisão (DSS) da irrigação. Fonte: Autora.	49
Figura 6 – Interação entre as camadas de Aplicação e Serviços. Fonte: Autora	50
Figura 7 – Principais Funções do Serviço de Fusão de Dados do Solo. Fonte: Autora.	52
Figura 8 – Detalhes do campo experimental pertencente à Embrapa Agroindústria Tropical (EMBRAPA, 2019) – (a) tensiômetros a 15, 45 e 75 cm de profundidade do solo perto de uma muda de cajueiro e (b) mudas de caju.	57
Figura 9 – Análise da relevância de atributos – abordagem de previsão (a) Local e (b) Global.	65
Figura 10 – Dados reais de potencial mátrico ψ_c resultantes da Fusão Redundante Mínimo e usados no manejo do Potencial Mátrico	71
Figura 11 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da Fusão Redundante Mínimo — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico	71
Figura 12 – Dados reais de potencial mátrico ψ_c resultantes da Fusão Redundante Máximo e usados no manejo do Potencial Mátrico	72
Figura 13 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da Fusão Redundante Máximo — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico	72
Figura 14 – Dados reais de potencial mátrico ψ_c resultantes da Fusão Redundante Mediana e usados no manejo do Potencial Mátrico	73
Figura 15 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da Fusão Redundante Mediana — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico	73
Figura 16 – Dados reais de potencial mátrico ψ_c resultantes da Fusão Redundante Média e usados no manejo do Potencial Mátrico	74

Figura 17 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da Fusão Redundante Média — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico .	74
Figura 18 – Dados estimados de potencial mátrico ψ_p da abordagem de predição Local resultantes da Fusão Redundante Média e usados no manejo do Potencial Mátrico	77
Figura 19 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição Local e fusionados pela Média — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico	77
Figura 20 – Dados estimados de potencial mátrico ψ_p da abordagem de predição Local resultantes da Fusão Redundante Mediana e usados no manejo do Potencial Mátrico	78
Figura 21 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição Local e fusionados pela Mediana — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico	78
Figura 22 – Dados estimados de potencial mátrico ψ_p da abordagem de predição Local resultantes da Fusão Redundante Máximo e usados no manejo do Potencial Mátrico	79
Figura 23 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição Local e fusionados pelo Máximo — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico .	79
Figura 24 – Dados estimados de potencial mátrico ψ_p da abordagem de predição Local resultantes da Fusão Redundante Mínimo e usados no manejo do Potencial Mátrico	80
Figura 25 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição Local e fusionados pelo Mínimo — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico .	80
Figura 26 – Dados estimados de potencial mátrico ψ_p da abordagem de predição Global usados no manejo do Potencial Mátrico	81
Figura 27 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição Global — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico	81

Figura 28 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da Fusão Média . . .	84
Figura 29 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da Fusão Mínimo . . .	84
Figura 30 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da Fusão Mediana . . .	84
Figura 31 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da Fusão Máximo . . .	84
Figura 32 – Balanço Hídrico – Umidade do solo real $\theta_c(\psi_c)$ resultante da Fusão Média . . .	85
Figura 33 – Balanço Hídrico – Umidade do solo real $\theta_c(\psi_c)$ resultante da Fusão Mínimo . . .	85
Figura 34 – Balanço Hídrico – Umidade do solo seal $\theta_c(\psi_c)$ resultante da Fusão Mediana . . .	85
Figura 35 – Balanço Hídrico – Umidade do solo real $\theta_c(\psi_c)$ resultante da Fusão Máximo . . .	85
Figura 36 – Balanço Hídrico: ψ_p estimado pela abordagem Local e fusionado pela Média . . .	88
Figura 37 – Balanço Hídrico: ψ_p estimado pela abordagem Local e fusionado pelo Mínimo . . .	88
Figura 38 – Balanço Hídrico: ψ_p estimado pela abordagem Local e fusionado pela Mediana . . .	88
Figura 39 – Balanço Hídrico: ψ_p estimado pela abordagem Local e fusionado pelo Máximo . . .	88
Figura 40 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem Local e fusão Média	89
Figura 41 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem Local e fusão Mínimo	89
Figura 42 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem Local e fusão Mediana	89
Figura 43 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem Local e fusão Máximo	89
Figura 44 – Balanço Hídrico – Dados de potencial mátrico ψ_p estimado pela abordagem Global	91
Figura 45 – Balanço Hídrico – Umidade do solo θ_p deduzida com dados ψ_p estimados pela abordagem Global	91
Figura 46 – Curva do coeficiente de cultivo (Kc) do coqueiro anão-verde. Fonte: (MI- RANDA <i>et al.</i> , 2007)	106
Figura 47 – Curva característica de solo franco-arenoso. Fonte: (SILVA, 2017)	107

LISTA DE TABELAS

Tabela 1 – Palavras usadas na busca (<i>search query</i>). Fonte: Autora.	31
Tabela 2 – Algoritmo da Busca: (Grupo 1) AND (Grupo 2) AND (Grupo 3) AND (Grupo 4) AND (Grupo 5). Fonte: Autora.	31
Tabela 3 – Forma usada para extrair dados de cada estudo. Fonte: (Autora).	32
Tabela 4 – Os dados usados pelos manejos de irrigação são providos por várias fontes de um campo monitorado. Fonte: Autora.	44
Tabela 5 – Informações usadas para a escolha dos manejos da irrigação	46
Tabela 6 – Estrutura do Conjunto de Dados do Campo Experimental	57
Tabela 7 – Contantes de Van Genutchen do Solo do Campo Experimental	58
Tabela 8 – Grupos de Atributos	58
Tabela 9 – Grupo de Atributos Seleccionados	59
Tabela 10 – Abordagem Local, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 1 – média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	62
Tabela 11 – Abordagem Local, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 2 – média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	62
Tabela 12 – Abordagem Local, Avaliação dos Algoritmos para Predição de Potencial Mátrico ψ_p quando usam os Atributos do Grupo 3 – média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	62
Tabela 13 – Abordagem Local, Avaliação dos Algoritmos para Predição de Potencial Mátrico ψ_p quando usam os Atributos do Grupo 4– média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	63
Tabela 14 – Abordagem Global, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 1 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	63

Tabela 15 – Abordagem Global, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 2 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	64
Tabela 16 – Abordagem Global, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 3 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	64
Tabela 17 – Abordagem Global, Avaliação dos Algoritmos de Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 4 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em negrito	64
Tabela 18 – ITN fornecida pelo manejo do Potencial Mátrico usando dados reais $\theta_c(\psi_c)$ de umidade do solo processados pelo Serviço de Fusão.	70
Tabela 19 – ITN fornecida pelo manejo do Potencial Mátrico usando dados estimados $\theta_p(\psi_p)$ de umidade do solo e processados pelo Serviço de Fusão.	76
Tabela 20 – Equivalência de desempenho entre o manejo do Potencial Mátrico que usa dados reais $\theta_c(\psi_c)$ e o mesmo que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição — Métrica: ITN média diferente em até ± 2 mm/dia.	81
Tabela 21 – Economia de ITN do manejo do Potencial Mátrico que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição de umidade do solo com desempenho equivalente ao mesmo manejo que usa dados reais $\theta_c(\psi_c)$	82
Tabela 22 – ITN média obtida pelo Balanço Hídrico usando dados reais $\theta_c(\psi_c)$ de umidade do solo processados pelo Serviço de Fusão e economia de ITN em comparação ao Balanço Hídrico que não usa umidade do solo	83
Tabela 23 – ITN média fornecida pelo Balanço Hídrico usando dados de umidade do solo estimados $\theta_p(\psi_p)$ pelos Modelos de Predição e processados pelo Serviço de Fusão. A economia de ITN é obtida comparando-se ao Balanço Hídrico que não usa umidade do solo	87
Tabela 24 – Equivalência de Desempenho entre o manejo do Balanço Hídrico que usa dados reais $\theta_c(\psi_c)$ e o mesmo que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição — Métrica: Diferença de ITN de até ± 2 mm/dia	90

Tabela 25 – Economia de ITN do manejo do Balanço Hídrico que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição de umidade do solo com desempenho equivalente ao mesmo manejo que usa dados reais $\theta_p(\psi_p)$ 90

SUMÁRIO

1	INTRODUÇÃO	18
1.1	Contextualização	18
1.2	Motivação	19
1.2.1	<i>Problema 1: Integração de diferentes fontes de dados usados nos manejos de irrigação</i>	20
1.2.2	<i>Problema 2 - O tratamento de outliers nos dados de umidade do solo não pode prejudicar o tradeoff entre a economia de ITN e o desenvolvimento saudável do cultivo</i>	22
1.2.3	<i>Problema 3 - Os dados de umidade do solo deixam o manejo de irrigação mais exato, no entanto, não estão sempre disponíveis</i>	23
1.3	Questões de Pesquisa e Hipóteses	24
1.4	Objetivos	25
1.5	Metodologia de Pesquisa	26
1.6	Contribuição	27
1.7	Organização da Tese	28
2	REVISÃO DA LITERATURA	30
2.1	Mapeamento Sistemático da Literatura - Planejamento	30
2.2	Mapeamento Sistemático da Literatura - Resultado	33
2.2.1	<i>Coleta de dados</i>	34
2.2.2	<i>Preprocessamento de dados</i>	34
2.2.3	<i>Manejos de Irrigação</i>	36
2.2.4	<i>Predição de umidade dos solo</i>	37
2.2.5	<i>Plataformas IoT para Agricultura Inteligente</i>	38
2.3	Síntese do Capítulo	39
3	PROPOSTA	40
3.1	Visão Geral do <i>framework</i> Smart&Green	40
3.2	Camada de Aplicação	41
3.2.1	<i>Automatização dos Manejos de Irrigação</i>	41
3.2.2	<i>Predição de Umidade do Solo</i>	48
3.3	Camada de Serviços	49

3.3.1	<i>Armazenamento</i>	49
3.3.2	<i>Sincronização de dados</i>	49
3.3.3	<i>Fusão de dados de umidade do solo</i>	51
3.4	Camada de Comunicação	53
3.5	Síntese do Capítulo	54
4	AVALIAÇÃO DOS MODELOS COMPUTACIONAIS PARA PREDIÇÃO DE UMIDADE DO SOLO	56
4.1	Materiais e Métodos para Avaliação dos Modelos Computacionais	56
4.1.1	<i>Conjunto de Dados do Campo Experimental da Embrapa</i>	56
4.1.2	<i>Métodos de Preprocessamento dos Dados</i>	58
4.1.3	<i>Métodos de Aprendizagem de Máquina.</i>	59
4.2	Resultados da Avaliação dos Modelos Computacionais	61
4.2.1	<i>Avaliação do Modelo da Abordagem de Predição Local</i>	61
4.2.2	<i>Avaliação do Modelo da Abordagem de Predição Global</i>	63
4.2.3	<i>Relevância de atributos</i>	64
4.3	Síntese do Capítulo	65
5	AVALIAÇÃO DOS MANEJOS DE IRRIGAÇÃO DO SMART&GREEN	67
5.1	Implementação do framework Smart&Green	67
5.2	Materiais e Métodos da Avaliação de Manejos de Irrigação	68
5.3	Resultados da Avaliação do manejo do Potencial Mátrico	69
5.3.1	<i>Avaliação do manejo do Potencial Mátrico que usa dados reais θ_c de umidade do solo</i>	70
5.3.1.1	<i>Impacto da Fusão Redundante no manejo do Potencial Mátrico</i>	70
5.3.1.2	<i>Impacto da Remoção de Outliers no manejo do Potencial Mátrico</i>	75
5.3.2	<i>Avaliação do manejo do Potencial Mátrico que usa dados estimados θ_p de umidade do solo</i>	75
5.4	Resultados da Avaliação do manejo do Balanço Hídrico	82
5.4.1	<i>Avaliação do manejo do Balanço Hídrico que usa dados reais θ_c de umidade do solo</i>	83
5.4.2	<i>Avaliação do manejo do Balanço Hídrico que usa dados estimados θ_p de umidade do solo</i>	86
5.5	Síntese do Capítulo	91

6	CONCLUSÃO	94
6.1	Respostas às Questões de Pesquisa (QP) e às respectivas hipóteses	94
6.1.1	<i>QP#1 - confirmação da Hipótese #1</i>	94
6.1.2	<i>QP#2 - confirmação parcial da Hipótese #2</i>	94
6.1.3	<i>QP#3 - confirmação da Hipótese #3</i>	95
6.2	Publicações	95
6.3	Trabalhos Futuros	96
	REFERÊNCIAS	97
	APÊNDICES	105
	APÊNDICE A–MANEJOS DA IRRIGAÇÃO	105
A.1	Dados climáticos, do cultivo e do sistema de irrigação	105
A.2	Dados do solo	106
A.3	Métodos para manejo da irrigação	109
A.3.1	<i>Forma de Distribuição de Água</i>	109
A.3.2	<i>Método do Turno de Rega Calculado</i>	109
A.3.3	<i>Método do Balanço Hídrico</i>	111
A.3.4	<i>Método do Potencial Mátrico</i>	112
	APÊNDICE B–PROBLEMA: PREDIÇÃO DE UMIDADE DO SOLO .	114
	APÊNDICE C–BANCO DE DADOS DO SMART&GREEN	116
C.1	Regras de Negócio Principais do Smart&Green	116
C.2	Diagrama de Entidade e Relacionamento	118
C.3	Dicionário de dados	118

1 INTRODUÇÃO

Esta tese de doutorado propõe o Smart & Green, um *framework* computacional baseado no paradigma de Internet das Coisas (IoT) para aplicação na agricultura e que provê suporte inteligente à atividade de irrigação. Este capítulo de Introdução contém a motivação e a descrição do problema tratado, hipóteses, questões de pesquisa, objetivos, metodologia da pesquisa, escopo, contribuições e estrutura do manuscrito de tese.

1.1 Contextualização

A agricultura já usa 70% de água doce retiradas no mundo (FAO, 2014b). A irrigação fornece artificialmente água para suprir necessidades específicas de recursos hídricos de uma cultura nos diferentes estágios de seu desenvolvimento em um determinado local (TESTEZLAF, 2017). No Brasil, em 2015, a irrigação consumiu cerca de 46% do total de retiradas de água, i.e. 969 mil litros por segundo (l/s). Em 2030, estudos prevêem um aumento de 38% desse consumo, ou seja, mais de 1.338 milhão l/s (ANA, 2017). Logo, é fundamental a atuação do manejo de irrigação, uma das atividades da agricultura de precisão, especialmente, nas regiões semiáridas e com escassez de chuvas.

Dentre os vários conceitos para a agricultura de precisão (AP), destacamos o da gestão dos cultivos que busca o máximo rendimento das culturas com o mínimo consumo de insumos (MOLIN *et al.*, 2015), isto é, água, fertilizantes, pesticidas, entre outros. A origem do termo “agricultura de precisão” tem a ver com a variabilidade espaço-temporal das lavouras e remonta ao início dos anos 1990 com uso civil do Sistema de Posicionamento Global (GPS) (MOLIN *et al.*, 2015).

Entretanto, se no início dos anos 1990 a agricultura de precisão (AP) estava fortemente associada tão-somente ao uso do GPS nas lavouras, atualmente o entendimento é bem mais abrangente. AP tem a ver por exemplo com a análise de dados do cultivo de vegetais, do solo e do clima em um campo (já que variam no espaço e no tempo) com o objetivo de gerenciar/otimizar a quantidade de recursos/insumos empregados visando o aumento do retorno financeiro, a sustentabilidade e a minimização de efeitos nocivos ao meio ambiente (BERNADI *et al.*, 2014). A AP refere-se a uma “*agricultura com exatidão maior do que aquela que já é praticada*” (MOLIN *et al.*, 2015). Por exemplo, a aplicação exata de pesticidas reduz a poluição e aumenta a eficiência dos fertilizantes (SUNDMAEKER *et al.*, 2016).

Um conceito de agricultura mais recente e dentro do qual insere-se esta tese é o da agricultura inteligente (*smart farming, smart agriculture*) ao qual baseia-se em tarefas de gerenciamento (não só da localização geográfica, mas também dos dados), enriquecidas pela sensibilidade ao contexto e inicializadas por eventos em tempo real. A agricultura inteligente é habilitada pelos paradigmas da Internet das Coisas (*Internet of Things, IoT*) e da computação em nuvem (*Cloud Computing*) (WOLFERT *et al.*, 2017).

O paradigma da IoT, por sua vez, está baseado na comunicação contínua entre os chamados objetos inteligentes, os quais podem ser entendidos como sistemas embarcados com capacidade de sensoriamento e de conexão em rede (normalmente sem fio). Neste contexto de IoT, uma “coisa” significa um objeto real/físico ou mesmo virtual/digital, o qual pode ser monitorado (lido, reconhecido, localizado) e eventualmente controlado, tudo isso remotamente via Internet (BORGIA, 2014).

As coisas ou objetos ou dispositivos inteligentes (*smart devices*) estendem as ferramentas convencionais (pluviômetros, tratores, notebooks), visto que, a coleta de dados em tempo real, a ciência de contexto e a inteligência computacional as tornam capazes de analisar dados da cadeia produtiva transformando-os em conhecimento essencial à tomada de decisão importantes para agricultura (ISSAD *et al.*, 2019). Desta maneira, os objetos inteligentes podem executar ações autônomas ou fazê-las remotamente, tais como o monitoramento de recursos naturais (e.g. água, solo, ar), a predição de condições climáticas/solo, de produção, de crescimento das culturas e a logística (produção, comércio e transporte) (TALAVERA *et al.*, 2017), cabendo ao agricultor acompanhar, monitorar e endossar os processos em curso.

1.2 Motivação

Os manejos de irrigação controlam a quantidade de água e a frequência da irrigação através do monitoramento da umidade do solo¹ e da evapotranspiração² do cultivo (ANA, 2017). Esta tese aborda a implementação dos seguintes manejos de irrigação: Turno de Rega, Balanço Hídrico e Potencial Mátrico (detalhes no Apêndice A.3), os quais consomem dados do clima, do cultivo, do sistema de irrigação (Apêndice A.1) e do solo (Apêndice A.2). Nesta seção, destacamos três problemas relacionados aos serviços considerados importantes no suporte dos manejos de irrigação: integração de diferentes fontes de dados (Seção 1.2.1), tratamento de

¹ nível de água retida no solo

² consumo de água da cultura de acordo com seu estágio de desenvolvimento e as condições climáticas

outliers (Seção 1.2.2) e predição da umidade do solo (Seção 1.2.3).

1.2.1 Problema 1: Integração de diferentes fontes de dados usados nos manejos de irrigação

Os campos agrícolas considerados nesta tese possuem um tipo de cultivo, solo e sistema de irrigação conforme a ilustração da Figura 1. A disponibilidade de dados do clima, cultivo, sistema de irrigação (Apêndice A.1) e solo (Apêndice A.2) de um campo determina a escolha do manejo de irrigação. Por exemplo, os três manejos de irrigação tratados nesta tese usam necessariamente dados do cultivo e do sistema de irrigação, mas, apenas o Campo 2 ou 3 da Figura 1 possuem fontes dados de umidade do solo, fundamentais para execução do manejo do Potencial Mátrico, o qual não pode ser aplicado no Campo 1.

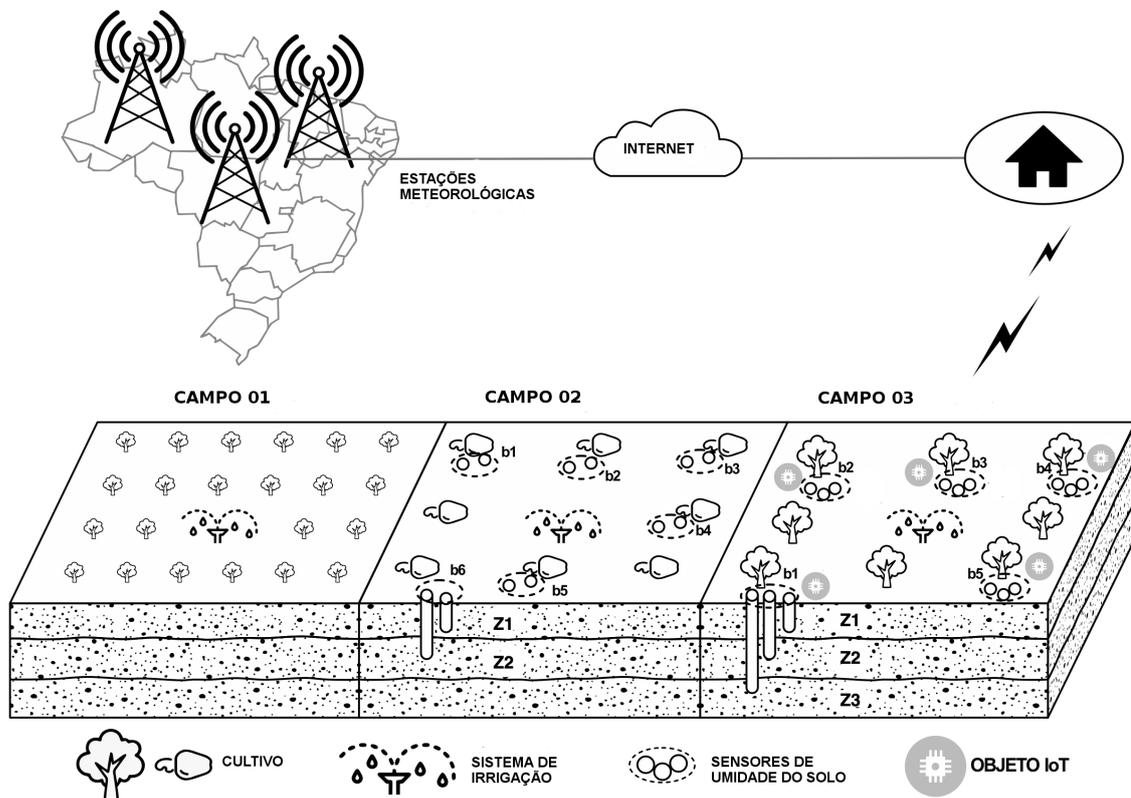


Figura 1 – Propriedade rural com diferentes tipos de campos. Fonte: Autora.

Quanto mais dados disponíveis, mais consistente será o diagnóstico da variabilidade do campo (MOLIN *et al.*, 2015). No contexto da irrigação, essa variabilidade está relacionada à necessidade hídrica do cultivo, medida pela irrigação total necessária (ITN), provida pelos manejos de irrigação. Se comparamos os manejos do Turno de Rega e do Balanço Hídrico, já que ambos precisam dos mesmos tipos de dados, podemos afirmar que o do Balanço Hídrico

fornece uma ITN mais consistente, pois, usa a média diária dos dados (Eq. A.7), enquanto, que o do Turno de Rega usa a média mensal histórica, podendo provocar déficit ou excesso de água ao cultivo (Apêndice A.3.2).

As fontes de dados necessários aos manejos de irrigação são heterogêneas. No Brasil, os dados climáticos são coletados por estações meteorológicas automáticas e disponibilizados em páginas web na Internet (INMET, 2018). Os sensores analógicos, presentes em campos como o Campo 2 da Figura 1, fornecem dados de umidade do solo os quais devem ser coletados/lidos por qualquer pessoa e armazenados por exemplo em planilhas eletrônicas. Uma outra alternativa de instrumentação de campo é a do Campo 3 da Figura 1, em que nós sensores e atuadores podem ser configurados como objetos inteligentes capazes de interagir entre si através protocolos de comunicação da IoT (ex. CoAP, MQTT, Zigbee, LoRa, entre outros).

Agrônomos são considerados fontes de dados, pois, podem fornecer informações específicas sobre vários tipos de cultivo, características do solo e os critérios de detecção e remoção imediata de *outliers* para o planejamento dos manejos de irrigação. *Outliers* são dados fora da faixa de valores possíveis ou valores incoerentes com outros valores dentro da mesma série de dados. Assim, os critérios de detecção e remoção imediata devem ser associados a um tipo de dado climático e/ou umidade do solo para tratar valores que estejam fora de um determinado intervalo esperado.

Por exemplo, um critério para dados de potencial mátrico³ pode ser definido de acordo com um intervalo fora dos valores possíveis que são detectados pelo tipo de sensor instalado no campo (TORRES, 2017). Enquanto critérios relacionados aos dados climáticos podem ser definidos com uma faixa de valores incomuns à região geográfica de onde são coletados (CAMINHA *et al.*, 2017).

A fusão de dados ajuda na construção de conhecimento sobre certos eventos e ambientes os quais não é possível usando sensores individuais separadamente (WANG *et al.*, 2016). Em outras palavras, pode-se dizer que os métodos de manejos da irrigação realizam a fusão cooperativa (ELMENREICH, 2002 apud NAKAMURA *et al.*, 2007), pois, esses métodos usam diferentes informações (clima, solo, cultivo) providas por diferentes fontes (estações climáticas, sensores de solo, agrônomos) para determinar a informação da ITN.

Portanto, é fundamental a integração dessas várias fontes de dados, para que seja escolhido o método de manejo de irrigação mais adequado para um campo e os dados necessários

³ o potencial mátrico ψ é usado para deduzir a umidade do solo $\theta(\psi)$ de uma camada de solo monitorada por tensiômetros (Apêndice A.2) instalados em diferentes locais de um campo (ex. Campo 2 ou 3 da Figura 1).

sejam disponibilizados de maneira sincronizada/alinhada de acordo com o método escolhido. Por exemplo, o método do Turno de Rega (Seção A.3.2) calcula a água facilmente disponível (Eq. A.3) de acordo com a fase de desenvolvimento do cultivo determinado pelo coeficiente k_c (Eq. A.4). Assim, k_c deve estar alinhado com a idade (quantidade de dias) do cultivo. No método do Balanço Hídrico (Seção A.3.3), quando fontes de dados da umidade do solo estão disponíveis (Campo 2 e 3 da Figura 1), é necessário sincronizar, adequar na mesma janela temporal (intervalo de tempo) de coleta, os dados climáticos fornecidos pelas estações meteorológicas. Isto porque dados climáticos podem ser amostrados de hora em hora e dados do solo, diariamente (Campo 2, Figura 1) ou em um intervalo de tempo menor (Campo 3, Figura 1).

1.2.2 Problema 2 - O tratamento de outliers nos dados de umidade do solo não pode prejudicar o tradeoff entre a economia de ITN e o desenvolvimento saudável do cultivo

Sempre haverão questionamentos, especialmente em relação às técnicas e tecnologias empregadas na agricultura, pois, elas podem ser fontes de incertezas no processamento de dados (MOLIN *et al.*, 2015). Sensores são as principais fontes de dados do clima e do solo necessários aos manejos de irrigação. Eles podem estar mal calibrados ou apresentar falhas que adicionem imperfeições, ambiguidades ou inconsistências aos dados gerados, levando-os a valores discrepantes que podem prejudicar a exatidão da irrigação total necessária (ITN). Portanto, o pré-processamento de dados é um serviço necessário aos manejos de irrigação para tratar *outliers*.

Enquanto os critérios de remoção imediata de *outliers* tratam os dados de maneira individual, as técnicas de remoção de *outliers* processam uma série de dados de umidade do solo $\theta(\psi)$ provenientes de vários sensores de um mesmo tipo instalados em diferentes pontos/locais de um campo (como os sensores de solo dos Campos 2 e 3 da Figura 1). Em Torres (2017), existem recomendações de técnicas eficientes, por exemplo, a técnica MZscore (*Modified Zscore*) trata o potencial mátrico ψ para remoção de *outlier* quando $\psi \in [10; 100]$ kPa. A técnica GESD (*Generalized Extreme Studentized Deviate*) é mais adequada para tratar *outliers* quando $\psi \cong 1000$ kPa .

A fusão de dados é outro meio de tratar *outliers*, pois, consiste em um método que combina dados de vários sensores para produzir informação mais acurada, mais completa, mais confiável do que seria possível alcançar com um único sensor (HALL; LLINAS, 1997). Em Torres (2017), usou a média aritmética como técnica de fusão redundante (ELMENREICH, 2002

apud NAKAMURA *et al.*, 2007) para agregar uma série de dados de potencial mátrico ψ de uma mesma camada de solo monitorada em um campo IoT (ex. Campo 3, Figura 1). O dado ψ_f resultante é usado para se obter a umidade do solo $\theta(\psi_f)$ necessária aos manejos de irrigação.

No entanto, quando se usa uma técnica de remoção de *outliers* em conjunto com uma técnica de fusão redundante no pré-processamento de dados de umidade do solo, a ITN pode ser aumentada ou diminuída dependendo do método de manejo de irrigação (Potencial Mátrico ou Balanço Hídrico) adotado para o campo. Tal situação pode prejudicar a relação custo benefício desejada entre a economia de ITN e o desenvolvimento saudável do cultivo. Isto é, o uso de tais técnicas em conjunto pode reduzir drasticamente a ITN, induzindo o déficit de água ao cultivo e prejudicar o desenvolvimento do cultivo em consequência. Ou, por outro lado, o uso dessas técnicas pode gerar uma economia insignificante ou não alterar em nada a ITN fornecida por um manejo de irrigação, ou ainda pode provocar o fornecimento de água em excesso ao cultivo, e, assim, agravar o desperdício de água no mundo.

1.2.3 Problema 3 - Os dados de umidade do solo deixam o manejo de irrigação mais exato, no entanto, não estão sempre disponíveis

O solo funciona como um reservatório de água que também supre as necessidades hídricas do cultivo (Apêndice A.2). Assim, quando os dados de umidade do solo $\theta(\psi)$ são acrescentados ao manejo de irrigação do Balanço Hídrico, a irrigação total necessária (ITN) é reduzida em média de 76,5% a 90,4% (Seção 5.4.1). Porém, nem todo proprietário rural possui orçamento para adquirir sensores de solo (ex. tensiômetros) e, assim, obter dados $\theta(\psi)$ para otimizar a ITN. Além dos cenários de campos totalmente sem sensores (ex. Campo 1 da Figura 1), há também casos em que alguns dos sensores instalados em um campo (ex. Campo 2 ou 3 da Figura 1) podem apresentar defeitos e, por isso, serem isolados, deixando o campo parcialmente sem monitoramento. Como também, problemas de conectividade podem impedir que um campo envie os dados $\theta(\psi)$ a tempo para a execução do manejo de irrigação.

Partindo do pressuposto que exista um conjunto de dados histórico do clima, do cultivo, da irrigação e do solo coletados diariamente em um campo (ex. Campo 2 ou 3 da Figura 1), métodos de aprendizagem de máquina podem ajudar na análise da associação entre a umidade do solo θ e os outros dados usados nos manejos de irrigação. O Apêndice B formaliza esse problema definindo tipos de dados que podem ser aplicados a uma função f de previsão do potencial mátrico ψ atual para, conseqüentemente, obter a umidade do solo $\theta_p(\psi_p)$ atual para

campos sem sensores de solo (ex. total como Campo 1 da Figura 1 ou parcial).

1.3 Questões de Pesquisa e Hipóteses

A partir dos três problemas levantados na Seção 1.2, seguem as questões de pesquisa desta tese e suas respectivas hipóteses:

1. **Questão de Pesquisa #1 (QP#1):** Sabe-se que a escolha do manejo de irrigação mais adequado (Turno de Rega, Balanço Hídrico e Potencial Mátrico) a um campo (Figura 1) depende da disponibilidade de vários tipos de dados (clima, cultivo, solo, irrigação) provenientes de diferentes fontes (estações meteorológicas, agrônomos, proprietários rurais, planilhas eletrônicas, sensores analógicos, objetos da Internet das Coisas - IoT). Como a agricultura inteligente pode integrar essas diferentes fontes de dados e, assim, fornecer dados: 1) sincronizados/alinhados de acordo com a especificação de cada manejo de irrigação; e 2) preprocessados com tratamento de remoção de *outliers* adequado ao tipo de sensor de umidade do solo e aos valores incomuns relacionados ao clima da região do campo monitorado?
 - **Hipótese #1:** *A agricultura inteligente fornece objetos inteligentes virtuais/digitais com conhecimento sobre os principais tipos de dados usados nos métodos de manejos de irrigação. Esses objetos inteligentes são capazes de 1) interagir com as diferentes fontes de dados associadas a um campo monitorado, 2) analisar os tipos de dados disponíveis e, conseqüentemente, 3) escolher o manejo de irrigação mais adequado ao campo. Tais objetos promovem a coleta, a sincronização e a remoção de outliers de dados para execução em tempo real dos manejos de irrigação do campo monitorado.*
2. **Questão de Pesquisa #2 (QP#2):** Os dados de umidade do solo, provenientes de várias camadas monitoradas em diferentes pontos/locais de um campo (ex. Campo 2 ou 3 da Figura 1), não estão isentos de *outliers* que podem interferir na exatidão da irrigação total necessária (ITN) fornecida pelos manejos de irrigação do Potencial Mátrico e Balanço Hídrico (Apêndice A). Então, como a agricultura inteligente realiza o serviço de preprocessamento em tempo real, customizado para a infraestrutura atual de um campo, levando em consideração 1) o reuso desse serviço por diferentes manejos de irrigação e 2) sem prejudicar a relação custo-benefício entre a economia de ITN e o desenvolvimento saudável do cultivo?

- **Hipótese #2:** *A agricultura inteligente fornece o serviço de pré-processamento de dados de umidade do solo baseado em técnicas de remoção de outliers e fusão redundante de dados genéricas o suficiente para serem reaproveitadas pelos diferentes métodos de manejos de irrigação. No entanto, esse serviço tem conhecimento sobre como o uso em conjunto dessas técnicas interferem na ITN provida pelo método de manejo da irrigação a ser executado. Assim, ele escolhe uma técnica de remoção de outlier e outra de fusão redundante de dados de maneira que seja garantida uma boa relação custo-benefício entre a economia de água (ITN) e o desenvolvimento saudável do cultivo.*
3. **Questão de Pesquisa #3 (QP#3)** Dados dois campos com mesmo tipo de solo: um campo sem sensores de umidade do solo (ex. Campo 1 da Figura 1) e outro com esses sensores (ex. Campo 2 ou Campo 3 da Figura 1), como a agricultura inteligente estima a umidade do solo $\theta_p(\psi_p)$ atual para o campo sem sensores usando um conjunto de dados históricos do clima, do solo, do cultivo e da irrigação coletados diariamente do campo com sensores?
- **Hipótese #3:** *Na agricultura inteligente, as técnicas de aprendizagem de máquina amplamente empregadas na solução de problemas de regressão são usadas para a análise da influência de dados do clima, do cultivo e da irrigação sobre os dados de umidade do solo. Logo, um serviço da IoT de previsão de umidade do solo possui uma função f relacionada a uma técnica de aprendizagem de máquina que o habilita a estimar dados $\theta_p(\psi_p)$ sem o uso de sensores. Assim, os manejos de irrigação que usam dados $\theta_p(\psi_p)$, obtêm desempenho semelhante àquele de quando utilizam dados de umidade do solo reais fornecidos por sensores de solo.*

1.4 Objetivos

O objetivo geral desta tese é contribuir no suporte à tomada de decisão no nível dos serviços de monitoramento, sincronização, pré-processamento e fusão de dados necessários no contexto da agricultura inteligente, de modo que seja possível a escolha e execução em tempo real dos manejos de irrigação (Apêndice A) adequados a diferentes campos agrícolas (Figura 1). Para atingir esse objetivo geral, traçamos os seguintes objetivos específicos:

1. Integrar as diferentes fontes de dados usados nos manejos de irrigação, como as estações meteorológicas brasileiras (INMET, 2018), sensores de umidade do solo (tensiômetros analógicos e digitais) e profissionais com especialidade agrônoma;

2. Prover a escolha do manejo de irrigação mais adequado para diferentes tipos de campos (Figura 1) de maneira transparente aos seus usuários;
3. Sincronizar os dados do clima, solo e cultivo de acordo com o método do manejo de irrigação escolhido para diferentes tipos de campos;
4. Auxiliar na definição de critérios de remoção imediata de *outliers* para dados do clima e umidade do solo de acordo com a infraestrutura atual e clima de diferentes tipos de campos;
5. Investigar a interferência de técnicas de remoção de *outliers* e fusão redundante aplicadas aos dados de umidade do solo na irrigação total necessária (ITN) fornecida pelos manejos de irrigação do Potencial Mátrico e Balanço Hídrico;
6. Oferecer uma metodologia para concepção de modelos computacionais de previsão de umidade do solo para campos sem sensores (total ou parcialmente) a partir de técnicas de aprendizagem de máquina amplamente usadas para solução de problemas de regressão. Esses modelos computacionais incorporam uma função f de previsão que estima dados de umidade do solo a partir de dados do clima, cultivo e irrigação (Apêndice B) de maneira que os manejos de irrigação obtenham desempenho semelhante àquele de quando usam dados reais providos por sensores.

1.5 Metodologia de Pesquisa

A metodologia aplicada ao longo de toda a pesquisa deste doutorado pode ser sintetizada como segue:

1. Estudo sobre manejos de irrigação — para conhecer os tipos de método e dados, a evapotranspiração de cultivos (Apêndice A) e a retenção e disponibilidade de água no solo;
2. Estudo sobre fusão de dados multinível para sistemas da Internet das Coisas (IoT) descrita em Torres (2017), Torres *et al.* (2017) — entendimento da arquitetura de fusão de dados do solo para o tratamento de *outliers* (Seção 3.3.3);
3. Mapeamento sistemático de literatura — levantamento do estado da arte sobre plataformas e serviços IoT para irrigação na agricultura inteligente (Capítulo 2);
4. Engenharia de requisitos para a execução automática dos manejos de irrigação com previsão de umidade do solo — envolveu entrevistas com um pesquisador da Embrapa⁴ e estudos para identificar e definir os serviços IoT necessários à execução em tempo real dos manejos de irrigação abordados nesta tese. Envolveu também a concepção de modelos

⁴ Rubens Sonsol Gondim, email: rubens.gondim@embrapa.br

- predictivos (Seção 3.2.2) e a metodologia de seleção da melhor técnica de aprendizagem de máquina (Seção 4.1) para solução do problema de umidade do solo (Apêndice B);
5. Modelagem de requisitos da proposta — cenário (Figura 1), dados (Apêndice A.1 e A.2) e classes de serviços IoT (Capítulo 3) necessários à execução em tempo real dos manejos de irrigação que usam dados de umidade do solo reais $\theta_c(\psi_c)$ e estimados $\theta_p(\psi_p)$ pelos modelos de predição (Capítulo 5);
 6. Implementação da proposta para automatização dos manejos de irrigação (Apendice A.3) através de serviços IoT (Seção 5.1). A predição de umidade de solo também foi implementada usando a melhor técnica de aprendizagem computacional daquelas avaliadas nesta tese (Seção 4.2) que soluciona o problema formalizado no Apêndice B;
 7. Análise do impacto das técnicas de remoção de *outliers* e fusão redundante de dados de umidade do solo no desempenho dos manejos de irrigação. Avaliou-se o impacto na irrigação total necessária (ITN) do manejo do Potencial Mátrico (Seção 5.2) e do Balanço Hídrico (Seção 5.2) causado pelo uso de dados de umidade do solo reais $\theta_c(\psi_c)$ ou estimados $\theta_p(\psi_p)$.

1.6 Contribuição

A principal contribuição desta tese é o *framework* Smart&Green, uma solução da Internet das Coisas (*Internet of Things — IoT*) para agricultura inteligente suportar os tradicionais métodos de manejos de irrigação do Turno de Rega, Potencial Mátrico e Balanço Hídrico (Apendice A), com serviços IoT para irrigação, tais como monitoramento, sincronização, preprocessamento, fusão, armazenamento de dados e predição de umidade do solo (Figura 2). Esses serviços IoT são componentes modulares para facilitar o reuso e customização do código, o que é considerado um desafio da implantação da IoT na agricultura (PANG *et al.*, 2015).

A integração de diferentes fontes de dados é viabilizada pelos componentes para:

- o cadastro de informação fornecidas por proprietários rurais e agrônomos sobre a infraestrutura do campo (camada de aplicação da Figura 2);
- o monitoramento em tempo real do clima e do solo através de clientes, servidores de dados e gateway (camada de comunicação da Figura 2);

O *framework* analisa dados e informações fornecidas pelos seus usuários para conhecer a infraestrutura do campo monitorado e escolher/sugerir o manejo de irrigação mais adequado. Os serviços de sincronização e fusão usados pela implementação de acordo com o tempo de

execução e método de cada manejo de irrigação.

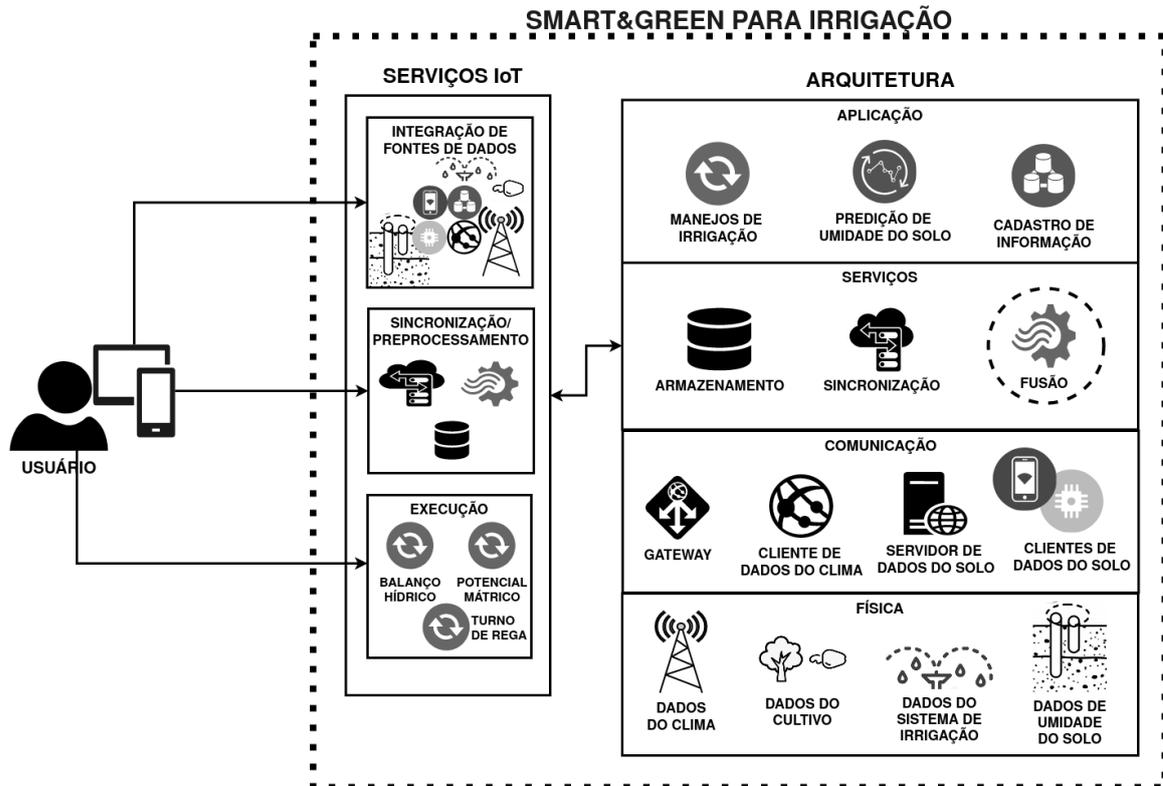


Figura 2 – Visão Geral do Smart&Green. Fonte: Autora

O componente de cadastro de informação também permite que o tratamento de *outliers* seja customizado de acordo com as fontes de dados do clima e solo. O serviço de Fusão é uma implementação da fusão de dados multinível para sistemas IoT descrita em Torres (2017), Torres *et al.* (2017), que adota técnicas de remoção de *outliers* e fusão redundante de dados de umidade do solo.

1.7 Organização da Tese

O Capítulo 2 apresenta uma revisão de literatura sobre plataformas IoT para a agricultura inteligente que suporta a irrigação. No Capítulo 3, são descritos a estrutura e serviços do *framework* Smart&Green. O Capítulo 4 relaciona os materiais, métodos e resultados da avaliação dos modelos computacionais para predição de umidade do solo.

O Capítulo 5 apresenta os materiais usados na implementação do *framework* Smart&Green, a metodologia e resultados da avaliação dos manejos de irrigação do Potencial Mátrico e Balanço Hídrico, pois, estes são impactados pelas técnicas de fusão redundante e remoção de *outliers* aplicadas aos dados de umidade solo reais ou estimados pelo módulo de predição.

A conclusão e trabalhos futuros estão no Capítulo 6. O Apêndice A detalha os tipos de dados e manejos de irrigação. No Apêndice B, o problema de predição de umidade do solo está formalizado. E por último, o Apêndice C possui o projeto de banco de dados do serviço de Armazenamento do Smart&Green.

2 REVISÃO DA LITERATURA

Neste capítulo, estão descritos os trabalhos relacionados ao *framework* Smart&Green proposto. Primeiramente, realizamos o mapeamento sistemático sobre os principais serviços necessários aos manejos de irrigação e que são oferecidos pela agricultura inteligente através da Internet das Coisas (*Internet of Things — IoT*) (Seção 2.1). Um mapeamento sistemático de literatura facilita a identificação e classificação de tópicos relacionados à pesquisa e consiste em um método de extração e resumo de conhecimento em atividades de pesquisa baseada em evidências (SILVA NETO *et al.*, 2019). O reporte do mapeamento está dividido em termos de coleta (Seção 2.2.1) e processamento (Seção 2.2.2) de dados usados pelos manejos de irrigação (Seção 2.2.3), estudos sobre a predição de umidade do solo (Seção 2.2.4) e plataformas IoT para agricultura inteligente que suporta à irrigação (Seção 2.2.5).

2.1 Mapeamento Sistemático da Literatura - Planejamento

Estão descritos nesta seção, o planejamento de um mapeamento sistemático de literatura para estabelecer o estado da arte relacionado ao software da agricultura inteligente de suporte à irrigação. As perguntas deste mapeamento são:

1. Como o software da agricultura inteligente automatiza a **coleta**, a **sincronização/alinhamento**, o **preprocessamento** e a **fusão** dos dados usados pelos métodos de manejos de irrigação (Apêndice A.3) para diferentes infraestruturas de campos agrícolas (Figura 1)?
2. Existe software que recomenda a **escolha do manejo de irrigação adequado** de acordo com a infraestrutura de um campo?
3. Existe software que realiza a **predição de umidade do solo** em campos sem tensiômetros (ex. total, como o Campo 1 da Figura 1 ou parcialmente)?

As fontes de pesquisas são motores de busca web que acessam diversas bibliotecas digitais na Internet: ISI Web¹, Scopus² e Google Scholar³. Definimos cinco grupos de termos de buscas relacionadas às perguntas de pesquisa (Tabela 1). O Grupo 1 reúne palavras-chaves sobre artefatos de software e a IoT. O Grupo 2 possui termos para gestão da agricultura, enquanto que o Grupo 3 concentra serviços e funções desejáveis em um software da agricultura inteligente. O Grupo 4 reúne termos específicos usados nos manejos de irrigação e o Grupo 5 agrupa possíveis

¹ <https://tinyurl.com/yaz2jfle>

² <http://www.scopus.com>

³ <https://scholar.google.com>

tipos de dados e sensores usados no monitoramento climático e do solo. Manualmente, deve ser inserido em cada fonte de busca o algoritmo definido pelos termos dos grupos e operadores lógicos (Tabela 2).

Tabela 1 – Palavras usadas na busca (*search query*). Fonte: Autora.

<p>Grupo 1: framework, API, information technology, Internet of Things, middleware, mobile application, software, Web of Things</p>
<p>Grupo 2: smart agriculture, precision agriculture</p>
<p>Grupo 3: automation methods, control, data aggregation, data fusion, data gathering, data science, data synchronization, forecasting, management, monitoring, prediction, processing, recommendation, registration, selection, storage, visualization</p>
<p>Grupo 4: irrigation management, critical matric potential, crop evapotranspiration, field capacity, irrigation frequency, soil, soil characteristic curve, soil matric potential</p>
<p>Grupo 5: sensors, GPS, humidity, meteorological station, weather station, tensiometer, watermark, localization, temperature, radiation</p>

Tabela 2 – Algoritmo da Busca: (Grupo 1) AND (Grupo 2) AND (Grupo 3) AND (Grupo 4) AND (Grupo 5). Fonte: Autora.

```
TITLE-ABS-KEY("framework"OR "API"OR "application"OR "information technology"OR
"Internet of Things"OR "middleware"OR "mobile application"OR "software"OR "Web of
Things") AND ("smart agriculture"OR "precision agriculture") AND ("automation methods"OR
"control"OR "data aggregation"OR "data fusion"OR "data gathering"OR "data science"OR
"data synchronization"OR "forecasting"OR "management"OR "monitoring"OR "prediction"
OR "processing"OR "recommendation"OR "registration"OR "selection"OR "storage"OR
"visualization") AND ("irrigation management"OR "critical matric potential"OR
"crop evapotranspiration"OR "field capacity"OR "irrigation frequency "OR "soil"OR
"soil characteristic curve"OR "soil matric potential") AND ("sensors"OR "GPS"OR "humidity"
OR "meteorological station"OR "weather station"OR "tensiometer"OR "watermark"OR
"localization"OR "temperature"OR "radiation")
```

Garante-se a qualidade dos artigos selecionados para extração de informações (leitura) somente se os seguintes critérios forem atendidos:

1. Documentos publicados em conferências e revistas que adotam a revisão por pares (*peer-reviewed*); artigos de organizações associadas às áreas da ciência da computação, engenharias; patentes e notas técnicas.
2. Documentos escritos em inglês e português.
3. Documentos publicados no período de 2008 a 2019.

4. Documentos que podem ser obtidos pela rede da UFC sem necessitar de pagamento;

Caso o tema principal do artigo selecionado seja irrelevante ou estiver fora do escopo da pesquisa, o artigo será removido da lista de leituras. Com o propósito de reduzir o número de artigos encontrados, no entanto, mantendo a alta qualidade na resposta das perguntas da pesquisa, os **critérios de seleção** de artigos são subdivididos em critérios de inclusão e critérios de qualidade. Assim, o processo de seleção de trabalhos relacionados ocorre em três fases:

1. **critério de inclusão baseado no resumo**, ou seja, após a leitura do resumo, decidimos se lemos o restante do artigo ou excluimos da lista de potenciais trabalhos relacionados. O resumo deve conter informações sobre software que auxilie na automatização da coleta, pré-processamento, fusão e sincronização/alinhamento de dados usados nos métodos de manejos de irrigação (Seção A.3) e predição de umidade no solo.
2. **critério de inclusão baseado na leitura completa**, isto é, o artigo deve conter pelo menos um termo de cada grupo apresentados na Tabela 1. Caso contrário, ele será excluído.
3. **critério de inclusão baseado na análise de qualidade** determina que os artigos que passaram pelas duas primeiras fases de seleção devem atender todos os seguintes critérios de qualidade ou serão excluídos:
 - a) esse estudo apresenta algum software que automatiza o processo de planejamento ou programação ou execução dos manejos de irrigação (Seção A.3) ou realiza a predição de umidade do solo (Apêndice B)?
 - b) o estudo detalha a infraestrutura e/ou tecnologias de coleta, pré-processamento, fusão e sincronização de dados implementados pela solução proposta?

Na fase de extração de informações dos artigos selecionados, com base na Tabela 3, elaboramos resumos para cada estudo que respondam da maneira mais objetiva possível as três perguntas desta pesquisa desta seção.

Tabela 3 – Forma usada para extrair dados de cada estudo. Fonte: (Autora).

Dados obtidos	Descrição
Título	Título do estudo principal
Ano	Ano de publicação do estudo
Instituição	Nome da Instituição que está liderando a pesquisa
Fonte	Conferência, revista ou livro contendo o estudo principal
Solução	Nome da solução de IoT descrita
Modelo da arquitetura	Descrição da arquitetura usada, seu escopo e limitação

Tabela 3 (cont.)

Dados obtidos	Descrição
Sensores	Informação sobre os tipos de sensores e contagem de sensores por nó presentes na solução
Conectividade e Comunicação	Tecnologias usadas na transmissão de dados
Tratamento de outliers	Algoritmos e metodologias de detecção e remoção de outliers de dados do clima e do solo
Sincronização/Alinhamento de dados	Métodos de sincronização de dados do clima, do solo e do cultivo
Manejos da Irrigação	Planejamento, programação e execução do turno de rega, do balanço hídrico e do potencial mátrico
Predição de umidade so solo	Algoritmos de aprendizagem de máquina, reconhecimento de padrões para predição de umidade do solo

2.2 Mapeamento Sistemático da Literatura - Resultado

A maioria dos artigos encontrados respondem parcialmente às perguntas do mapeamento (Seção 2.1). Muitos artigos tratam apenas da coleta e visualização de dados climáticos e do solo (Seção 2.2.1). Outros trabalhos envolvem pré-processamento de dados aplicando algoritmos de detecção de *outliers* (Seção 2.2.2). Algumas abordagens usam métodos específicos para decidir quando irrigar, porém, outros implementam os métodos tradicionais do Balanço Hídrico e Potencial Mátrico (Seção 2.2.3). A predição de umidade do solo aplica técnicas de aprendizagem de máquina aos dados do cultivo, imagens de satélite, e informações dos manejos de irrigação (Seção 2.2.4).

Assim, quase não existem plataformas da Internet das Coisas (IoT) para agricultura inteligente que possuem os serviços necessários à irrigação, como a coleta, pré-processamento, sincronização e fusão de dados; o planejamento e execução dos manejos de irrigação; como também o serviço de predição de umidade do solo. As plataformas IoT devem ser flexíveis, permitindo que os usuários as configurem para diferentes casos/cenários de monitoramento no contexto da agricultura inteligente para a irrigação (Seção 2.2.5).

2.2.1 Coleta de dados

Geralmente, a grande maioria dos nós sensores são construídos a partir do Arduino (ABAYA *et al.*, 2017; MATH; DHARWADKAR, 2017; RAJKUMAR *et al.*, 2017; HAMPANAVAR; YARAGATTI, 2015; MESAS-CARRASCOSA *et al.*, 2015) ou Raspberry Pi (BALAMURUGAN; SATHEESH, 2017; MAIA *et al.*, 2017; FLORES *et al.*, 2016) conectados a sensores que detectam temperatura T , umidade θ , potencial mátrico ψ do solo, como também temperatura T e umidade relativa RH do ar. A comunicação na rede de sensores sem fio se dá por meio do Zigbee (HEBLE *et al.*, 2018; MAIA *et al.*, 2017; MATH; DHARWADKAR, 2017; KANNAN; THILAGAVATHI, 2013), LoRa (FLORES *et al.*, 2016), Wifi (KAMELIA *et al.*, 2018), Bluetooth (MESAS-CARRASCOSA *et al.*, 2015), GSM (RAJKUMAR *et al.*, 2017) e GPRS (BYISHIMO; GARBA, 2016; NAVARRO-HELLIN *et al.*, 2015).

Este trabalho permite que usuários escolham entre os protocolos CoAP (SHELBY *et al.*, 2014) e MQTT (OASIS, 2014) para enviar dados para que o *framework* Smart&Green armazene e processe durante os manejos de irrigação. Para equipamentos que não podem ser adaptados a nós sensores, como os tensiômetros analógicos, o *framework* disponibiliza uma aplicação móvel para que usuários colem manualmente os dados. O *framework* ainda permite que os usuários selecionem como uma fonte de dados climáticos uma das estações automáticas brasileiras mais próximas do campo monitorado para coletar dados por meio do HTTP.

2.2.2 Preprocessamento de dados

Outras abordagens realizam também o preprocessamento de dados do solo que consiste na calibração de dados brutos – *raw data* (POPOVIC *et al.*, 2017) e tratamento de dados através da filtragem de ruídos (LE; TAN, 2015). Em um sistema que também planeja e executa o manejo de irrigação (HAMOUDA; MSALLAM, 2018), a aplicação do EKF (*Extended Kalman Filter*) aos dados de umidade do solo, evita que os atuadores acionem a irrigação sem necessidade.

O reconhecimento de padrões de consumo de água pelo sistema radicular do cultivo é uma técnica de detecção de *outliers* baseada na análise de séries temporais de dados de umidade coletados em várias profundidades do solo antes, durante e após a irrigação. Ela permite a identificação de mau funcionamento de sensores sinalizando a necessidade de calibragem ou troca (FIGUEROA; POPE, 2017).

Em Torres *et al.* (2017), técnicas de remoção de *outliers* e fusão redundante tratam a série de dados de potencial mátrico ψ (Apêndice A.2) das camadas de solo de um campo IoT (ex. Campo 3 da Figura 1). Primeiro, são retirados da série coletada em uma camada os dados ψ que estejam fora da faixa de valores possíveis detectados pelo tipo de tensiômetro do campo (ex. 0 a 200 kPa). Segundo, aplica-se à série de dados ψ uma das seguintes técnicas: Zscore, MZscore (*Modified Zscore*), GESD (*Generalized Extreme Studentized Deviate*) e Chauvenet. Por fim, agrega-se os dados ψ da série usando a média aritmética como fusão redundante. Para o campo ser irrigado, o dado ψ fusionado deve ser maior que 60 kPa.

O Zscore de um dado ψ (amostra) é a pontuação em unidades (escores) de quantos desvios padrão (abaixo ou acima) esse dado ψ está da média da série de dados ψ (população) coletados em uma mesma camada de solo (NIST/SEMATECH, 2013a). Zscores variam de -3 desvios padrão (que podem estar mais à esquerda da curva da distribuição normal) à +3 desvios padrão (que podem estar mais a direita da curva de distribuição normal). No MZscore, a pontuação é estimada a partir do desvio absoluto da mediana e um dado ψ é considerado um *outlier* quando o valor absoluto do seu MZscore é maior que 3,5 (NIST/SEMATECH, 2013a).

O GESD pode detectar um ou mais *outliers* em um conjunto de dados univariado que segue aproximadamente a distribuição normal. Para isto, o GESD necessita apenas do limite superior para a especificação do número de *outliers* suspeitos (NIST/SEMATECH, 2013b). O critério de Chauvenet assume que os dados estão distribuídos normalmente e cria uma faixa de valores aceitáveis ao redor da média. Assim, qualquer valor fora dessa faixa (< 0.5) são considerados *outliers* (BARNETT; LEWIS, 1978).

O *framework* proposto nesta tese permite que os usuários definam critérios para remoção de *outliers* imediata para cada tipo de dado climático e de umidade do solo usados pelos manejos de irrigação. O serviço de Fusão do *framework* executa em tempo real a fusão de dados multinível para sistemas de Internet das Coisas (TORRES, 2017; TORRES *et al.*, 2017) para os manejos de irrigação (Apêndice A). O serviço de Fusão aplica 1) os critérios de remoção imediata de *outliers* para dados do clima e de umidade do solo; 2) diferentes técnicas de remoção de *outliers* (Zscore, MZscore, GESD e Chauvenet) e 3) fusão redundante de dados (média, mínimo, mediana e máximo) aos dados de potencial mátrico ψ usados na dedução da umidade do solo $\theta(\psi)$. Esta tese traz também um estudo que analisa o impacto do uso em conjunto dessas técnicas no desempenho dos manejos de irrigação do Balanço Hídrico e Potencial Mátrico (Capítulo 5).

2.2.3 Manejos de Irrigação

Outros trabalhos iniciam automaticamente a irrigação a partir da implementação de algoritmos próprios para análise dos dados coletados por uma rede de sensores sem fio. Assim, em Ferrandez *et al.* (2018), o momento da irrigação, bem como a quantidade de água a ser usada são determinados por algoritmos baseados em árvore de decisão que processam dados de umidade e temperatura do solo. Regras rígidas (IF-THEN) comparam valores limites (*threshold*) com valores coletados de sensores de umidade do solo para, então, ligar ou não microaspersores em outras abordagens (PATOKAR; GOHOKAR, 2018; VAISHALI *et al.*, 2017).

Existem plataformas que suportam apenas a decisão sobre quando executar os manejos de irrigação tradicionais (Apêndice A.3). Estas ferramentas permitem que agrônomos e proprietários rurais gerenciem componentes da Internet das Coisas que monitoram seus campos. Assim, dados de sensores de solo e imagens capturadas por um AUV (*Unmanned Aerial Vehicle*) são usados em métodos de manejo como o do Potencial Mátrico (Seção A.3.4) (PAVON-PULIDO *et al.*, 2017) em favor da correta estimativa da quantidade de água da irrigação.

Sistemas web realizam o planejamento e execução do manejo irrigação. Além de permitirem a visualização de dados climáticos e de umidade do solo coletados pela rede de sensores sem fio, esse sistema usa regras rígidas na análise de dados e, quando necessário, envia SMS para seus usuários notificando o momento de irrigar (KARIMI *et al.*, 2018) ou ativa automaticamente a irrigação (MAT *et al.*, 2015; MAT *et al.*, 2014). Em outras abordagens, o sistema leva em consideração a previsão de chuva disponível na Internet para ativar ou não o sistema de irrigação. Assim, mesmo se a umidade do solo estiver baixa, mas a previsão indica chuva para os próximos três dias, o sistema não ativará a irrigação (CAETANO *et al.*, 2015). A própria rede de sensores sem fio também pode planejar e executar o manejo de irrigação (BHANU *et al.*, 2014). Neste caso, os nós coletam e analisam dados de umidade do solo e acionam a irrigação até que a umidade atinja a capacidade de campo.

Um sistema SCADA (*Supervisory Control and Data Acquisition*) pode atuar na coleta, planejamento e execução automática do Balanço Hídrico (CAPRARO *et al.*, 2014). Neste trabalho, os nós sensores de dados do solo são também atuadores que controlam a irrigação e enviam informações da irrigação continuamente a um sistema de gerência. O sistema recebe dados de uma estação meteorológica para calcular a evapotranspiração de referência ETo . O sistema também possui dados do coeficiente de cultivo k_c e escolhe o seu valor de acordo com o estágio de desenvolvimento do cultivo.

A sincronização de dados do cultivo (k_c) com os dados climáticos e do solo é levada em consideração pelos manejos de irrigação do Balanço Hídrico e Turno de Rega. Algumas aplicações fazem a sincronização usando dados de cultivos pré-cadastrados para se determinar a quantidade de água correta para o estágio de desenvolvimento do cultivo (MILLER *et al.*, 2018; SAWANT *et al.*, 2017; MAUGET; LEIKER, 2010; CARLESSO *et al.*, 2009).

Os sistemas descritos nesta seção não são flexíveis o suficiente para deixar que os usuários configurem os atributos de um campo monitorado, tais como, tipo de cultivo, tipo e número de sensores do solo e parâmetros dos sistemas da irrigação que são usados pelos manejos de irrigação. Esses sistemas geralmente trabalham com configurações e infraestrutura predefinidas. Assim, o reuso de código é impraticável para manejos de irrigação de campos que não se encaixam na configuração do sistema.

O *framework* Smart&Green proposto permite que os usuários registrem informações relevantes para os manejos de irrigação: diferentes tipos de cultivo, características do solo, sistema de irrigação, fontes de dados (climática ou de umidade do solo) e critérios de remoção de *outlier*. Neste trabalho, o *framework* planeja os manejos de irrigação do Potencial Mátrico (Eq. A.12), do Balanço Hídrico (Eq. A.7) e do Turno de Rega (Eq. A.2).

Além do mais, nenhum dos sistemas descritos realiza a predição da umidade do solo $\theta_p(\psi_p)$ para campos sem sensores (total como o Campo 1 da Figura 1 ou parcialmente). O *framework* aqui proposto implementa modelos computacionais para estimar a umidade do solo baseando-se em dados climáticos, informações do cultivo e da irrigação. Neste trabalho, os dados estimados $\theta_p(\psi_p)$ pelos modelos de predição foram incluídos no manejo de irrigação do Balanço Hídrico, que obteve uma economia média de 56.4% a 90% da irrigação total necessária (Seção 5.4.2).

2.2.4 Predição de umidade dos solo

Com o objetivo de estimar o melhor plano de irrigação para o cultivo de jojoba (GOLDSTEIN *et al.*, 2017), os dados climáticos e do solo (umidade e estresse hídrico) são fornecidos com entrada de técnicas de aprendizagem de máquina de regressão linear, árvore de decisão, *random forest* e GBRT (*Gradient Boosting with Regression Tree*). Foi criado um módulo com o modelo computacional do GBRT, pois, obteve melhor desempenho para funcionar em um serviço de suporte à decisão.

A umidade do solo dos próximos dias pode ser estimada usando algoritmo baseado

nas técnicas de support vector regression (SVR) e k-means que usa dados do solo (temperatura e umidade) dados do clima (temperatura, umidade, radiação solar) e dados de previsão climática (temperatura, umidade, radiação solar, precipitação) fornecidos por sites web de previsão do tempo (GOAP; SHARMA, 2018). O planejamento da irrigação é executado por outro algoritmo que leva em consideração tanto os dados de umidade do solo atuais como aqueles da previsão, além dos dados de precipitação dos sites de previsão do tempo.

Um sistema integrado monitora a seca no nordeste da China analisando imagens de satélite (NOAA, MODIS e FY3A), dados de estações climáticas e dados do solo (LUAN *et al.*, 2015). A previsão de seca usa dados de potencial mátrico na profundidade das raízes do cultivo, índices de depleção, medidas do estresse do cultivo e o manejo de irrigação do Balanço Hídrico para estimar a umidade do solo e a irrigação total necessária para um longo período.

Neste trabalho, dados climáticos, do cultivo e da irrigação são usados para a previsão do potencial mátrico ψ_p a uma profundidade de 15cm do solo (Apendice B). O valor do potencial mátrico ψ_p fornecido pelo modelo de previsão é aplicado ao modelo de Van Genutchen (Eq. A.1) para se obter a umidade do solo $\theta_p(\psi_p)$ usada nos manejos de irrigação (Eq. A.7 e A.12) de campos que não possuem sensores de solo (ex. Campo 1 da Figura 1).

São apresentados duas abordagens de previsão: uma local e outra global. Na abordagem local, um modelo de previsão é criado para cada ponto de monitoramento de um campo experimental (Seção 4.1.2). Na abordagem global, cria-se apenas um modelo de previsão generalizado para qualquer ponto de monitoramento do campo.

Foi avaliado o desempenho de um conjunto de técnicas de aprendizagem de máquina populares que são amplamente aplicadas para resolver problemas de regressão: regressão linear (FREEDMAN, 2009), *Decision Stump* (IBA; LANGLEY, 1992), *M5 Model Trees and Rules* (QUINLAN, 1992; WANG; WITTEN, 1997), *Random Tree, Random Forest* (BREIMAN, 2001), *RepTree* e GBRT (FRIEDMAN, 2000; FRIEDMAN, 2002). Os resultados mostram que o GBRT supera as outras técnicas avaliadas em ambas abordagens.

2.2.5 Plataformas IoT para Agricultura Inteligente

A plataforma SWAMP (KAMIENSKI *et al.*, 2019) é um dos trabalhos relacionados que mais se assemelha a proposta desta tese, pois, reúne vários componentes que possibilitam a implementação de diferentes serviços IoT para viabilizar os manejos de irrigação. Os serviços de coleta (MQTT ou LoRA), processamento e sincronização de dados podem ser customizados a

diferentes tipos de cultivo, clima e país. Serviços de armazenamento e análise de dados podem ser reaproveitados em diferentes configurações de campos agrícolas, o que a torna uma plataforma com excelente flexibilidade. No entanto, as técnicas de processamento e fusão de dados não foram detalhadas e não existe aplicação/serviço de predição de umidade do solo.

O *framework* Agro-IoT (KAMILARIS *et al.*, 2016) também se assemelha a esta proposta, pois, pode ser configurado para coletar, agregar (fusionar) e analisar dados em tempo real no contexto das aplicações de fazendas inteligentes (*smart farming*). O Agro-IoT reusa vários componentes que viabilizam a implementação de aplicações IoT de acordo com a necessidade da infraestrutura do campo agrícola monitorado. Entre os serviços ofertados, estão o gerenciamento de dispositivos (como cadastro/registo de sensores e atuadores) e a detecção de eventos a partir da análise de dados coletados. Porém, não é mencionado a facilitação dos serviços de sincronização de dados e remoção de *outliers* ou um componente que suporte a predição de umidade do solo.

2.3 Síntese do Capítulo

Existem muitas iniciativas de implantação da Internet das Coisas (*Internet of Things* — IoT) para agricultura inteligente com suporte à irrigação. No entanto, a grande maioria destes trabalhos preocupa-se com a implementação de alguns dos serviços IoT que o *framework* Smart&Green aborda, tais como: monitoramento, sincronização, preprocessamento associado a remoção de *outliers* e a fusão redundante de dados de umidade do solo usados em manejos de irrigação executados em tempo real. Estas iniciativas adequam-se apenas a tipos predeterminados de cultivo, de solo e sistema de irrigação e, portanto, não são flexíveis o suficiente para serem reusadas de acordo com a infraestrutura de diferentes tipos de campos (Figura 1). Outro aspecto importante foi que a predição de umidade do solo não foi tratada por nenhuma das duas plataformas IoT para agricultura inteligente consideradas mais correlatas ao Smart&Green (Seção 2.2.5).

3 PROPOSTA

Este capítulo apresenta os detalhes do *framework* Smart&Green. Primeiro, contextualizamos o conceito de *framework* na agricultura inteligente para irrigação (Seção 3.1). Em seguida, descrevemos o funcionamento das aplicações de automatização dos manejos de irrigação e de predição de umidade do solo (Seção 3.2), serviços de armazenamento, sincronização, fusão redundante de dados de umidade do solo (Seção 3.3) e comunicação do *framework* com as fontes de dados climáticas e de umidade do solo (Seção 3.4).

3.1 Visão Geral do *framework* Smart&Green

Um **framework** consiste em uma "*implementação específica do esqueleto de uma infraestrutura usada para concepção de um trabalho*" (PRESSMAN, 2010). O trabalho pode ser qualquer atividade a ser desempenhada por pessoas ou software. Nesta proposta, o trabalho consiste no planejamento, programação e execução dos manejos de irrigação do Turno de Rega, Potencial Mátrico e Balanço Hídrico (Seção A.3), os quais podem ser enriquecidos pela predição de umidade do solo. Proprietários rurais e agrônomos vão desempenhar esse trabalho de gestão da irrigação para diferentes tipos de campos (Figura 1) sendo auxiliados por qualquer software da agricultura inteligente que utilizar o *framework* Smart&Green.

Em engenharia de software, um *framework* é ainda definido como uma "*mini-arquitetura que pode ser reusada e que provê um comportamento e uma estrutura genérica para uma família de abstrações de software em conjunto com um contexto que especifica a colaboração e uso deles dentro de um dado domínio*" (AMBLER, 1998). Assim, dado o domínio da agricultura inteligente e o contexto da irrigação, o *framework* Smart&Green especifica a estrutura de um conjunto de classes¹ concretas e abstratas para implementação de serviços modularizados de coleta, armazenamento, sincronização e processamento de dados do clima, solo, cultivo e sistema de irrigação.

O *framework* Smart&Green também implementa os modelos matemáticos que regem os diferentes métodos de manejos de irrigação os quais determinam o comportamento das classes e módulos do *framework*. Como mini-arquitetura, o *framework* Smart&Green pode ser reusado por um software da agricultura inteligente que necessite das funcionalidades genéricas dos manejos de irrigação (Apêndice A) e predição de umidade do solo (Apêndice B). A Figura 2 apresenta

¹ conceito do paradigma da orientação a objetos

a arquitetura conceitual do *framework* dividida em quatro camadas: **Aplicação, Serviços, Comunicação e Física**. A arquitetura em camadas permite a discussão das principais funções que compõem os serviços IoT para irrigação: 1) de integração de fontes de dados, 2) a sincronização e o pré-processamento de dados customizados/adequados às diferentes infraestruturas de campos para 3) a execução em tempo real de diferentes manejos de irrigação.

3.2 Camada de Aplicação

A camada de Aplicação fornece meios para a execução em tempo real dos manejos de irrigação (Seção 3.2.1) e a predição de umidade do solo (Seção 3.2.2).

3.2.1 Automatização dos Manejos de Irrigação

Para que os manejos do Turno de Rega, Balanço Hídrico e Potencial Mátrico (Apêndice A.3) sejam executados em tempo real, os usuários precisam fornecer informações que caracterizam a infraestrutura atual do campo agrícola (Figura 1). Essas informações são usadas para o *framework* integrar as diferentes fontes de dados e escolher o manejo de irrigação mais adequado para o campo monitorado. Seguem as **funções** relacionados a coleta de informações sobre a infraestrutura do campo a ser monitorado:

1. **Cadastrar Usuário:** o *framework* Smart&Green permite os tipos de usuário especialista e regular. Usuários especialistas podem fornecer informações técnicas sobre o cultivo, sistemas de irrigação (Apêndice A.1), tipos de solo (ex. as constantes do modelo Van Genuchten usadas na Eq. A.1). Usuários especialistas também cadastram critérios de remoção de *outliers* para dados climáticos e do solo e escolhem técnicas do serviço de Fusão de dados de umidade do solo. Usuários regulares podem registrar propriedades rurais e campos.
2. **Cadastrar Propriedade Rural:** deve-se fornecer o endereço da propriedade (cidade e estado brasileiros) ou coordenadas geográficas de latitude e longitude para que *framework* possa selecionar a estação meteorológica mais próxima de sua localização.
3. **Cadastrar Cultivo:** um usuário especialista fornece a curva do coeficiente do **cultivo** $\{(K_{c, inicial}, dias), (K_{c, intermediario}, dias), (K_{c, final}, dias)\}$ e umidade da condição crítica θ_{cr} do cultivo (Seção A.2). É necessário também informar a **idade do cultivo** em dias desde que ele foi plantado;

4. **Cadastrar Sistema de Irrigação:** um usuário especialista insere tipos de sistemas de irrigação (microaspersão, gotejamento, etc.).
5. **Cadastrar Estação Climática:**
 - a) **Selecionar Estação Climática:** O Smart&Green seleciona uma estação climática automática (INMET, 2018) próxima da propriedade rural. Usuários podem confirmar ou escolher outra.
 - b) **Cadastrar Critério de Remoção de *Outlier* da Estação:** um usuário especialista pode associar valores mínimo e máximo para cada tipo de dado climático de acordo com a região geográfica da estação ms , que provê um conjunto de dados $M_{ms} = \{T_{max}, T_{min}, RH_{max}, RH_{min}, R_n, u_2, P, PREC\}$, cujos tipos estão definidos no Apêndice A.1.
6. **Cadastrar Sensor de Umidade do Solo:** o usuário pode inserir tipos de sensores do solo. O *framework* já traz tensiômetros analógicos e tensiômetros digitais cadastrados.
7. **Cadastrar Campo**, no qual usuários podem:
 - a) **Área:** tamanho do campo (m^2).
 - b) **Distribuição de água:** (1) rotação, (2) demanda (Apêndice A.3);
 - c) **Escassez de água:** (1) sim, (2) não, (para ajudar na seleção do manejo de irrigação mais adequado);
 - d) **Inserir cultivo:** deve-se **escolher um cultivo cadastrado**, se o cultivo do campo que está sendo configurado já estiver na base de dados do *framework*. Ou, então, deve-se **Cadastrar Cultivo**.
 - e) **Inserir Sistema de Irrigação:** fornecendo um tipo já cadastrado e informações específicas da precipitação efetiva P_{irri} e eficiência do sistema Ef (Apêndice A.1).
 - f) **Inserir Pontos de Monitoramento do Solo:** cada ponto de monitoramento consiste em uma bateria B de um ou mais tensiômetros que monitoram o potencial mátrico ψ_m de uma ou mais camadas de solo (Apêndice A.2). Logo, o número de camadas monitoradas varia conforme o número de tensiômetros em um ponto de monitoramento do campo. Portanto, deve-se fornecer o número de pontos de monitoramento e o número de camadas de solo do campo. Para cada camada monitorada, deve-se informar:
 - i. a profundidade de instalação do tensiômetro;
 - ii. largura z da camada monitorada;

- iii. as constantes de Van Genuchten (Eq. A.1);
- iv. a umidade da condição de campo θ_{cc} (definição no Apêndice A.2);
- v. critérios de remoção imediata de *outlier* (valores mínimo e máximo detectados pelo sensor);
- vi. técnica de remoção de *outlier*: (1) *Zscore*, (2) *MZscore*, (3) GESD, (4) *Chauvenet*;
- vii. técnica de fusão redundante: (1) Média, (2) Mínimo, (3) Mediana e (4) Máximo;
- viii. tipo de sensor do solo cadastrado (ex. tensiômetro analógico ou digital);
- ix. método de coleta de dados do solo: (1) planilha eletrônica, (2) aplicação móvel HTTP, (3) protocolo de comunicação IoT (CoAP ou MQTT).
- x. Endereços de Servidores: os endereços IP ou URL do servidor (HTTP, CoAP ou MQTT) e do gateway do campo monitorado (detalhes na Seção 3.4).

Sobre os métodos de coleta de dados do solo, as opções 1 e 2 são indicadas para tensiômetros analógicos em campos sem Internet (ex. Campo 2 da Figura 1), cujo envio dos dados de umidade do solo para o armazenamento do *framework* não é feito no momento da coleta manual, mas, posteriormente via protocolo HTTP. A opção 3 é indicada para campos IoT (ex. Campo 3 da Figura 1), no qual um gateway habilitado a CoAP e MQTT envia os dados de umidade do solo. A velocidade/tecnologia de acesso à Internet deve ser informada se o usuário desejar que o *framework* escolha o protocolo IoT mais adequado. Por exemplo, como o protocolo CoAP possui baixo overhead, é mais indicado quando o acesso é GPRS/GSM.

Todas informações coletadas caracterizam a infraestrutura do campo e são armazenadas (Seção 3.3.1), pois, vão ser usadas por outros módulos/serviços do Smart&Green. A Tabela 4 mostra como o Smart&Green relaciona os tipos de dados usados pelos manejos de irrigação às suas fontes, que podem estar presentes ou não, nas diferentes infraestruturas de campos (Figura 1). O Campo 1 da Figura 1 não possui sensores de umidade do solo em sua infraestrutura, logo, os manejos de irrigação mais indicados não usam dados de umidade do solo $\theta_c(\psi_c)$ provenientes de sensores, como:

1. Balanço Hídrico — BH;
2. Turno de Rega Fixo — TR_f ;
3. Turno de Rega Variável — TR_v ;
4. Balanço Hídrico com dados de umidade do solo estimados pelo modelo de predição — $BH + \theta_p(\psi_p)$

5. Potencial Mátrico com dados de umidade do solo estimados pelo modelo de predição —
 $PM + \theta_p(\psi_p)$

Tabela 4 – Os dados usados pelos manejos de irrigação são providos por várias fontes de um campo monitorado. Fonte: Autora.

Dados			Campo		
Descrição	Tipos	Fonte	1	2	3
Clima	variáveis: $M_{ms} = \{T_{max}, T_{min}, RH_{max}, RH_{min}, R_n, u_2, P, PREC\}$	Estação Meteorológica (HTTP)	X	X	X
Cultivo	variáveis: K_c , Fase (dias) constante: θ_{cr}	Usuário (Cadastro)	X	X	X
Sistema de Irrigação	constantes: P_{irri}, Ef	Usuário (Cadastro)	X	X	X
Solo	variáveis: $BT_c = \{B_{1,c} = \{\psi_{m_1,c,z_1}, \psi_{m_1,c,z_2}\}, \dots, B_{n,c} = \{\psi_{m_n,c,z_1}, \psi_{m_n,c,z_2}\}\}$	tensiômetros analógicos (planilha ou aplicação móvel)	-	X	-
		objeto IoT (CoAP ou MQTT)	-	-	X
	constante: θ_{cc}	Usuário (Cadastro)	X	X	X

Os Campos 2 e 3 da Figura 1 requerem métodos de manejos que usem todos tipos de dados (clima, solo, irrigação, cultivo) para que o manejo de irrigação forneça com maior exatidão a irrigação total necessária (ITN). Assim, os métodos mais indicados para estas infraestruturas de campos são:

1. Potencial Mátrico com dados de umidade do solo reais $\theta_c(\psi_c)$ e
2. Balanço Hídrico com dados de umidade do solo reais $\theta_c(\psi_c)$.

O serviço de integração de fontes de dados checa a disponibilidade dos dados usados pelos manejos de irrigação usando módulos/componentes dos serviços de Armazenamento (Seção 3.3.1) e Comunicação (Seção 3.4) da arquitetura do Smart&Green (Figura 2). A Figura 3 apresenta exemplos de integração de fontes de dados que são variáveis (Tabela 4) do clima M_{ms} , do cultivo $\{k_c$ e fase} (Apêndice A.1) e do solo BT_c (Apêndice A.2). Os dados do cultivo estão no banco de dados, logo, o serviço de integração deve consultar sempre o armazenamento.

A integração de dados de umidade do solo verifica primeiro o armazenamento e solicita dados BT_c ao usuário quando a fonte de dados forem sensores analógicos (Tabela 4). Campos IoT enviam BT_c assíncronamente para o armazenamento, logo, se não existirem no banco de dados, os nós sensores não os coletaram ainda, então, realmente, não estão disponíveis.

A integração dos dados do clima também verifica primeiro se o banco de dados do Smart&Green possui o conjunto M_{ms} . O cliente de dados meteorológicos (Seção 3.4) é utilizado

para obter dados das estações automáticas (INMET, 2018) somente quando os dados não estão disponíveis localmente. A integração de fontes de dados climáticos deve ser feita tanto para 1) dados diários $M_{ms,D}$, pois, podem ser usados pelo manejo do Balanço Hídrico (Eq. A.7); e, 2) dados mensais $M_{ms,M}$, para o Turno de Rega (Apêndice A.3.2).

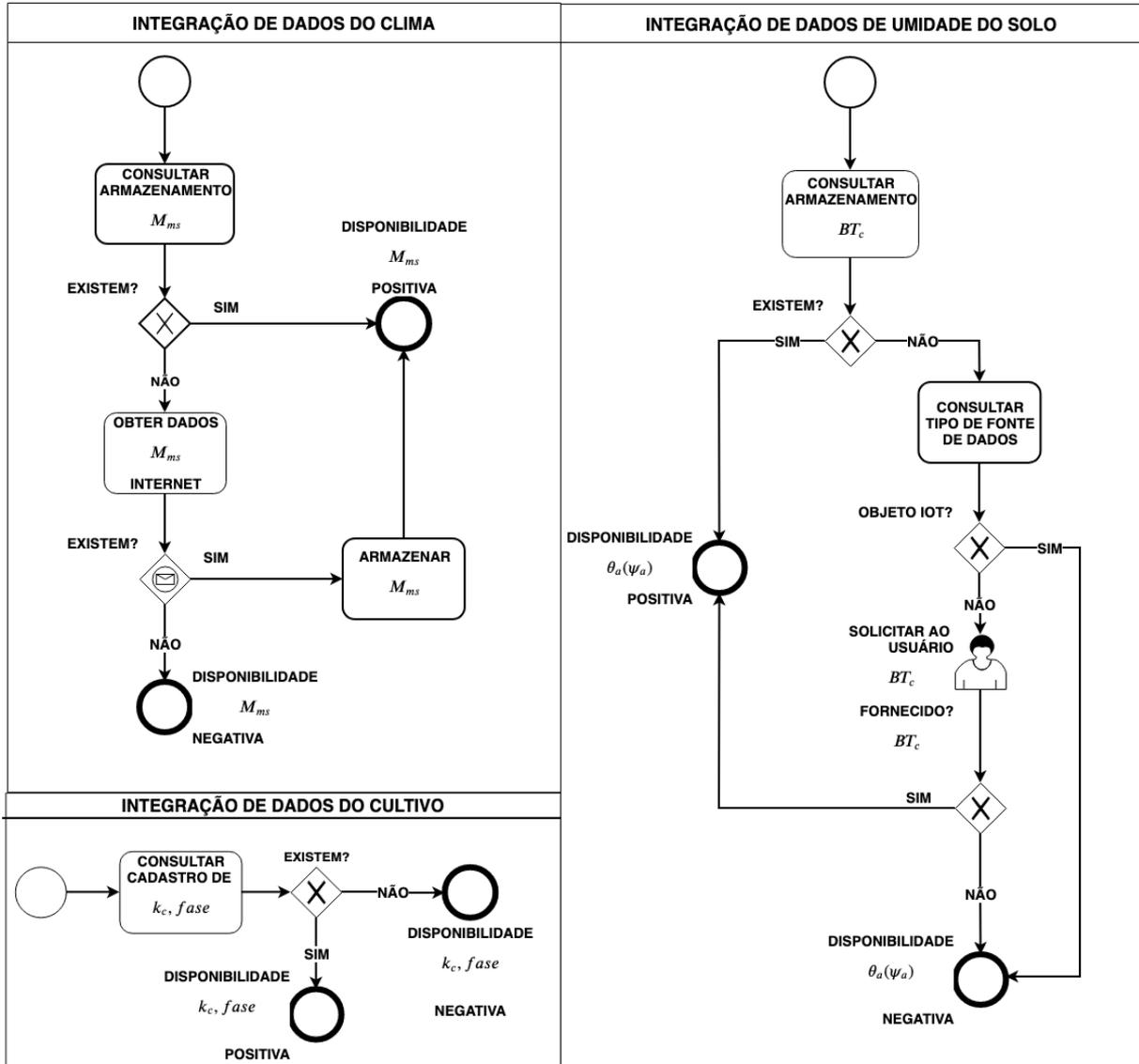


Figura 3 – Exemplos de integração de fontes de dados usados pelos manejos de irrigação. Fonte: Autora.

Cada manejo de irrigação usa um conjunto de vários tipos de dados do clima, do cultivo, da irrigação e do solo (Apêndice A) presentes na Tabela 5. Existe ainda a indicação de qual manejo usar de acordo com a forma de distribuição de água (Seção A.3.1) e quais manejos podem usar dados de umidade do solo $\theta_p(\psi_p)$ estimados pelas abordagens de previsão (Seção 3.2.2). A Figura 4 apresenta as etapas, informações e as regras rígidas (IF-THEN) que fazem com que o Smart&Green ofereça o suporte de tomada de decisão sobre o manejo mais adequado

para diferentes infraestruturas de campos agrícola (Figura 1).

Tabela 5 – Informações usadas para a escolha dos manejos da irrigação

Domínio da Informação		TR _f	TR _v	BH	BH + $\theta_c(\psi_c)$	BH + $\theta_p(\psi_p)$	PM $\theta_c(\psi_c)$	PM $\theta_p(\psi_p)$
clima	$M_{ms,D}$			x	x	x		x
	$M_{ms,M}$	x	x					
cultivo	k_c	x	x	x	x	x		x
	$\theta_{cr}(\psi_{cr})$		x				x	x
irrigação	P_{irri}	x	x	x	x	x	x	x
	Ef	x	x	x	x	x	x	x
	Distr. água por rotação	x		x	x	x		
	Distr. água por demanda		x	x	x	x	x	x
solo	$\theta_a(\psi_a)$				x		x	
	$\theta_{cc}(\psi_{cc})$		x		x	x	x	x
	z		x		x	x	x	x
	ctc_{VG}		x		x	x	x	x
	Mod. Pred. $\theta_p(\psi_p)$					x		x

Considera-se que as informações do cultivo, da irrigação, do solo (exceto, θ_a, θ_p) e os dados climáticos mensais $M_{ms,M}$ estão sempre disponíveis (Tabela 5), pois, foram armazenadas (Seção 3.3.1) durante o processo de cadastro de informação sobre o campo. A escolha de um método de manejo começa com a integração de fontes de dados (Figura 4). Em seguida, o campo a ser monitorado é classificado em Campo 1 ou Campo 2/3 (Figura 1) de acordo com a disponibilidade da umidade do solo $\theta_a(\psi_a)$.

Na seleção de manejos para o Campo 1, a não disponibilidade dos dados climáticos $M_{ms,D}$ diários é decisiva para a escolha do manejo de irrigação do Turno de Rega –TR (Apêndice A.3.2), que usa dados mensais $M_{ms,M}$. A forma de distribuição de água determina a frequência de execução do manejo do Turno de Rega: 1) variável, quando a água é distribuída por demanda (e adiciona os dados da Eq. A.3); e, 2) fixa, quando a água vem por rotação (Figura 4).

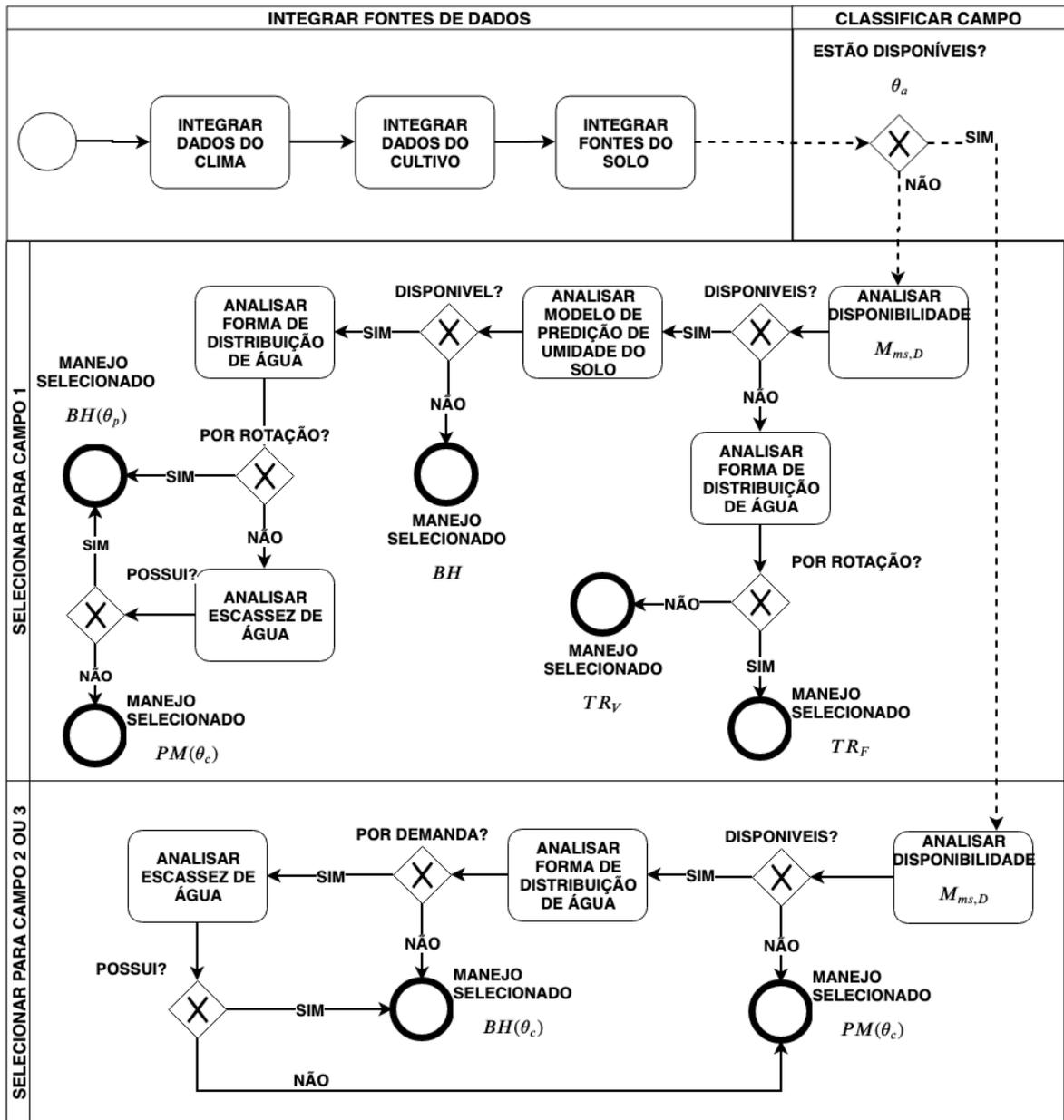


Figura 4 – Seleção de manejos de irrigação. Fonte: Autora.

O campo monitorado precisa ter as mesmas contantes de Van Genuchten (Eq. A.1, ct_{VG} da Tabela 5) que o modelo de predição de umidade do solo. Ainda analisando a Figura 4, a escassez de água confirmada pelo usuário é usada para selecionar o manejo do Balanço Hídrico (BH), que é mais exato no cálculo da irrigação total necessária (Capítulo 5). No caso da seleção de manejos para o Campo 2 ou 3 que recebem água por demanda, a informação sobre a escassez de água positiva determina a seleção do manejo do BH, visto que, ele tende a ser mais exato, já que usa mais tipos de dados que o método do Potencial Mátrico (PM).

3.2.2 Predição de Umidade do Solo

O problema de predição de umidade dos solo está definido no Apêndice B. Neste trabalho, são apresentadas duas abordagens de predição, uma local e outra global, que podem aprender uma função de predição \hat{f} para estimar o potencial mátrico ψ_p de tensiômetros que monitoram a camada mais superficial do solo de um grande campo (maior que $7500 m^2$).

A abordagem local cria diferentes funções de predição \hat{f}_i para cada tensiômetro da camada superficial de solo do campo. A abordagem local usa as observações registradas por cada tensiômetro b_i e pela estação climática ms como também pela quantidade de água w_i dada ao cultivo e o coeficiente de cultivo k_c relacionado ao estágio de vida atua do cultivo. A abordagem local define a predição \hat{f} em termos de n diferentes funções \hat{f}_i para predição local. No entanto, se o campo possuir um grande número de tensiômetros, vai ser necessário o treinamento de um grande número de diferentes modelos de predição.

A abordagem global cria uma única função de predição \hat{f} usando informação de todos os tensiômetros do campo. Assim, o modelo de predição global pode não se ajustar a alguns tensiômetros de maneira individual. O Capítulo 4 reúne detalhes da metodologia de treinamento e seleção da melhor técnica de aprendizagem de máquina associada a função preditiva \hat{f} para ambas abordagens.

O *framework* proposto usa os modelos de predição como um sistema de suporte a decisão da irrigação (Figura 5). Para isto, realiza-se a coleta e armazenamento automático do conjunto de dados climáticos M_{ms} (definidos no Apêndice B) publicados na Internet por uma estação climática ms próxima ao campo. Os dados M_{ms} são processados para remoção imediata de possíveis *outliers*.

A predição de umidade do solo estima o potencial mátrico ψ_p de um campo baseando-se em (M'_{ms}, W, k_c) (Figura 5). O manejo de irrigação calcula a umidade do solo θ_p aplicando ψ_p ao modelo de umidade do solo (GENUCHTEN, 1980). Depois, $\theta_p(\psi_p)$ é usado na Equação A.12 ou A.7 para calcular a irrigação total necessária (ITN).

No contexto da Internet das Coisas (IoT), o *framework* pode mandar o valor de ITN para um atuador iniciar o sistema de irrigação. Assim, o sistema de irrigação só será desligado quando o campo receber a quantidade de água especificada por ITN, por exemplo. Embora, esta proposta de tese não lide com sistemas de irrigação baseados em atuadores, essa questão é de interesse de uma completa implementação de um sistema IoT para agricultura inteligente.

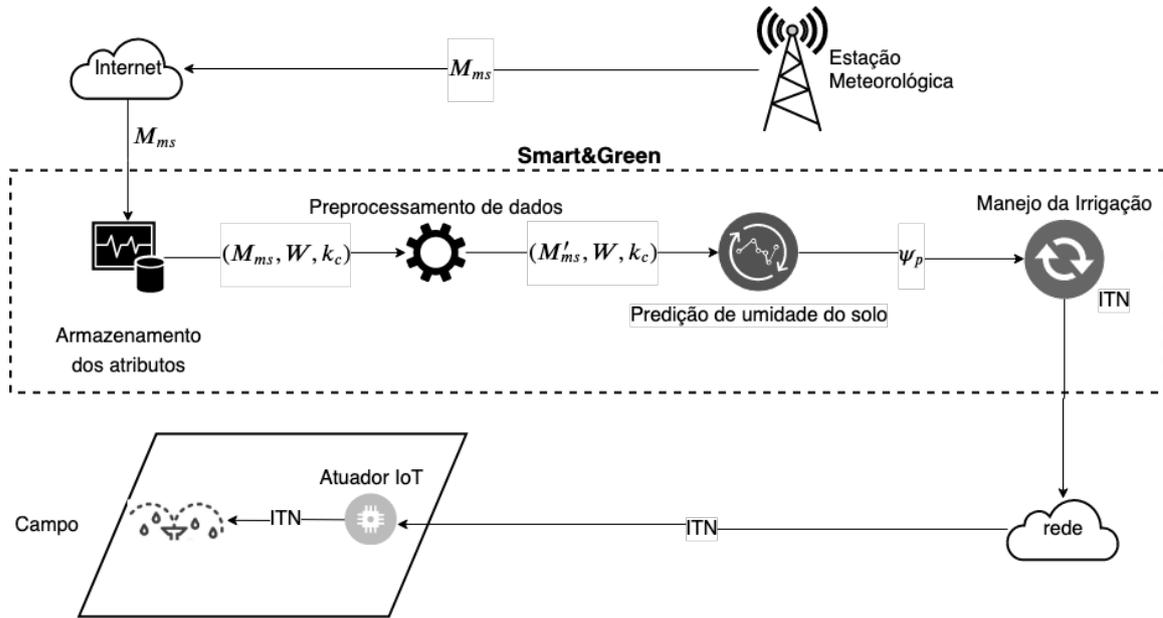


Figura 5 – Smart&Green - modelo de predição de umidade do solo como um sistema de suporte a decisão (DSS) da irrigação. Fonte: Autora.

3.3 Camada de Serviços

A camada de Serviços dá suporte a camada de Aplicação armazenando dados fornecidos pelos usuários do *framework*, sincronizando dados do cultivo, de estações meteorológicas e do solo quando solicitado pela aplicação (Figura 6). Ela também realiza a fusão redundante de dados do solo para campos agrícolas com tensiômetros (Campo 2 e 3 da Figura 1).

3.3.1 Armazenamento

O serviço de armazenamento permite a **persistência centralizada** de dados de monitoramento climático, cultivo, sistema de irrigação e solo presente em campos agrícolas de uma propriedade rural usados em aplicações de manejos de irrigação e predição de umidade do solo. O projeto do banco de dados do Smart&Green está no Apêndice C detalhando as principais regras de negócio, dicionário de dados com relacionamentos e tipos de dados.

3.3.2 Sincronização de dados

Quando são executadas, as aplicações de automatização de manejos de irrigação e predição de umidade do solo solicitam ao serviço de sincronização de dados informações (Figura 6, fluxo 1) sobre o cultivo, o clima e o solo relacionado a um campo cadastrado. Por último (Figura 6, fluxo 3), o serviço de sincronização entrega a aplicação o conjunto de dados composto

por $\{K_{c,atual}, M_{ms,período}, BT_{c,período}\}$.

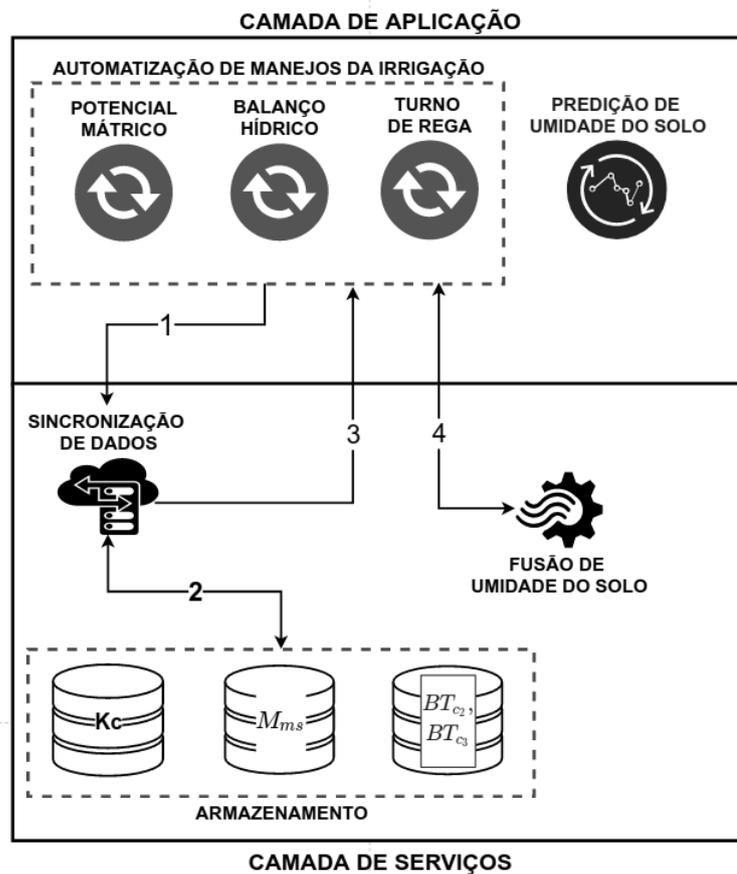


Figura 6 – Interação entre as camadas de Aplicação e Serviços. Fonte: Autora

A **sincronização de dados do cultivo** seleciona o coeficiente da cultura do estágio de desenvolvimento atual $K_{c,atual}$ do cultivo. Para isto, a curva do coeficiente do cultivo deve estar cadastrada e a idade do cultivo (dias) atualizada no *framework*.

A **sincronização de dados climáticos** busca no armazenamento local o conjunto de dados M_{ms} (definidos na Seção A.1) da estação climática mais próxima da propriedade rural. Caso o M_{ms} disponível localmente esteja fora do período de tempo necessário para executar a aplicação, a sincronização invoca os serviços da camada de comunicação para baixar os dados do período correto pela Internet.

No caso da **sincronização de dados do solo**, tomando como exemplo o cenário da Figura 1, o *framework* deve levar em consideração o período de tempo necessário para a execução do manejo de irrigação (ex. diário ou a cada três dias). Assim, o serviço de sincronização busca no banco de dados local (Figura 6, fluxo 2) o conjunto BT_c de dados de potencial mátrico ψ do campo em questão.

Caso não existam dados para o período necessário, a sincronização solicita ao usuário a entrada de dados quando o campo possui tensiômetros analógicos (ex. Campo 2 da Figura 1). Se não houver BT_c para o período, o serviço de sincronização, sinaliza à camada de Aplicação para que seja escolhido outro tipo de manejo de irrigação que não necessite de dados reais de umidade do solo $\theta(\psi)$ (Seção 3.2.1).

3.3.3 Fusão de dados de umidade do solo

O serviço de fusão realiza o pré-processamento de dados, isto é, trata os dados a serem utilizados pelos manejos de irrigação através de critérios e técnicas de remoção de *outliers*, como também, fusão redundante de dados de umidade do solo. O serviço de fusão de dados de umidade do solo (Figura 6, fluxo 4) é usado quando o campo possuir sensores de umidade do solo coletando dados em pelo menos uma camada de solo em diferentes pontos de monitoramento (ex. Campos 2 ou 3 da Figura 1). O serviço de fusão também pode tratar os dados estimados pelo módulo de previsão de umidade do solo para os manejos de irrigação em campos sem sensores de umidade do solo (ex. total como o Campo 1 da Figura 1, ou parcialmente). O Serviço de Fusão executa em tempo real a Fusão de Dados Multinível para Sistemas de Internet das Coisas de Torres *et al.* (2017).

Tendo o Campo 2 da Figura 1 como exemplo, o *framework* sincroniza/alinha previamente, dentro de um período de tempo, um conjunto de dados de potencial mátrico ψ_m de duas camadas monitoradas (z_1 e z_2) em seis diferentes pontos (locais): $BT_{c_2, periodo} = \{A_{c_2, z_1}(\psi_{m_1, c_2, z_1}, \dots, \psi_{m_6, c_2, z_1}), A_{c_2, z_2}(\psi_{m_1, c_2, z_2}, \dots, \psi_{m_6, c_2, z_2})\}$. A fusão redundante de dados do solo (Figura 7) consiste na submissão do conjunto de dados alinhados A_{c_2} de cada camada de solo monitorada aos critérios e/ou técnicas de detecção e remoção de *outliers* (DRO) e fusão redundante (FR) conforme uma arquitetura de fusão de dados multinível (TORRES *et al.*, 2017).

Os **critérios de remoção imediato de outliers** são específicos para um tipo de sensor cadastrado (Seção 3.2.1) e consistem em valores mínimos e máximos possíveis de leitura do sensor. Assim, dados que estejam fora desses limites, são considerados *outliers* e retirados do conjuntos de dados A_{c_2} a ser fusionado. O *framework* traz a implementação de **técnicas de remoção de outliers** Zscore, MZscore (*Modified Zscore*), GESD (*Generalized Extreme Studentized Deviate*) e Chauvenet, visto que, foram testados por Torres *et al.* (2017). Por exemplo, pode-se configurar o *framework* para aplicar a técnica Zscore para dados da camada z_1 do Campo 2, ou seja, $DRO_{c_2, z_1} = Zscore(A_{c_2, z_1}(\{\psi_{m_1, c_2, z_1}, \dots, \psi_{m_6, c_2, z_1}\}))$. Enquanto,

que a técnica GESD pode ser aplicada para dados coletados na camada z_2 , isto é, $DRO_{c_2,z_2} = GESD(A_{c_2,z_2}(\{\psi_{m_1,c_2,z_2}, \dots, \psi_{m_6,c_2,z_2}\}))$.

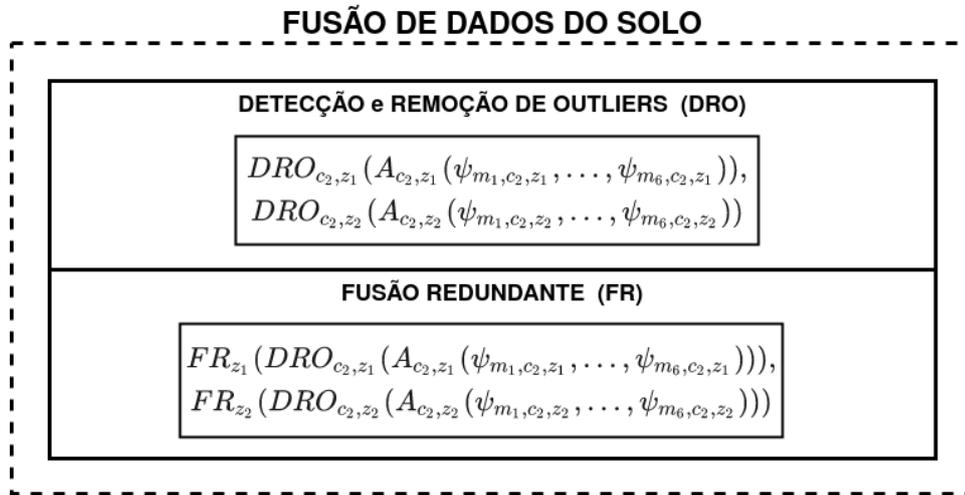


Figura 7 – Principais Funções do Serviço de Fusão de Dados do Solo. Fonte: Autora.

A **Fusão Redundante (FR)** fusiona um mesmo **tipo** de dado de **locais diferentes** dentro de um mesmo **cenário** (NAKAMURA *et al.*, 2007). Por exemplo, no Campo 2 (cenário), monitora-se o dado de potencial mátrico ψ_m do solo em duas camadas diferentes, z_1 e z_2 em seis pontos de monitoração (locais) diferentes. Assim, para cada camada z , a implementação de FR agrega o conjunto de dados em um só valor de potencial mátrico $\psi_{f,z}$. O *framework* traz como fusão redundante a implementação da média, recomendada por Torres (2017), e as técnicas mínimo, mediana e máximo. Logo, para o Campo 2, o resultado da fusão consiste em dois valores de potencial mátrico:

- $\psi_{f,z_1} = MEDIA(DRO_{c_2,z_1}(A_{c_2,z_1}(\psi_{m_1,c_2,z_1}, \dots, \psi_{m_6,c_2,z_1})))$
- $\psi_{f,z_2} = MEDIA(DRO_{c_2,z_2}(A_{c_2,z_2}(\psi_{m_1,c_2,z_2}, \dots, \psi_{m_6,c_2,z_2})))$

O valor fusionado de potencial mátrico ψ_f são usados para deduzir a umidade do solo θ pelo modelo de Van Genuchten (Eq. A.1) para cada camada z monitorada. A decisão de irrigar ou não o campo é função dos manejos de irrigação que usam dados de umidade do solo $\theta(\psi_f)$ (Eq. A.6 e A.12).

O *framework* também recomenda o uso em conjunto de técnicas de remoção de *outliers* e fusão redundante de dados de umidade do solo que poderá otimizar a relação custo-benefício entre a economia da irrigação total necessária fornecida pelos manejos de irrigação do Potencial Mátrico e Balanço Hídrico e o desenvolvimento saudável do cultivo (detalhes da análise no Capítulo 5).

3.4 Camada de Comunicação

A camada de Comunicação reúne um conjunto de software para obter via protocolos de comunicação em redes de computadores dados meteorológicos e do solo para o serviço de Armazenamento do *framework* (Figura 6). Ela contém um cliente de dados meteorológicos, quatro clientes e três servidores de dados do solo e um gateway.

O **cliente de dados meteorológicos** consiste em um cliente HTTP que baixa dados de estações meteorológicas próximas da propriedade rural e os armazena no banco de dados local do *framework*. No caso das estações nacionais brasileiras (INMET, 2018), esse cliente web utiliza técnicas de *web scrapping* já que não existe API oficial para o download do conjunto de dados M_{ms} (definido na Seção A.1).

O **cliente de dados do solo** coleta dados de potencial mátrico ψ_m quando existem tensiômetros no campo monitorado. Para tensiômetros analógicos, uma **aplicação móvel para coleta de dados offline** auxilia os usuários na inserção de dados em campos sem nós sensores (ex. Campo 2 da Figura 1). Ele permite o envio posterior dos dados ao servidor de dados HTTP para que este os armazenem no banco de dados do *framework*.

Em campos com nó sensores (ex. Campo 3 da Figura 1), os dados do solo são enviados para um gateway, que consiste aplicação executando em uma máquina próxima ao campo monitorado. Ao receber os dados do solo, o gateway invoca um cliente MQTT *Publisher* ou um cliente CoAP *Observer* para enviar tais dados ao servidor adequado do *framework*, e este, por sua vez, possa armazená-los.

As aplicações clientes de dados do solo usam um arquivo de configuração (json) fornecido pelo módulo de cadastro de informações agrícolas do *framework* (Seção 3.2.1) para auxiliar na construção do *payload* das mensagens. O arquivo de configuração contém:

1. `farm_id`: identificador da propriedade rural no banco de dados do *framework*;
2. `field_id`: identificador do campo agrícola no banco de dados do *framework*;
3. `monitoring_point_number`: número de pontos de monitoramento,
4. `mp_info`: informações sobre cada ponto de monitoramento, como identificador do ponto, profundidade de instalação dos sensores umidade do solo e suas respectivas camadas de solo monitoradas;
5. `coap_server_address`: endereço IP ou URL do servidor CoAP;
6. `mqtt_server_address`: endereço IP ou URL do servidor MQTT;
7. `http_server_address`: endereço IP ou URL do servidor HTTP;

8. `mqtt_topic`: tópico para publicação do payload de uma mensagem MQTT.

Para um ponto de monitoramento de um campo, o *payload* de dados de umidade do solo consiste em:

1. `farm_id`: identificador da propriedade rural no banco de dados do *framework*;
2. `field_id`: identificador do campo agrícola no banco de dados do *framework*;
3. `mp_id`: identificador do ponto de monitoramento do campo agrícola;
4. `matric_potential`: valores de potencial mátrico ψ_m associados às profundidades de instalação dos tensiômetros do ponto de monitoramento;
5. `power_level`: nível de energia do nós sensor (apenas para campos IoT como o Campo 3 da Figura 1).

3.5 Síntese do Capítulo

O *framework* Smart&Green pode ser classificado como um conjunto de objetos inteligentes virtuais/digitais da agricultura inteligente, pois, é capaz de prover serviços necessários aos manejos de irrigação (Apêndice A) em tempo real através da coleta de dados/informações de diferentes fontes associadas a diferentes infraestruturas de campos (Figura 1). Assim, este trabalho contribui parcialmente na solução do desafio/limitação da implantação da Internet das Coisas (*Internet of Things* — IoT) relacionado a integração dos códigos IoT com a infraestrutura existente (equipamentos e máquinas agrícolas, como também com software legado) (TALAVERA *et al.*, 2017).

Os serviços de monitoramento (camada de Comunicação), sincronização, pré-processamento que trata da remoção de *outliers* e fusão redundante de umidade do solo (camada de Serviços), predição de umidade do solo são modulares e podem ser reusados por qualquer método de manejo de irrigação (camada de Aplicação) e estendidos com outras técnicas de pré-processamento, protocolos de comunicação e modelos computacionais. Desta maneira, este trabalho contribui na solução de outros desafios e limitações da implantação da IoT na agricultura, tais, como:

- o projeto de soluções de software modularizadas para propiciar reuso e customização (PANG *et al.*, 2015);
- a adoção de boas práticas de engenharia de software (projeto e documentação) para que não se aumente o tempo e esforço dedicados à análise de dados gerados, re-fatoração de código e adição de novas funcionalidades quando a solução implantada crescer em escala

e durabilidade (JAYARAMAN *et al.*, 2015).

4 AVALIAÇÃO DOS MODELOS COMPUTACIONAIS PARA PREDIÇÃO DE UMIDADE DO SOLO

Neste capítulo, são apresentados os materiais e métodos usados na avaliação dos modelos computacionais de previsão de umidade do solo (Seção 4.1). Em seguida, foram avaliados o MAE (*Mean Absolute Error*) e o RMSE (*Root Mean Squared Error*) das técnicas de aprendizagem de máquina testadas segundo as abordagens de previsão local e global (Seção 3.2.2) para a implementação do módulo de previsão de umidade de solo do Smart&Green (Seção 4.2).

4.1 Materiais e Métodos para Avaliação dos Modelos Computacionais

Nesta seção, são descritos os materiais e a metodologia de treinamento de modelos computacionais para o módulo de previsão de umidade de solo do Smart&Green em termos de conjunto de dados climáticos, de cultivo e umidade do solo (Seção 4.1.1), critérios para detecção e remoção imediata de *outliers* e engenharia de atributos (Seção 4.1.2) e técnicas de aprendizagem de máquina (Seção 4.1.3).

4.1.1 Conjunto de Dados do Campo Experimental da Embrapa

O conjunto de dados de umidade do solo foi obtido de um campo experimental de caju e outras culturas na cidade de Paraipaba, Ceará, Brasil (EMBRAPA, 2019). O campo possui mais de 250 cajueiros com nove pontos de monitoramento. Cada ponto possui três tensiômetros analógicos que permitem a leitura do potencial mátrico a 15, 45 e 75 cm de profundidade. A Figura 8 mostra detalhes de tensiômetros analógicos (sensores do solo) usados para compor o conjunto de dados de 234 amostras diárias do ano de 2016, 245 do ano de 2017 e 138 amostras do ano de 2018. Para prevenir o estresse hídrico em 2016, o campo teve a irrigação de salvação, na qual, foram fornecidos 5 litros de água para cada muda toda vez que o tensiômetro à 15 cm de profundidade detectava um potencial mátrico maior ou igual a 60 kPa.

O conjunto de dados meteorológicos é de uma estação climática da cidade de Itapipoca, Ceará (altitude 102 m; latitude 03°29'S; longitude 39°35'W), à 48,8 km de Paraipaba. No entanto, decidiu-se usar os dados do sensor pluviométrico do campo experimental, pois, o sensor de precipitação da estação climática falhou durante um grande período de tempo. A Tabela 6 relaciona os acrônimos, as descrições dos atributos e as unidades de medida dos dados usados

no problema de predição. O conjunto de dados foi divulgado ¹ para a comunidade científica a fim de assegurar a reprodução dos resultados e a promoção de desenvolvimento de pesquisa nessa área.



(a)



(b)

Figura 8 – Detalhes do campo experimental pertencente à Embrapa Agroindústria Tropical (EMBRAPA, 2019) – (a) tensiômetros a 15, 45 e 75 cm de profundidade do solo perto de uma muda de cajueiro e (b) mudas de caju.

Tabela 6 – Estrutura do Conjunto de Dados do Campo Experimental

Acrônimo	Descrição	Unidade
T _{x-y}	Leitura do tensiômetro na profundidade de {y} cm no ponto de monitoramento {x}	kPa
W _x	A quantidade de água irrigada no campo em um ponto de monitoramento {x}	L
T _{max}	Temperatura do ar máxima	°C
T _{min}	Temperatura do ar mínima	°C
RH _{max}	Umidade relativa do ar máxima	%
RH _{min}	Umidade relativa do ar mínima	%
R _n	Radiação solar líquida	MJm^2d^{-1}
U ₂	Velocidade do vento	m/s
P	Pressão atmosférica	kPa
Ri _f	Precipitação pluvial coletada no campo experimental	mm
K _c	Coefficiente do cultivo	

¹ *Experimental Cashew Field Dataset*. Disponível em: <http://smartgreen.great.ufc.br/publications>.

A Tabela 7 relaciona as constantes obtidas em laboratório e que caracterizam a curva de retenção de água do solo do campo experimental. Elas são usadas no modelo de Van Genuchten (Eq. A.1) para converter o potencial mátrico ψ_m (kPa) em umidade do solo θ (mm), que é usada nos manejos de irrigação do Potencial Mátrico (Eq. A.12) e Balanço Hídrico (Eq. A.7).

Tabela 7 – Contantes de Van Genuchten do Solo do Campo Experimental

θ_r	θ_s	α	n
0.14010	0.38839	0.022504	20.524

4.1.2 Métodos de Preprocessamento dos Dados

Detecção e Remoção de Outliers. Primeiro, as instâncias que apresentam $T_{max}, T_{min} < 20$ e $T_{max}, T_{min} > 39$, $RH_{max}, RH_{min} < 20$, e $R_i > 250$ foram removidas, pois, essas faixas de atributos são incomuns na cidade de Paraipaba. Depois disto, a média μ e o desvio padrão σ foram calculados para os atributos P , Rn , U_2 para cada mês. Em seguida, semelhante ao método de Magalhães (2018), foram filtradas as instâncias de dados com anomalias, isto é, foram removidas as instâncias que se distanciam da média em valores maiores ou iguais a $3 \times \sigma$.

Engenharia de Atributos. Algoritmos de seleção de atributos foram aplicados para encontrar os atributos que são mais relevantes ao modelo de predição, de tal maneira que, o modelo a ser criado a partir dos atributos selecionados tenha baixa taxa de erros. Os modelos foram treinados com **quatro** diferentes grupos de atributos escolhidos a partir dos algoritmos na Tabela 8 que são usados amplamente e disponíveis pela ferramenta Weka (HALL *et al.*, 2009).

Tabela 8 – Grupos de Atributos

Grupo	Algoritmo de Seleção de Atributos
1	Nenhum
2	<i>ScatterSearchV1</i> OU <i>Genetic Search</i> OU <i>BestFirst</i> OU <i>GreedyStepwise</i> OU <i>Exhaustive Search</i>
3	<i>Random Search</i>
4	Coefficiente de Correlação de Pearson Positivo

A Tabela 9 mostra os atributos selecionados. O Grupo 1 apresenta todos os atributos, pois, nenhum algoritmo de seleção de atributos foi aplicado. O Grupo 2 é um conjunto de atributos escolhidos por qualquer um dos algoritmos relacionados na Tabela 8. Independente da

execução de qualquer algoritmo do Grupo 2, os resultados foram sempre o mesmo conjunto de atributos (Tabela 9). O Grupo 3 é outro grupo de atributos de dados selecionados pelo *Random Search*. E, finalmente, o Grupo 4 apresenta os atributos que têm um coeficiente de correlação de Pearson positivo em relação ao potencial mátrico coletado pelo tensiômetro instalado à 15cm de profundidade do solo (o atributo saída dos modelos de predição, Apêndice B).

Tabela 9 – Grupo de Atributos Selecionados

Tensiômetro	Grupo 1	Grupo 2	Grupo 3	Grupo 4
1	Todos atributos	W, T_min, U2, Kc	W, Rn, Kc	Kc, T_max, T_min, U2, P
2	Todos atributos	W, T_min, U2, Kc	W, T_min, Rn, Kc, P, U2	Kc, T_max, T_min, U2, P, Rn, W
3	Todos atributos	W, T_min, U2, Kc	W, T_min, RH_max, U2, Kc	Kc, T_max, T_min, P, Rn, W
4	Todos atributos	W, T_min, U2, Kc	W, T_min, U2, Kc	Kc, T_max, T_min, P, Rn, W
5	Todos atributos	W, T_min, U2, Kc	W, T_min, RH_max, U2, Kc	Kc, T_max, T_min, Rn, W
6	Todos atributos	W, T_min, U2, Kc	W, T_min, Rn, Kc, P, U2	Kc, T_max, T_min, U2, P, Rn
7	Todos atributos	W, T_min, U2, Kc	W, T_min, Rn, Kc, P, U2	Kc, T_max, T_min, U2
8	Todos atributos	W, T_min, U2, Kc	W, T_min, U2, Kc	Kc, T_max, T_min, Rn
9	Todos atributos	W, T_min, U2, Kc	W, T_min, RH_max, U2, Kc	Kc, T_max, T_min, U2, P, Rn, W

4.1.3 Métodos de Aprendizagem de Máquina.

Os modelos de predição foram construídos empregando os algoritmos amplamente usados: Regressão Linear (FREEDMAN, 2009), *Decision Stump* (IBA; LANGLEY, 1992), M5P — *M5 Model Trees and Rules* (QUINLAN, 1992; WANG; WITTEN, 1997), *Random Tree*, *Random Forest* (BREIMAN, 2001), e *RepTree*. Todos são providos pelo Weka (HALL *et al.*, 2009). Outro algoritmo amplamente usado é o *Gradient Boosting Regression Tree* — GBRT (FRIEDMAN, 2000; FRIEDMAN, 2002), cuja implementação é provida pela biblioteca Python XGBoost (CHEN; GUESTRIN, 2016).

A Regressão Linear combina um conjunto de variáveis de entrada (neste trabalho M_{ms}, w_i, k_c do Apêndice B) em uma equação que estima uma variável de saída (potencial mátrico ψ_{15} coletado por tensiômetros à 15 cm de profundidade). Esta equação modela o relacionamento entre as variáveis de entrada (chamadas de variáveis explanatórias ou independentes) e a variável de saída (também chamada de resposta escalar ou variável independente) (FREEDMAN, 2009).

Decision Stump ou *One-Level Decision Tree* (árvore de decisão de um nível) seleciona um dos atributos de entrada (M_{ms}, w_i, k_c) para a predição do atributo de saída (ψ_{15}). A árvore tem uma raiz (nó interno) que conecta-se às folhas (nós terminais). Cada nó interno representa um teste realizado em um atributo de entrada, cada ramo representa a saída do teste e cada folha representa uma decisão tomada depois do processamento de todos os atributos de entrada. O teste dá escores a um atributo quanto maior for a probabilidade dos valores assumidos por esse

atributo pertecerem ao domínio do atributo de saída. O *Decision Stump* escolhe o atributo com maior pontuação e, no caso de empates, o algoritmo escolhe aleatoriamente um dos atributos com melhor pontuação (IBA; LANGLEY, 1992).

O M5P constrói um modelo baseado em uma árvore, cujos nós terminais são funções de regressão linear. Esse modelo de regressão múltipla é ajustado por um algoritmo guloso (*greedy algorithm*) que busca e remove atributos que contribuem pouco para que o modelo obtenha um erro de predição baixo. No M5P, o processo de suavização (*smoothing process*) faz com que um nó terminal realize a predição de um novo valor considerando os valores da predição feita por outros nós intermediários que localizam-se entre aquele nó terminal e o nó raiz da árvore. Assim, essa suavização melhora a acurácia de predição da árvore do M5P, tendo mais efeito quando os modelos (nós) ao longo do caminho estimam valores muito diferentes (QUINLAN, 1992; WANG; WITTEN, 1997).

O *Random Tree* implementado pelo Weka (HALL *et al.*, 2009) constrói uma árvore que usa K atributos escolhidos aleatoriamente em cada nó. No entanto, esse algoritmo não realiza a poda (*prunning*), isto é, a técnica de redução do tamanho da árvore através da remoção de seções da árvore que não contribuem para um erro de predição baixo. O *RepTree* é um algoritmo de árvore de decisão rápido do Weka (HALL *et al.*, 2009). O modelo usa informação de ganho/variância para construir uma árvore de regressão, como também, a poda com erro reduzido levando em consideração o método de *backfitting*.

Random Forest é um algoritmo que usa o *bagging* como método de *ensemble* e a árvore de decisão como modelo individual. Na aprendizagem por *ensemble*, o modelo realiza predições combinando a predição de outros modelos, por isso, o *Random Forest* consiste em uma coleção de árvores. O método *bagging* treina um conjunto de modelos individuais de modo paralelo dividindo aleatoriamente o dataset em vários n subconjuntos com os quais treina n árvores de decisões. Como cada árvore torna-se um *learner* (aprendiz) com dataset individual e, assim, realiza predição com baixa correlação, esse algoritmo limita os erros devido *bias* (enviesamento) ou da variância (BREIMAN, 2001).

Para fazer as predições, o GBRT usa árvores de decisão como *weak learners* (aprendizes fracos) e o método do *boosting*, no qual, treina-se um grupo de *weak learners* de maneira individual e sequencial de modo que um *weak learner* aprenda com o erro do outro. Logo, uma árvore gera um valor de saída que serve para a divisão dos ramos e também para a correção dos erros residuais na predição de árvores já existentes no modelo. O GBRT adiciona uma

árvore de regressão por vez e não modifica aquelas que já estão no modelo. A nova árvore é parametrizada com o procedimento do gradiente descendente funcional (*functional gradient descent*). Isto possibilita a redução da perda calculada pelo erro quadrático médio para problemas de regressão (*loss function*). A saída de uma nova árvore é adicionada à saída da sequência de árvores existentes para corrigir ou otimizar a saída (predição) do modelo gerado pelo GBRT (FRIEDMAN, 2000; FRIEDMAN, 2002).

Alguns hiperparâmetros são necessários para o *Random Tree*, *Random Forest*, *Rep-Tree* e *GBRT* e foram configurados com valores padrão do Weka (HALL *et al.*, 2009) e do XGboost (CHEN; GUESTRIN, 2016). A profundidade máxima da árvore (*max_depth*) foi configurada em 3 para GBRT e deixada sem restrição para *Random Tree*, *Random Forest* e *RepTree*. O número de árvores (*n_estimators*) é 100 para todos os métodos. Finalmente, o GBRT requer um terceiro hiperparâmetro: a taxa de aprendizagem (*learning_rate*) que foi configurada em 0.1.

Os conjuntos de treinamento e teste incluem, respectivamente, 80% e 20% dos dados de cada tensiômetro instalado à 15 cm de profundidade do solo (i.e., dados da camada mais superficial monitorada). Então, para cada grupo de atributos (Tabela 9), foi gerado um modelo de cada técnica de aprendizagem de máquina descrito. Foi escolhido como melhor aquele modelo que produz o menor MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*) que são métricas amplamente usadas para medir a qualidade de modelos de predição de potencial mátrico ψ dado em kPa. Neste trabalho, a umidade do solo θ dada em cm^3/cm^3 usada nos manejos de irrigação é estimada a partir do potencial mátrico ψ (Eq. A.1).

4.2 Resultados da Avaliação dos Modelos Computacionais

Nesta seção, são avaliadas e escolhida a melhor técnica de aprendizagem de máquina entre aquelas testadas para abordagem local (Seção 4.2.1) e global (Seção 4.2.2) de predição do potencial mátrico ψ dado em kPa (Apêndice B). Também é analisada a relevância de atributos da melhor técnica de aprendizagem de máquina (Seção 4.2.3).

4.2.1 Avaliação do Modelo da Abordagem de Predição Local

As Tabelas 10, 11, 12, e 13 apresentam os valores de MAE (*Mean Absolute Error*) e RMSE (*Root Mean Squared*) obtidos pela abordagem local usando cada grupo de atributos

e algoritmo para predição de potencial mátrico ψ em kPa. Os resultados da abordagem local são a média de MAE e RMSE obtidos pelos nove modelos construídos para cada ponto de monitoramento do campo experimental.

Tabela 10 – Abordagem Local, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 1 – média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1408	(0,1318; 0,1498)	0,1730	(0,1642; 0,1818)
Decision Stump	0,1798	(0,1632; 0,1965)	0,2196	(0,2031; 0,2360)
M5P	0,1288	(0,1159; 0,1416)	0,1722	(0,1576; 0,1868)
Random Tree	0,1443	(0,1319; 0,1567)	0,2120	(0,1877; 0,2363)
Random Forest	0,1189	(0,1025; 0,1352)	0,1551	(0,1393; 0,1709)
RepTree	0,1227	(0,1119; 0,1336)	0,1684	(0,1566; 0,1801)
Gradient Boosting	0,0752	(0,0683; 0,0822)	0,1038	(0,0939; 0,1137)

Tabela 11 – Abordagem Local, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 2 – média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1687	(0,1523; 0,1852)	0,2070	(0,1887; 0,2254)
Decision Stump	0,1791	(0,1613; 0,1970)	0,2212	(0,2049; 0,2374)
M5P	0,2095	(0,1750; 0,2440)	0,2557	(0,2205; 0,2910)
Random Tree	0,1880	(0,1670; 0,2089)	0,2573	(0,2312; 0,2834)
Random Forest	0,1512	(0,1355; 0,1668)	0,2020	(0,1837; 0,2203)
RepTree	0,1524	(0,1390; 0,1658)	0,1976	(0,1837; 0,2115)
Gradient Boosting	0,1129	(0,1031; 0,1227)	0,1512	(0,1371; 0,1653)

Tabela 12 – Abordagem Local, Avaliação dos Algoritmos para Predição de Potencial Mátrico ψ_p quando usam os Atributos do Grupo 3 – média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1641	(0,1472; 0,1811)	0,1973	(0,1783; 0,2163)
Decision Stump	0,1791	(0,1613; 0,1970)	0,2214	(0,2051; 0,2376)
M5P	0,1909	(0,1499; 0,2318)	0,2148	(0,1714; 0,2582)
Random Tree	0,1867	(0,1595; 0,2139)	0,2578	(0,2218; 0,2937)
Random Forest	0,1417	(0,1167; 0,1666)	0,1845	(0,1562; 0,2127)
RepTree	0,1464	(0,1269; 0,1659)	0,1881	(0,1656; 0,2105)
Gradient Boosting	0,1058	(0,0880; 0,1236)	0,1466	(0,1260; 0,1672)

Considerando todos os grupos, o *Gradient Boosting* supera todas as outras técnicas seguido pelo *Random Forest*. Ambas técnicas apresentam erros aceitáveis de acordo com um agrônomo especialista em agricultura de precisão. Grupo 1 obteve os melhores modelos de

predição com os mais baixos erros, logo, conclui-se que um bom modelo de predição para a abordagem local precisa usar todos os atributos.

Tabela 13 – Abordagem Local, Avaliação dos Algoritmos para Predição de Potencial Mátrico ψ_p quando usam os Atributos do Grupo 4– média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1472	(0,1372; 0,1573)	0,1802	(0,1699; 0,1904)
Decision Stump	0,1778	(0,1601; 0,1955)	0,2204	(0,2040; 0,2367)
M5P	0,1359	(0,1234; 0,1483)	0,1755	(0,1622; 0,1888)
Random Tree	0,1638	(0,1463; 0,1814)	0,2327	(0,2096; 0,2558)
Random Forest	0,1166	(0,1046; 0,1287)	0,1579	(0,1416; 0,1741)
RepTree	0,1299	(0,1190; 0,1408)	0,1731	(0,1583; 0,1880)
Gradient Boosting	0,0880	(0,0799; 0,0961)	0,1195	(0,1077, 0,1312)

4.2.2 Avaliação do Modelo da Abordagem de Predição Global

As Tabelas 14, 15, 16, e 17 apresentam os valores de MAE (*Mean Absolute Error*) e RMSE (*Root Mean Squared Error*) obtidos pela abordagem global para cada grupo de atributos e algoritmo para predição de potencial mátrico ψ dado em kPa. Novamente, o algoritmo *Gradient Boosting* obtém os melhores resultados quando comparados ao dos outros modelos avaliados. Ele alcança os mesmos valores de erro para os Grupos de atributos 1 e 4.

Tabela 14 – Abordagem Global, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 1 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1628	(0,1510; 0,1746)	0,1993	(0,1881; 0,2110)
Decision Stump	0,1938	(0,1812; 0,2063)	0,2335	(0,2220; 0,2450)
M5P	0,1461	(0,1348; 0,1573)	0,1824	(0,1706; 0,1942)
Random Tree	0,1494	(0,1413; 0,1574)	0,2094	(0,2004; 0,2183)
Random Forest	0,1406	(0,1317; 0,1494)	0,1873	(0,1769; 0,1977)
RepTree	0,1438	(0,1362; 0,1515)	0,1832	(0,1740; 0,1924)
Gradient Boosting	0,1382	não tem	0,1717	não tem

Percebe-se que o MAE e RMSE do modelo *Gradient Boosting* não varia como os outros modelos. Isto significa que o modelo global generaliza bem os dados de qualquer sensor do campo. A abordagem local do *Gradient Boosting* supera a abordagem global. Mesmo que a abordagem global aumente em 58,9% a média de MAE e em 45,7% a média de RMSE, os erros ainda são aceitáveis.

Tabela 15 – Abordagem Global, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 2 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1999	(0,1824; 0,2174)	0,2343	(0,2144; 0,2542)
Decision Stump	0,1998	0,1822; 0,2174)	0,2430	(0,2249; 0,2611)
M5P	0,3524	(0,3333; 0,3715)	0,6440	(0,5912; 0,6969)
Random Tree	0,1509	(0,1411; 0,1608)	0,1977	(0,1842; 0,2111)
Random Forest	0,1480	(0,1380; 0,1579)	0,1891	(0,1770; 0,2013)
RepTree	0,1641	(0,1536; 0,1745)	0,2100	(0,1979; 0,2221)
Gradient Boosting	0,1684	(0,1684; 0,1684)	0,2070	(0,2070; 0,2070)

Tabela 16 – Abordagem Global, Avaliação dos Algoritmos para Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 3 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1937	(0,1767; 0,2107)	0,2368	(0,2207; 0,2530)
Decision Stump	0,1998	(0,1822; 0,2174)	0,2430	(0,2249; 0,2611)
M5P	0,2685	(0,2569; 0,2801)	0,3735	(0,3578; 0,3891)
Random Tree	0,1496	(0,1397; 0,1595)	0,1952	(0,1814; 0,2089)
Random Forest	0,1478	(0,1376; 0,1580)	0,1893	(0,1766; 0,2019)
RepTree	0,1611	(0,1496; 0,1726)	0,2082	(0,1939; 0,2225)
Gradient Boosting	0,1653	(0,1653; 0,1653)	0,2031	(0,2031; 0,2031)

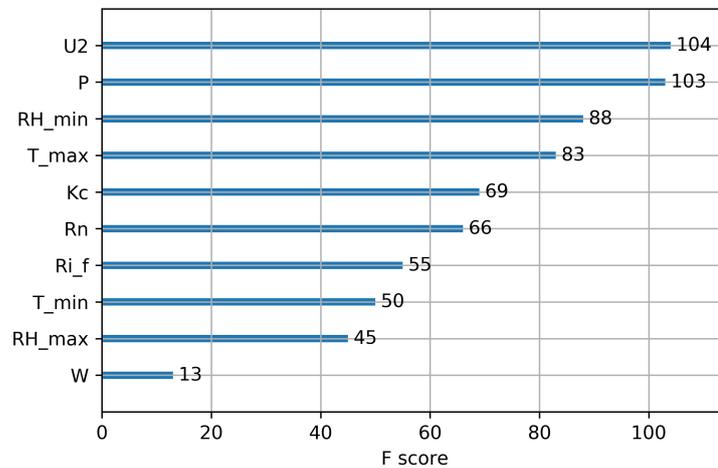
Tabela 17 – Abordagem Global, Avaliação dos Algoritmos de Predição do Potencial Mátrico ψ_p quando usam os Atributos do Grupo 4 - média de MAE e RMSE, unidade kPa, intervalo de confiança de 99%. Os melhores resultados estão destacados em **negrito**.

Algoritmo	MAE	Inter. Conf. MAE	RMSE	Inter. Conf. RMSE
Regressão Linear	0,1644	(0,1525; 0,1762)	0,2017	(0,1902; 0,2132)
Decision Stump	0,1938	(0,1812; 0,2063)	0,2335	(0,2220; 0,2450)
M5P	0,1464	(0,1339; 0,1589)	0,1840	(0,1712; 0,1968)
Random Tree	0,1495	(0,1405; 0,1584)	0,2093	(0,1968; 0,2219)
Random Forest	0,1409	(0,1316; 0,1501)	0,1887	(0,1777; 0,1997)
RepTree	0,1457	(0,1345; 0,1569)	0,1851	(0,1725; 0,1976)
Gradient Boosting	0,1382	não tem	0,1717	não tem

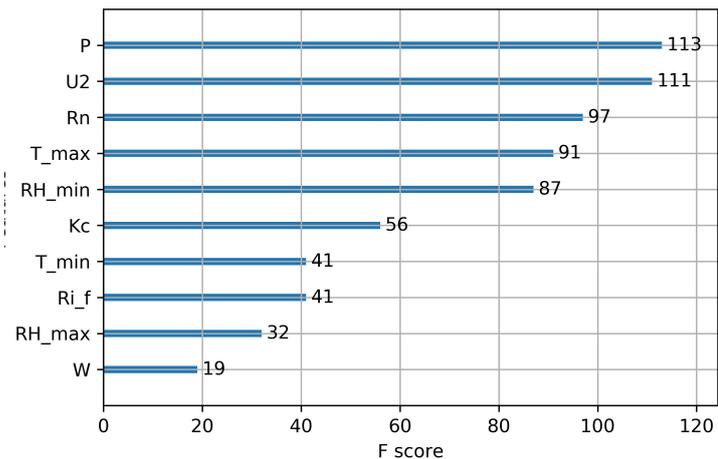
4.2.3 Relevância de atributos

A relevância de atributos do conjunto de dados (Tabela 6) foi introduzida na Seção 4.1.2. Para demonstrá-la, foram usados os modelos das abordagens de predição local e global (Seção 3.2.2) criados pelo *Gradient Boosting*, pois, obtiveram os menores MAE (*Mean Absolute Error*) e RMSE (*Root Mean Squared Error*) na predição do potencial mátrico ψ usado para a dedução da umidade do solo θ (Eq. A.1). A relevância de cada atributo é estimada pela contagem de vezes que ele é usado na divisão de nós de qualquer uma das árvores de decisão da floresta

criada pelo *Gradient Boosting* (CHEN; GUESTRIN, 2016). A Figura 9 apresenta os resultados. Quanto mais um atributo é usado nas árvores de decisão, maior é a sua importância. Ambos modelos de predição usam mais os atributos P , U_2 , RH_{min} e T_{max} .



(a)



(b)

Figura 9 – Análise da relevância de atributos – abordagem de predição (a) Local e (b) Global.

4.3 Síntese do Capítulo

Um conjunto de dados do clima, do solo, do cultivo, da irrigação do Campo Experimental do Curu (EMBRAPA, 2019) foi organizado para o treinamento das técnicas de aprendizagem de máquina e disponibilizado para a comunidade científica². Foram avaliadas as seguintes técnicas, visto que, são amplamente usadas na solução de problemas de regressão: Regressão Linear (FREEDMAN, 2009), *Decision Stump* (IBA; LANGLEY, 1992), M5P — M5

² *Experimental Cashew Field Dataset* disponível em: <http://smartgreen.great.ufc.br/publications>

Model Trees and Rules (QUINLAN, 1992; WANG; WITTEN, 1997), *Random Tree*, *Random Forest* (BREIMAN, 2001), *RepTree* (HALL *et al.*, 2009) e o *Gradient Boosting Regression Tree* — GBRT (FRIEDMAN, 2000; FRIEDMAN, 2002). A qualidade dos modelos foi medida sem e com a seleção de atributos usando o RMSE (*Root Mean Squared Error*) e MAE (*Mean Absolute Error*) como métricas. Os resultados mostram que o GBRT obteve os menores RMSE e MAE, logo, foi utilizado na implementação das abordagens local e global de predição de umidade do solo do Smart&Green (Seção 3.2.2).

5 AVALIAÇÃO DOS MANEJOS DE IRRIGAÇÃO DO SMART&GREEN

O *framework* Smart&Green implementa os tradicionais manejos de irrigação (Apêndice A) beneficiados pelo pré-processamento de dados do serviço de Fusão (Seção 3.3.3) e predição de umidade de solo (Seção 3.2.2). Neste capítulo, são listados os principais componentes e bibliotecas (Seção 5.1) usados no desenvolvimento do *framework*, bem como, materiais e métodos de avaliação dos manejos de irrigação do Smart&Green (Seção 5.2). É analisada a influência das técnicas do serviço de Fusão nos manejos de irrigação do Potencial Mátrico (Seção 5.3) e Balanço Hídrico (Seção 5.4) quando usam dados de umidade do solo reais ou estimados pelas abordagens de predição local e global.

5.1 Implementação do framework Smart&Green

O projeto Smart&Green (ROCHA, 2017) tem como objetivo conectar campos agrícolas à Internet por meio de tecnologia de baixo custo. O campo equipado com sistemas microcontrolados, sensores e atuadores, torna-se um objeto da Internet das Coisas (BORGIA, 2014) capaz de interagir com outros semelhantes e prestar **serviços** de maneira transparente aos seus usuários. Esses serviços podem estar relacionados a vários setores da agricultura de precisão: como **monitoramento** (água, solo, ar), **controle** (fertilizantes, pesticidas, iluminação, acesso a propriedade), **predição** (condições climáticas/solo, estimação de produção, crescimento das culturas) e logística (produção, comércio e transporte) (TALAVERA *et al.*, 2017).

O *framework* proposto neste trabalho compreende parte de duas metas do projeto Smart&Green:

1. **Especificar, modelar e implementar aplicações do domínio de agricultura de precisão** relacionadas aos serviços de **monitoramento** de dados climáticos, do cultivo e do solo para o **controle** de água fornecida a diferentes configurações de campos agrícolas de acordo com algum dos tipos de manejos de irrigação estudados. A **predição da umidade no solo** também é tratada.
2. **Construir uma camada de software para análise, processamento e inferência dos dados para suporte de tomada de decisão** relacionado aos diferentes manejos de irrigação cultivos, condições climáticas e do solo.

Nós implementamos o *framework* Smart&Green na linguagem Python 3. Na camada de aplicação (Seção 3.2), o manejo de irrigação do Balanço Hídrico usa a biblioteca PyETo (RI-

CHARDS, 2015) para calcular a evapotranspiração de referência para um cultivo de acordo com o modelo de Penman-Monteith (Eq. A.5). A camada de Serviço usa o banco de dados MySQL (ORACLE, 2019) como base para o serviço de Armazenamento. A biblioteca PyAstronomy (CZESLA, 2013) prover a implementação do GESD (*Generalized Extreme Studentized Deviate*) para o serviço de Fusão.

O pacote aiocoap (WASILAK; AMSUSS, 2012) suporta a implementação do cliente e servidor CoAP na camada de Comunicação (Seção 3.4). O cliente MQTT usa a biblioteca Paho MQTT da Eclipse (LIGHT, 2013) e, como servidor MQTT, adotamos o Mosquitto (LIGHT, 2017). Nossa API de aquisição de dados climáticos faz *web scrapping* da página web das estações meteorológicas brasileiras (INMET, 2018) usando bibliotecas como a requests (REITZ, 2011) e a BeautifulSoup4 (RICHARDSON, 2014) e, também, os *frameworks* Django (FOUNDATION, 2013) e Django Rest (ENCODE, 2011). Nós implementamos a aplicação móvel para coleta de umidade do solo em campos usando tensiômetros analógicos (ex. Campo 2 da Figura 1) em JavaScript através da plataforma Firebase (GOOGLE, 2016), do *framework* React Native (FACEBOOK, 2018), das bibliotecas Realm (REALM, 2014) e React Native Firebase (INVERTASE, 2016).

Nós também provemos uma interface web para o *framework* Smart&Green (COSTA, 2018) desenvolvida com o Django (FOUNDATION, 2013). Para campos como nós sensores como o campo 3 da Figura 1, nós usamos um Raspberry 3 com o Raspbian OS como o gateway que recebe dados de umidade do solo e usa clientes CoAP ou MQTT para enviá-los ao serviço de Armazenamento. O Smart&Green roda em uma máquina com 2GB de RAM, 20 GB de disco rígido e Ubuntu 18.04LTS.

5.2 Materiais e Métodos da Avaliação de Manejos de Irrigação

Os manejos do Balanço do Hídrico (Eq. A.7) e Potencial Mátrico (Eq. A.12) consomem dados de umidade do solo θ para calcularem a irrigação total necessária (ITN), isto é, a quantidade de água fornecida por um sistema de irrigação a um plantio. O desempenho do manejo de irrigação ao utilizar o Serviço de Fusão é avaliado em termos de quantidade (mm) e economia (%) de ITN quando são utilizados dados de umidade solo: 1) reais θ_c e 2) estimados θ_p por um modelo de predição de abordagem global ou local (Seção 3.2.2) baseado no *Gradient Boosting Regression Tree* — o melhor método preditivo avaliado (Seção 4.2).

Neste trabalho, os dados de umidade do solo são obtidos a partir do modelo de Van

Genutchen (Eq. A.1) que utiliza dados de potencial mátrico ψ e as constantes de propriedades do solo (Tabela 7) do Campo Experimental (EMBRAPA, 2019). Para os dados de umidade do solo reais $\theta_c(\psi_c)$, foram usados valores de potencial mátrico ψ_c de nove tensiômetros instalados à 15 cm de profundidade do solo do Campo Experimental (Seção 4.1.1). Para os dados de umidade do solo estimados $\theta_p(\psi_p)$ pelos modelos de predição, foram fornecidos dados climáticos, do cultivo e da irrigação (Apêndice B) do conjunto de dados do Campo Experimental. A abordagem de predição global estima apenas um valor de potencial mátrico ψ_p e a abordagem local, nove valores de ψ_p .

Os dados climáticos, do cultivo e da irrigação foram selecionados do conjunto de dados (Seção 4.1.1) do Campo Experimental de abril a outubro de 2017, pois, nesse período, o Campo Experimental não recebeu a irrigação da salvação, a qual não é considerada pelos manejos avaliados. Os dados de umidade do solo usados nesta avaliação são relacionados à zona radicular do cultivo entre 0 e 30 cm de profundidade da superfície do solo ($z = 30$ cm).

O método de avaliação de desempenho dos manejos de irrigação inicia com a execução do Serviço de Fusão do *framework* Smart&Green. Primeiro, os dados de potencial mátrico reais ψ_c ou estimados ψ_p são previamente processados (ou não) pelas técnicas de remoção de *outlier* implementadas pelo *framework*: Zscore, MZScore, GESD e Chauvenet. Segundo, os dados ψ_c/ψ_p são fusionados por uma técnica de fusão redundante: média, mínimo, mediana ou máximo. Por fim, o manejo de irrigação calcula a umidade do solo $\theta_c(\psi_c)$ ou $\theta_p(\psi_p)$ para usá-lo no cálculo da ITN. A decisão de irrigar é definido pelo manejo avaliado.

5.3 Resultados da Avaliação do manejo do Potencial Mátrico

O manejo de irrigação do Potencial Mátrico (Eq. A.12) usa exclusivamente dados de umidade do solo (reais θ_c ou estimados θ_p pelos modelos de predição propostos) para calcular diariamente a irrigação total necessária (ITN). Nesse manejo, a irrigação ocorre somente quando a umidade do solo atual θ_a (θ_c ou θ_p) na profundidade mais superficial monitorada (e.g., $z = 30$ cm) é menor ou igual a umidade da condição crítica $\theta_{cr}(\psi_{cr})$, onde, $\psi_{cr} = 60$ kPa.

O manejo do Potencial Mátrico considera θ_{cr} a umidade ideal para irrigação, pois, ela sinaliza quando a produtividade do cultivo começa a declinar. Nesta avaliação, nos dias que o manejo do Potencial Mátrico permite a irrigação, a $ITN = 90,7$ mm calculada pela água facilmente disponível (AFD), definida pela Equação A.3, onde, a umidade da capacidade de campo θ_{cc} é definida com potencial mátrico $\psi_{cc} = 20$ kPa e a eficiência do sistema de irrigação

$Ef = 0,8185$ (Apêndice A.1).

5.3.1 Avaliação do manejo do Potencial Mátrico que usa dados reais θ_c de umidade do solo

5.3.1.1 Impacto da Fusão Redundante no manejo do Potencial Mátrico

Para a análise do impacto das técnicas de fusão redundante no desempenho do manejo do Potencial Mátrico que usa dados reais $\theta_c(\psi_c)$ de umidade do solo, foi calculada ITN diária (mm/dia) fusionando os dados de potencial mátrico ψ_c diariamente no período definido na Seção 5.2. A Tabela 18 relaciona as técnicas de fusão redundante à ITN estimada pelo manejo do Potencial Mátrico. Nesta Seção, deve-se analisar o desempenho do manejo, apenas com o processamento de dados da técnica de fusão redundante e **nenhuma** técnica de remoção de *outlier*.

Tabela 18 – ITN fornecida pelo manejo do Potencial Mátrico usando dados reais $\theta_c(\psi_c)$ de umidade do solo processados pelo Serviço de Fusão.

Técnica de Fusão Redundante	Técnica de Remoção de <i>Outlier</i>	Média (mm/dia)	Intervalo de Confiança (90%)	Economia Média (%)
Média	Nenhuma	69,0	(60,0; 78,1)	N/A
	Zscore	54,7	(44,3; 65,2)	20,7
	MZscore	59,2	(49,1; 69,4)	14,1
	GESD	66,0	(56,5; 75,5)	4,3
	Chauvenet	65,2	(55,6; 74,9)	5,4
Mínimo	Nenhuma	21,7	(12,6; 38,9)	N/A
	Zscore	21,7	(12,6; 38,9)	0
	MZscore	21,7	(12,6; 38,9)	0
	GESD	21,7	(12,6; 38,9)	0
	Chauvenet	43,5	(32,8; 54,2)	-100,4
Mediana	Nenhuma	64,5	(54,8; 74,2)	N/A
	Zscore	51,7	(41,2; 62,3)	19,8
	MZscore	57,7	(47,5; 68,0)	10,5
	GESD	63,0	(53,2; 77,8)	2,3
	Chauvenet	63,0	(53,2; 77,8)	2,3
Máximo	Nenhuma	86,2	(81,6; 90,9)	N/A
	Zscore	71,2	(62,5; 80,0)	17,4
	MZscore	68,2	(59,0; 77,5)	20,9
	GESD	81,0	(74,4; 87,6)	6,1
	Chauvenet	85,5	(80,5; 90,5)	0,9

O manejo do Potencial Mátrico que usa a Fusão Mínimo fornece a menor ITN média (21.7 mm/dia). Pois, o Mínimo seleciona o menor ψ_c (Figura 10) dos nove pontos de monitoramento do Campo Experimental e, conseqüentemente, a maior θ_c (Figura 11). E, assim,

a irrigação foi inibida durante os primeiros 55 dias do período analisado, já que, a Fusão Mínimo manteve a $\theta_c > \theta_{cr}$.

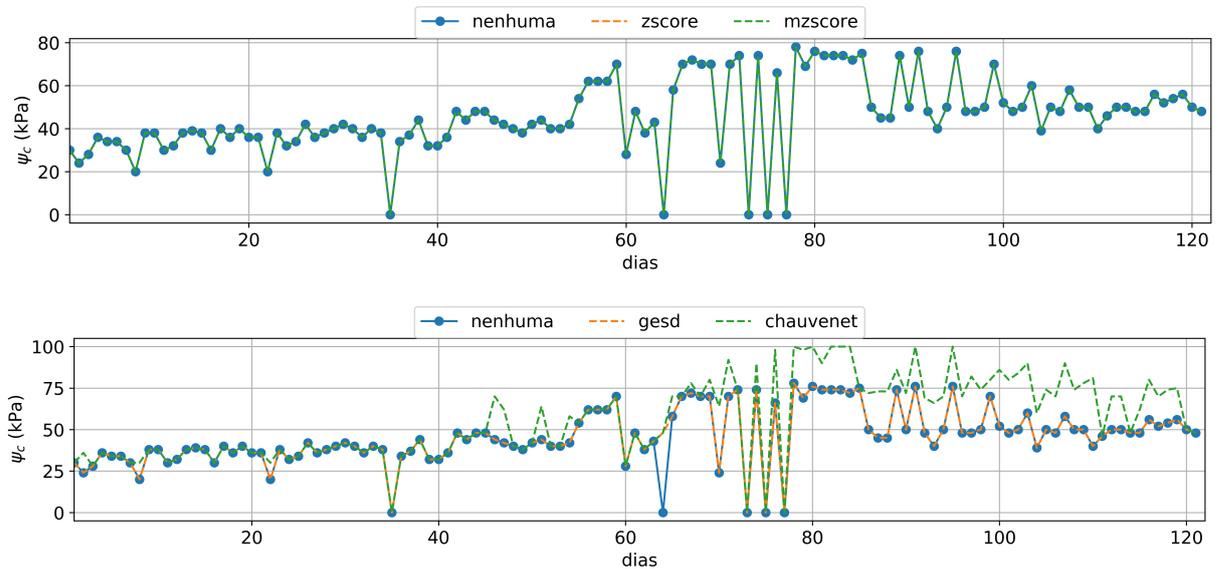


Figura 10 – Dados reais de potencial mátrico ψ_c resultantes da **Fusão Redundante Mínimo** e usados no manejo do Potencial Mátrico

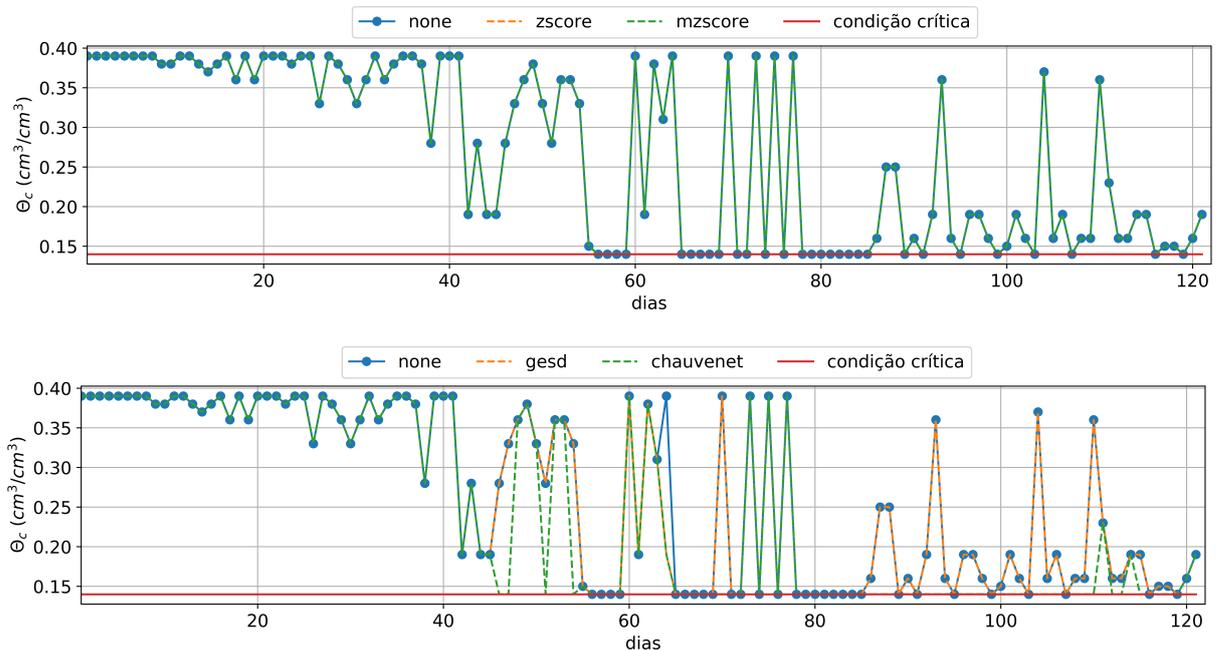


Figura 11 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da **Fusão Redundante Mínimo** — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

No outro extremo, a Fusão Máximo induz o manejo do Potencial Mátrico a fornecer a maior ITN média (86,2 mm/dia), visto que, o Máximo seleciona o maior valor de ψ_c daqueles providos pelos pontos de monitoramento (Figura 12), logo, a menor θ_c diária. A Figura 13 mostra a umidade atinge a condição crítica ($\theta_c \leq \theta_{cr}$) que por longos períodos. Por tanto, a

Fusão Máximo faz com que a ITN calculada pelo manejo do Potencial Mátrico seja em função do ponto de monitoramento com maior estresse hídrico (mais seco), porém, faz que o manejo gaste mais água diariamente.

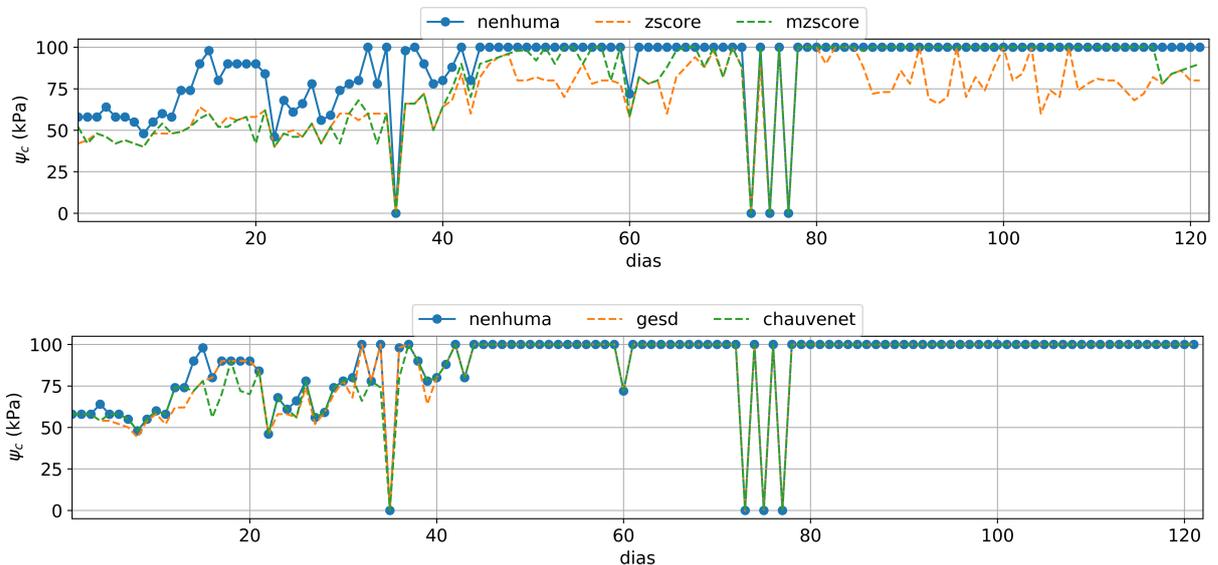


Figura 12 – Dados reais de potencial mátrico ψ_c resultantes da **Fusão Redundante Máximo** e usados no manejo do Potencial Mátrico

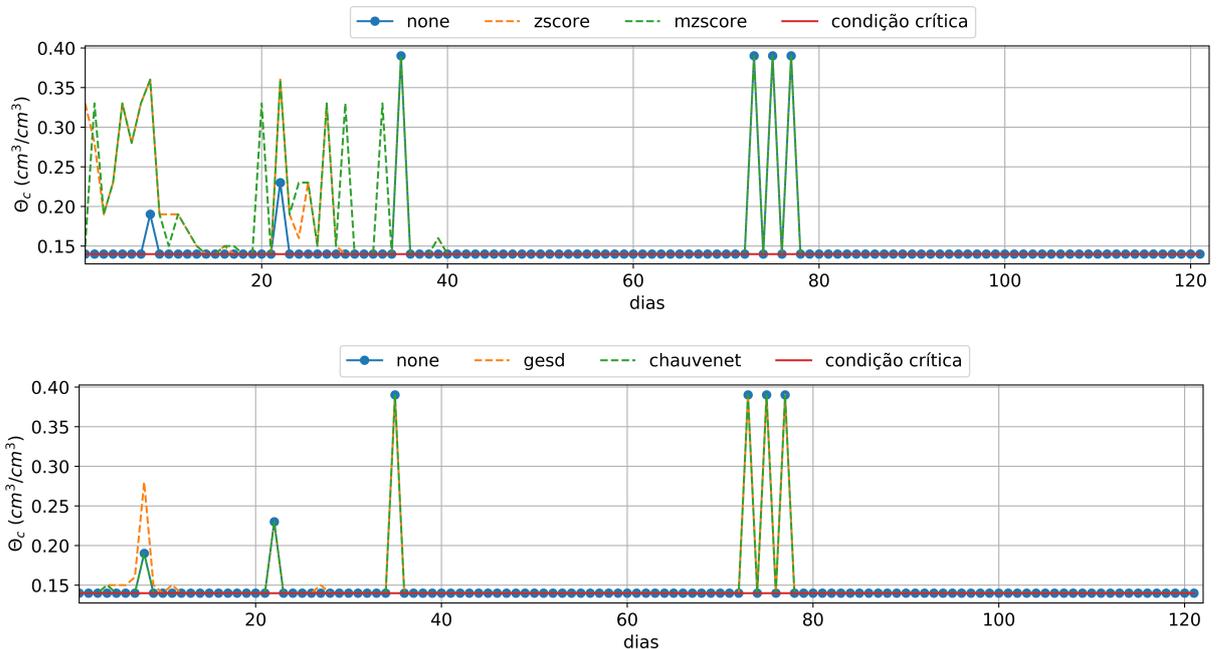


Figura 13 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da **Fusão Redundante Máximo** — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

O manejo de irrigação do Potencial Mátrico quando usa a Fusão Mediana provê uma ITN média de 64,5 mm/dia (Tabela 18). Isso significa que se o manejo do Potencial Mátrico usa a Fusão Mediana, obtém-se uma redução de 25,2% da ITN induzida pela Fusão Máximo e

um aumento de 66,4% da ITN induzida pelo Fusão Mínimo. Podemos observar os valores do potencial mátrico diário ψ_c fusionados pela Mediana na Figura 14 e os de umidade do solo na Figura 15.

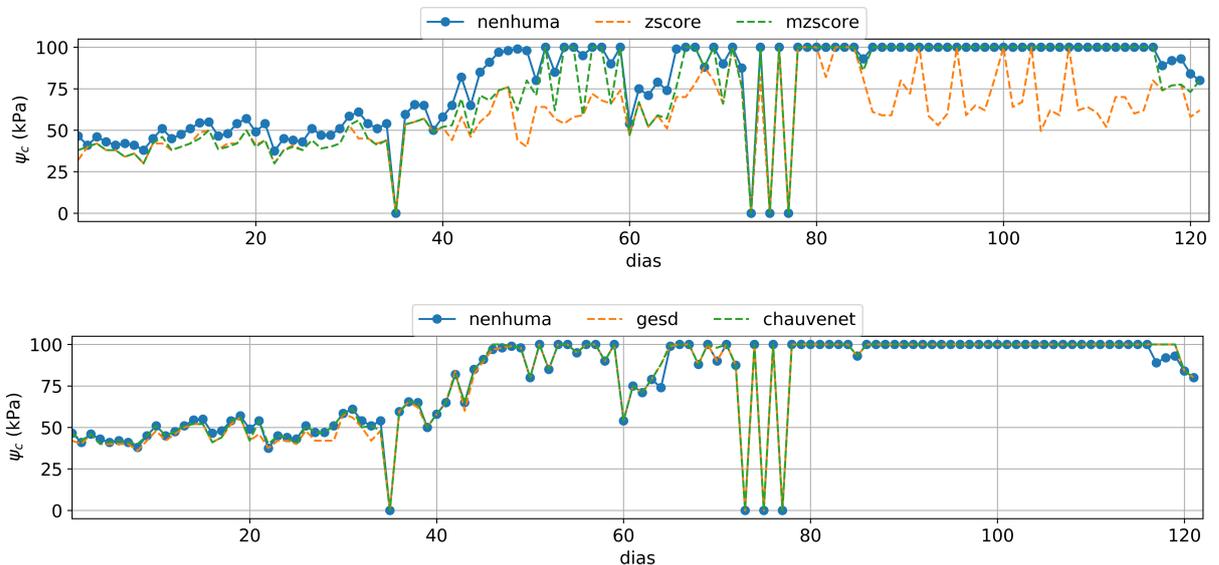


Figura 14 – Dados reais de potencial mátrico ψ_c resultantes da **Fusão Redundante Mediana** e usados no manejo do Potencial Mátrico

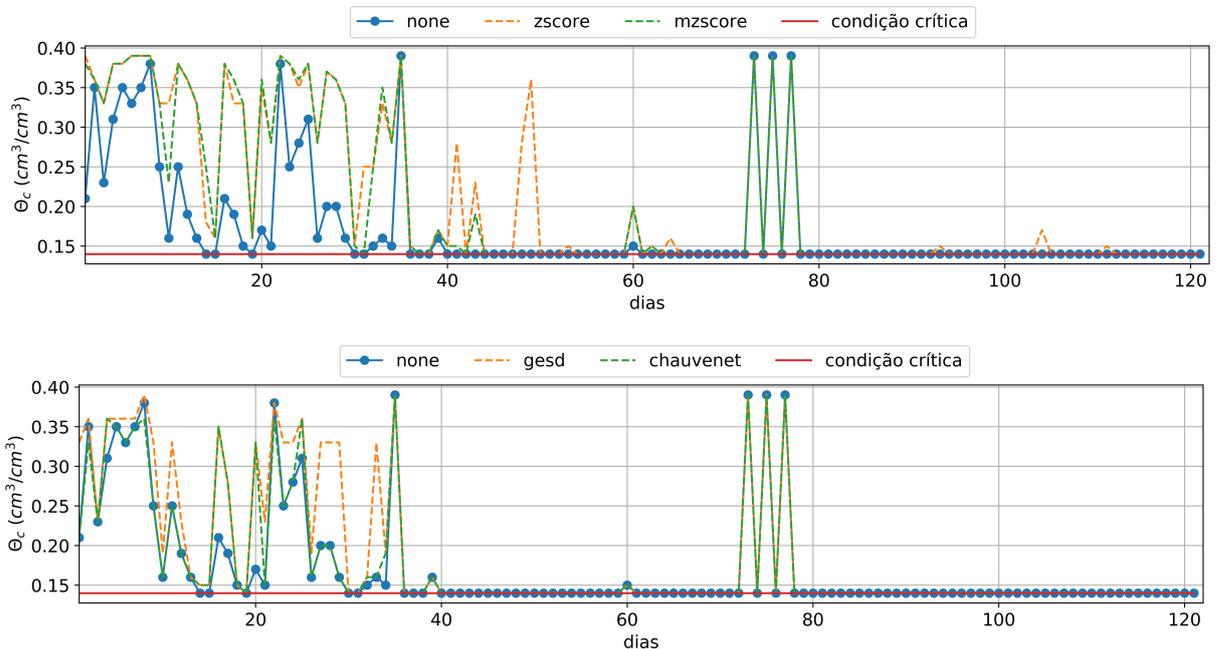


Figura 15 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da **Fusão Redundante Mediana** — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

A Fusão Média permite que o manejo do Potencial Mátrico planeje uma ITN média de 69,0 mm/dia (Tabela 18), o que corresponde a uma redução de 20% da ITN influenciada pela Fusão Máximo e um aumento de 68,6% da ITN quando o manejo usa a Fusão Mínima

para agregar dados ψ_c . As Figuras 16 e 17 apresentam os valores de ψ_c e θ_c da Fusão Média, respectivamente.

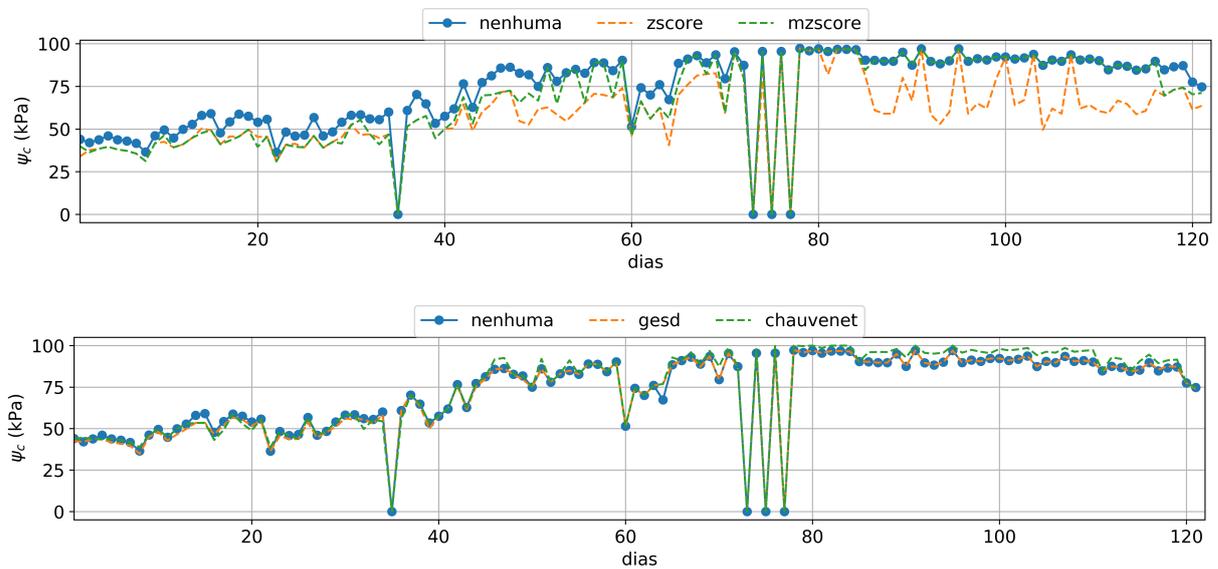


Figura 16 – Dados reais de potencial mátrico ψ_c resultantes da **Fusão Redundante Média** e usados no manejo do Potencial Mátrico

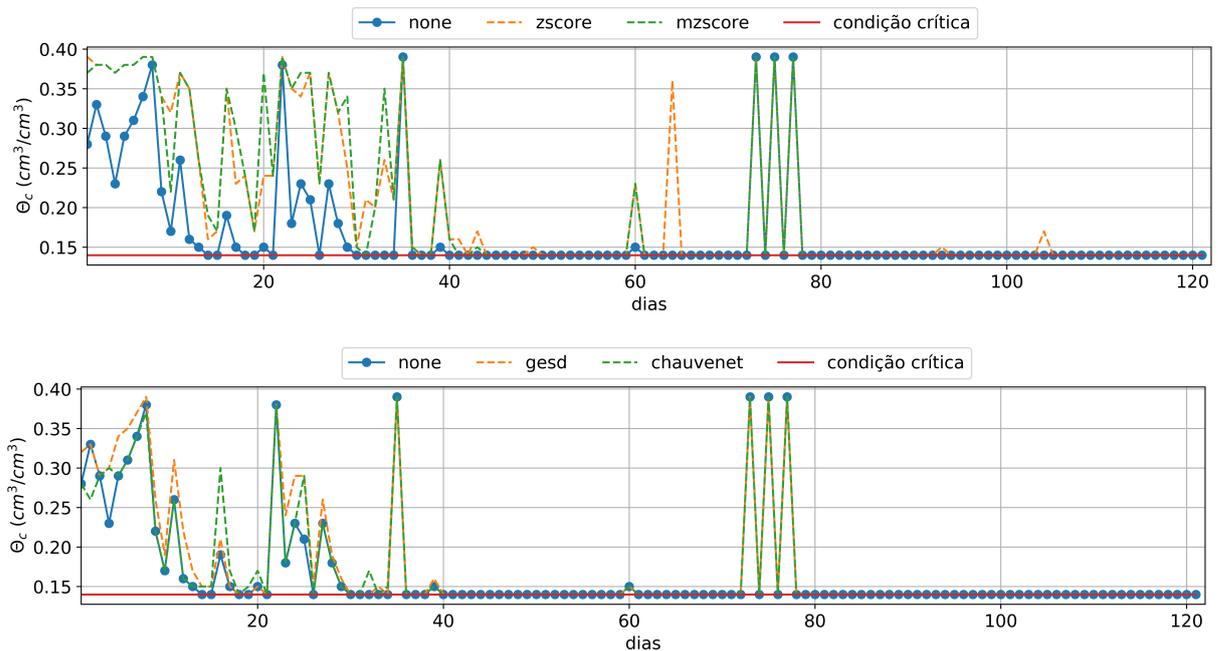


Figura 17 – Dados reais de umidade do solo $\theta_c(\psi_c)$ deduzidos usando dados ψ_c resultantes da **Fusão Redundante Média** — a irrigação ocorre quando $\theta_c \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

5.3.1.2 Impacto da Remoção de Outliers no manejo do Potencial Mátrico

A Tabela 18 também apresenta a percentagem de economia de ITN obtida pelo manejo quando usa cada técnica de remoção de *outlier* implementado pelo *framework* Smart&Green. Em geral, as técnicas Zscore e MZscore proporcionam maiores reduções: Zscore economiza em média de 17,4% (Fusão Máximo) a 20,7% (Fusão Média); e MZscore, de 10,5% (Fusão Mediana) a 20,9% (Fusão Máximo). Isto porque as técnicas Zscore e MZscore diminuem os valores potencial mátrico ψ_c com bem mostra a Figura 12 em seus primeiros quarenta dias. E, conseqüentemente, os valores de umidade do solo θ_c se tornam maiores que o da condição crítica θ_{cr} nesse período como mostra a Figura 13, fazendo que o manejo do Potencial Mátrico decida não irrigar.

As técnicas GESD e Chauvenet propiciam as menores economias de ITN ao manejo do Potencial Mátrico, porque, tanto GESD como Chauvenet aumenta os valores de ψ_c de maneira que fiquem maiores ou iguais a condição crítica ($\psi_{cr} = 60$ kPa) e, por consequência $\theta_c \leq \theta_{cr}$. No entanto, Chauvenet torna-se uma exceção ao ser usada em conjunto com a Fusão Mínimo, pois, faz com que o manejo do Potencial Mátrico duplique a ITN média para o período (Tabela 18). A Figura 10 mostra a eficácia de Chauvenet entre os dias 40 e 60, onde, após a fusão redundante, os valores de ψ_c atingem a condição crítica ($\psi_{cr} = 60$ kPa). Assim, com $\theta_c \leq \theta_{cr}$ (Figura 11), o manejo do Potencial Mátrico passa a permitir irrigação, colaborando para o aumento da ITN média.

5.3.2 Avaliação do manejo do Potencial Mátrico que usa dados estimados θ_p de umidade do solo

As abordagens de predição global e local são baseadas na técnica de aprendizagem de máquina do *Gradient Boosting Regression Tree*, pois, foi considerada a melhor dentre as outras técnicas avaliadas (Seção 4.2), já que obteve menores MSE (Mean Squared Error) e RMSE (Root Mean Squared Error). Esses modelos de predição usam dados climáticos, do cultivo e de irrigação (definidos no Apêndice B) para estimar o potencial mátrico ψ_p que permite a dedução de umidade do solo θ_p (Eq. A.1) para uma camada de solo de 0 a 30 cm de profundidade.

O modelo global estima apenas um valor diário de ψ_p e o modelo local, nove valores que são agregados por uma fusão redundante em um único valor ψ_p . No manejo do Potencial Mátrico, quando $\theta_p(\psi_p) \leq \theta_{cr}(\psi_{cr} = 60kPa)$, calcula-se irrigação total necessária (ITN) pela

água disponível do solo (Eq. A.3), onde, a umidade de capacidade de campo $\theta_{cc}(\psi_{cc} = 20 \text{ kPa})$ e a eficiência do sistema da irrigação $Ef = 0,8185$. Portanto, no dia que houver a necessidade de irrigação, $ITN = 90,7 \text{ mm}$.

O desempenho do manejo do Potencial Mátrico que usa $\theta_p(\psi_p)$ é medido em termos de ITN média diária ao longo do mesmo período (Seção 5.2), no qual o mesmo manejo foi avaliado usando dados reais $\theta_c(\psi_c)$ de umidade do solo. Foi avaliado também o impacto das técnicas de remoção de *outliers* na ITN média calculada pelo manejo do Potencial Mátrico quando usa $\theta_p(\psi_p)$. A Tabela 19 mostra a ITN diária obtida pelo manejo utilizando $\theta_p(\psi_p)$ estimados por a abordagem de predição, fusionados e tratados pelas técnicas de remoção de *outliers*.

Tabela 19 – ITN fornecida pelo manejo do Potencial Mátrico usando dados estimados $\theta_p(\psi_p)$ de umidade do solo e processados pelo Serviço de Fusão.

Predição	Técnica de Fusão Redundante	Técnica de Remoção de <i>Outlier</i>	Média (mm/dia)	Intervalo de Confiança (90%)
<i>Global</i>	<i>Nenhuma</i>	<i>Nenhuma</i>	42,1	(31,4; 52,8)
		<i>Nenhuma</i>	53,4	(42,9; 64,0)
<i>Local</i>	<i>Média</i>	Zscore	3,8	(-0,5; 8,0)
		MZscore	1,5	(-1,22; 4,23)
		GESD	53,4	(42,9; 64,0)
		Chauvenet	38,4	(27,8; 49,0)
		<i>Nenhuma</i>	0	(0,0)
<i>Local</i>	<i>Mínimo</i>	Zscore	0	(0,0)
		MZscore	0	(0,0)
		GESD	0	(0,0)
		Chauvenet	0	(0,0)
		<i>Nenhuma</i>	27,8	(18,0; 37,7)
<i>Local</i>	<i>Mediana</i>	Zscore	15,8	(7,7; 23,9)
		MZscore	11,3	(4,2; 18,4)
		GESD	27,8	(18,0; 37,7)
		Chauvenet	25,6	(15,9; 23,9)
		<i>Nenhuma</i>	91,0	(91,0; 91,0)
<i>Local</i>	<i>Máximo</i>	Zscore	58,7	(48,4; 69,0)
		MZscore	43,6	(32,9; 54,4)
		GESD	91,0	(91,0; 91,0)
		Chauvenet	63,2	(53,3; 73,1)

As técnicas Zscore e MZscore continuam a diminuir a ITN média, pois, tendem a reduzir os valores de ψ_p antes da Fusão Média (Figura 18), Mediana (Figura 20) e Máximo (Figura 22). Em consequência, os valores de θ_p ficam acima da condição crítica (Figura 19, 21 e 23), fazendo que o manejo do Potencial Mátrico não forneça ITN por vários dias. A técnica

Chauvenet diminuiu também ITN diária, pois, atua de maneira semelhante às técnicas Zscore e MZscore, principalmente, em conjunto com a Fusão Máximo (Figura 22 e 23).

A técnica GESD manteve a ITN média, pois, não modificou os valores de ψ_p e consequentemente, θ_p . A Fusão Mínimo selecionou o menor valor de ψ_p estimado pela abordagem de predição local (Figura 24), por isso, θ_p manteve-se sempre acima da condição crítica (Figura 25). Os dados diários ψ_p e θ_p estimados pela abordagem global podem ser visualizados nas Figuras 26 e 27.

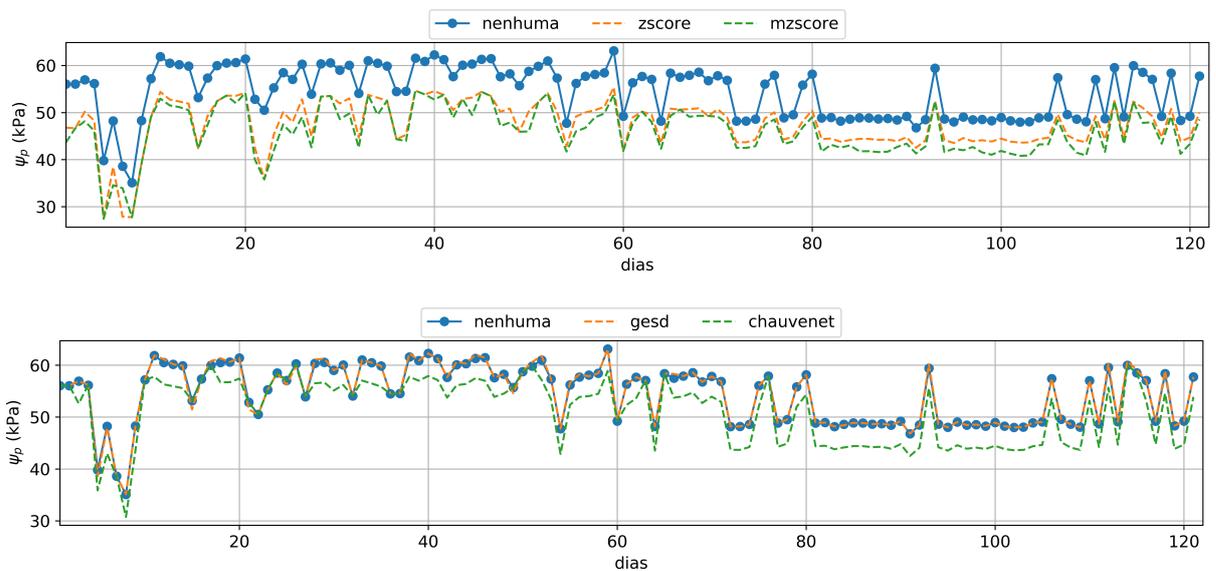


Figura 18 – Dados estimados de potencial mátrico ψ_p da abordagem de predição **Local** resultantes da **Fusão Redundante Média** e usados no manejo do Potencial Mátrico

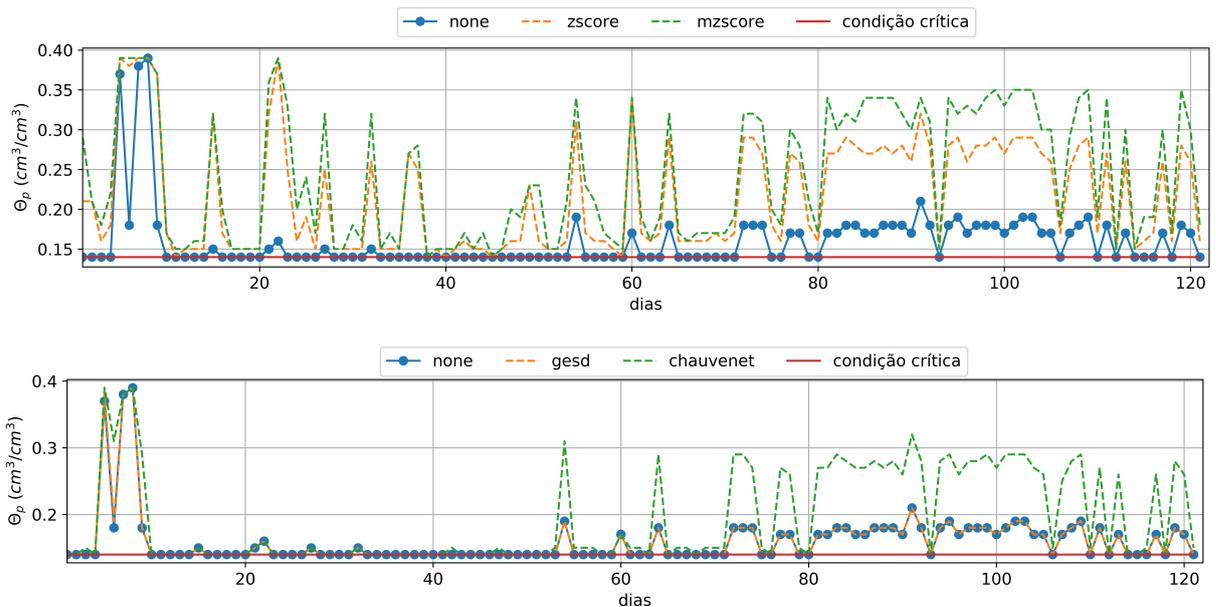


Figura 19 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição **Local** e fusionados pela **Média** — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

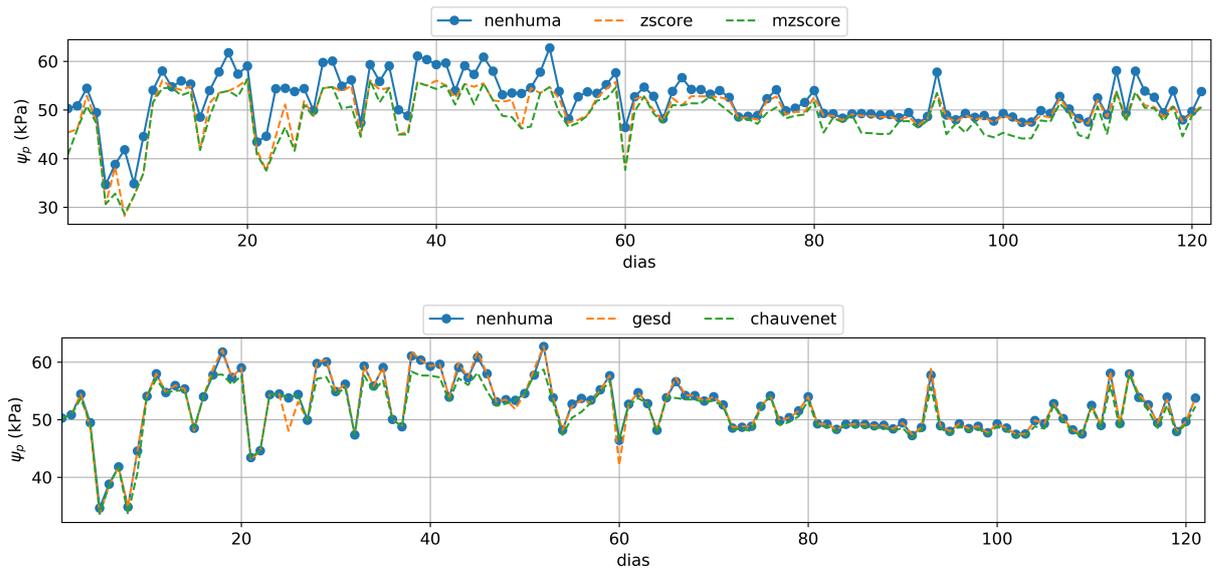


Figura 20 – Dados estimados de potencial mátrico ψ_p da abordagem de predição **Local** resultantes da **Fusão Redundante Mediana** e usados no manejo do Potencial Mátrico

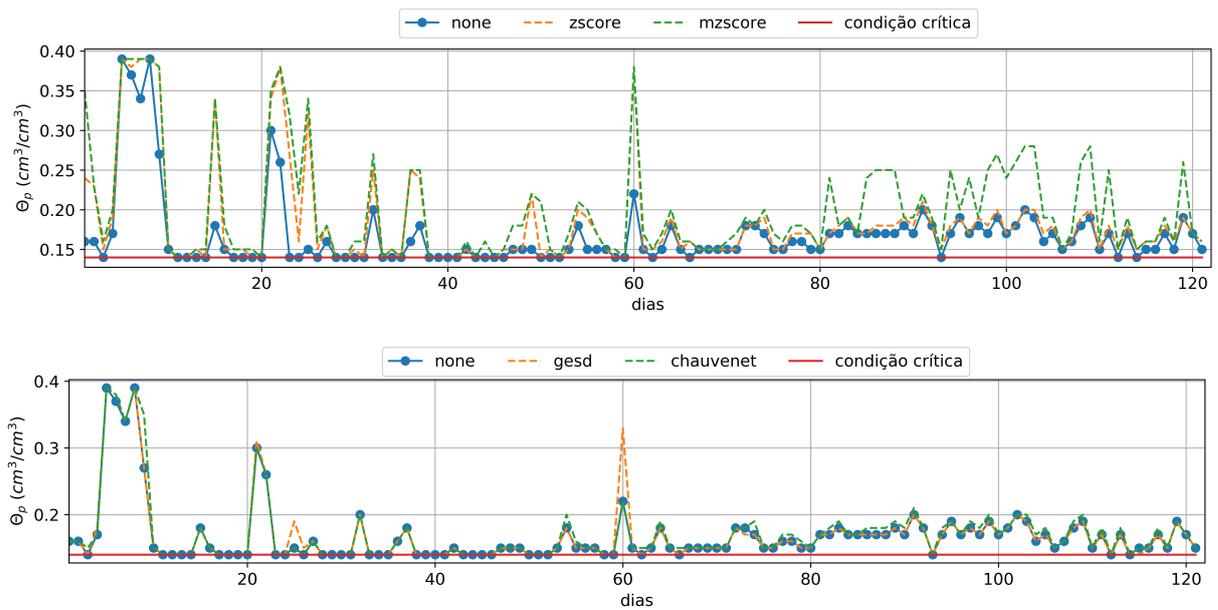


Figura 21 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição **Local** e fusionados pela **Mediana** — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

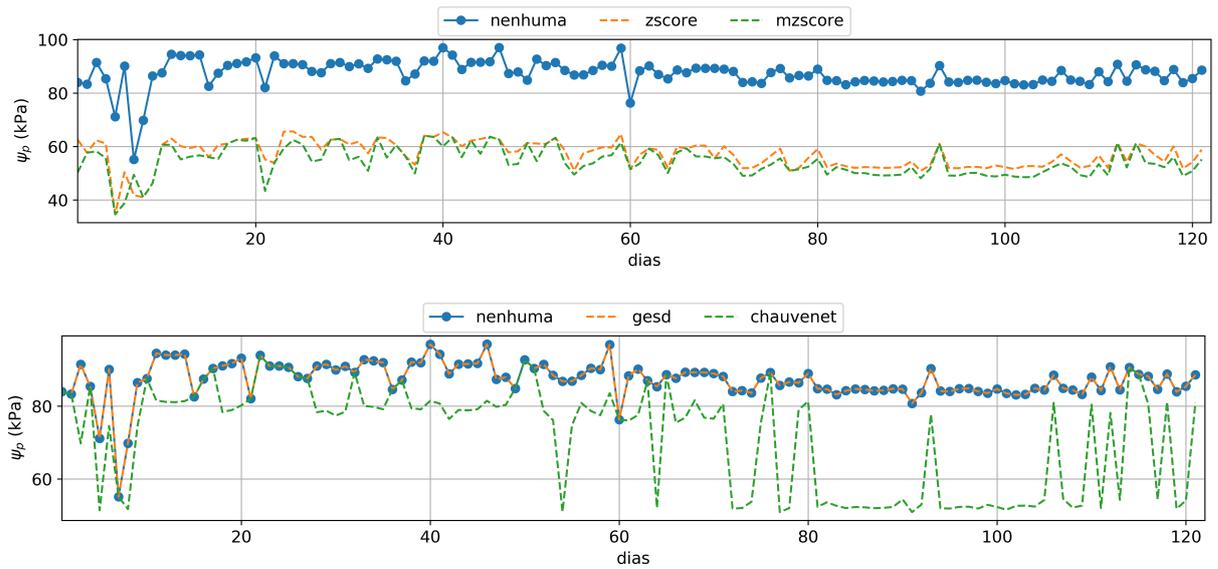


Figura 22 – Dados estimados de potencial mátrico ψ_p da abordagem de predição **Local** resultantes da **Fusão Redundante Máximo** e usados no manejo do Potencial Mátrico

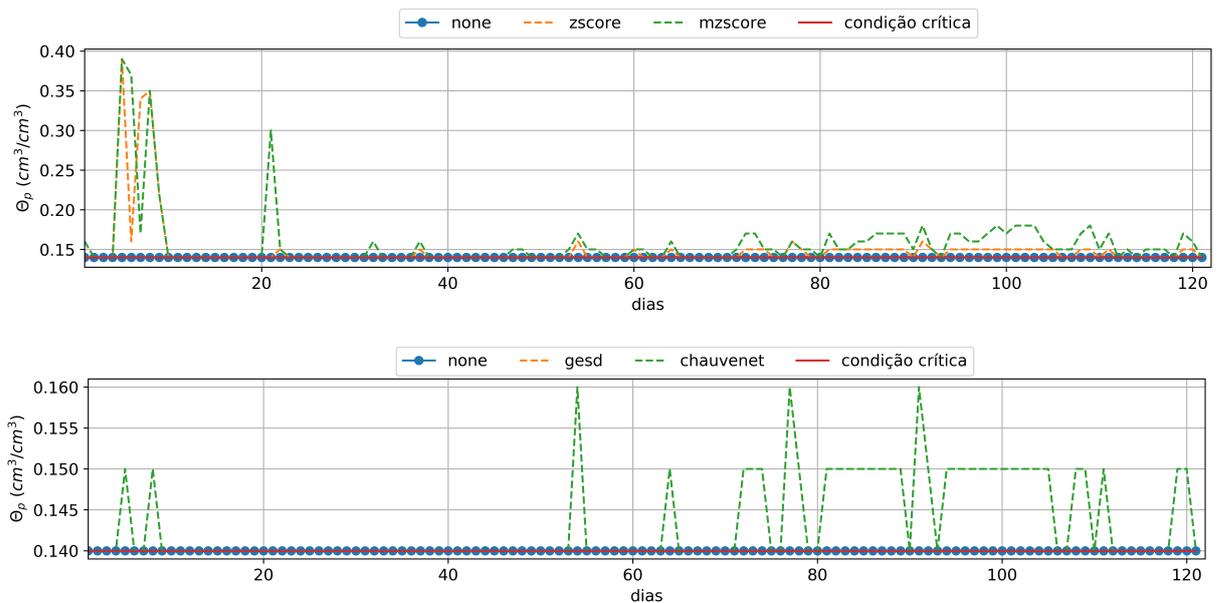


Figura 23 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição **Local** e fusionados pelo **Máximo** — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

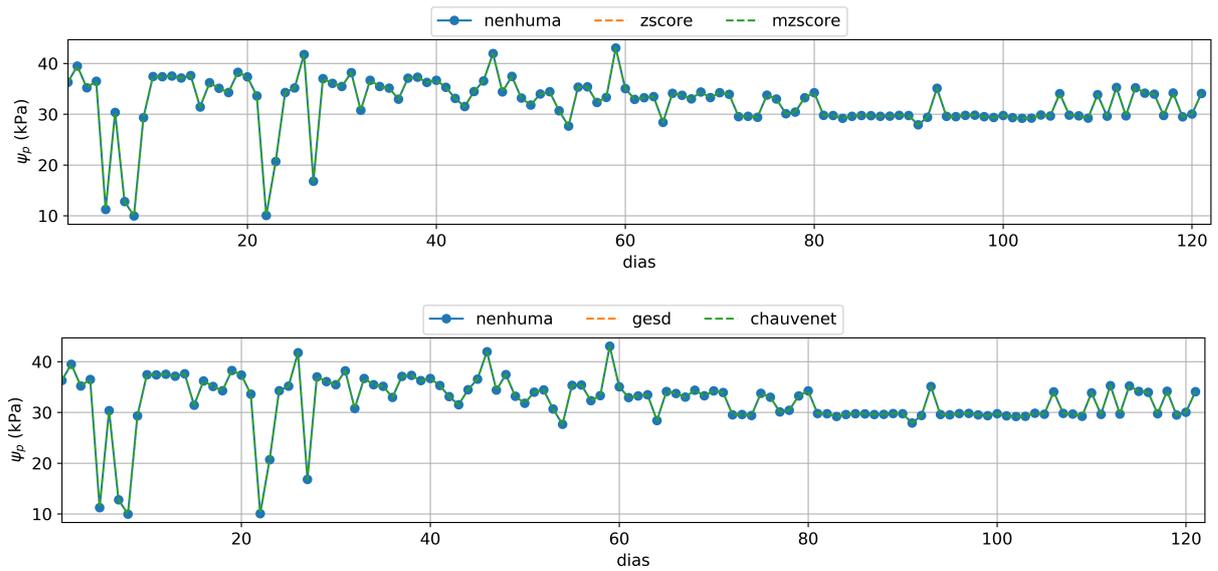


Figura 24 – Dados estimados de potencial mátrico ψ_p da abordagem de predição **Local** resultantes da **Fusão Redundante Mínimo** e usados no manejo do Potencial Mátrico

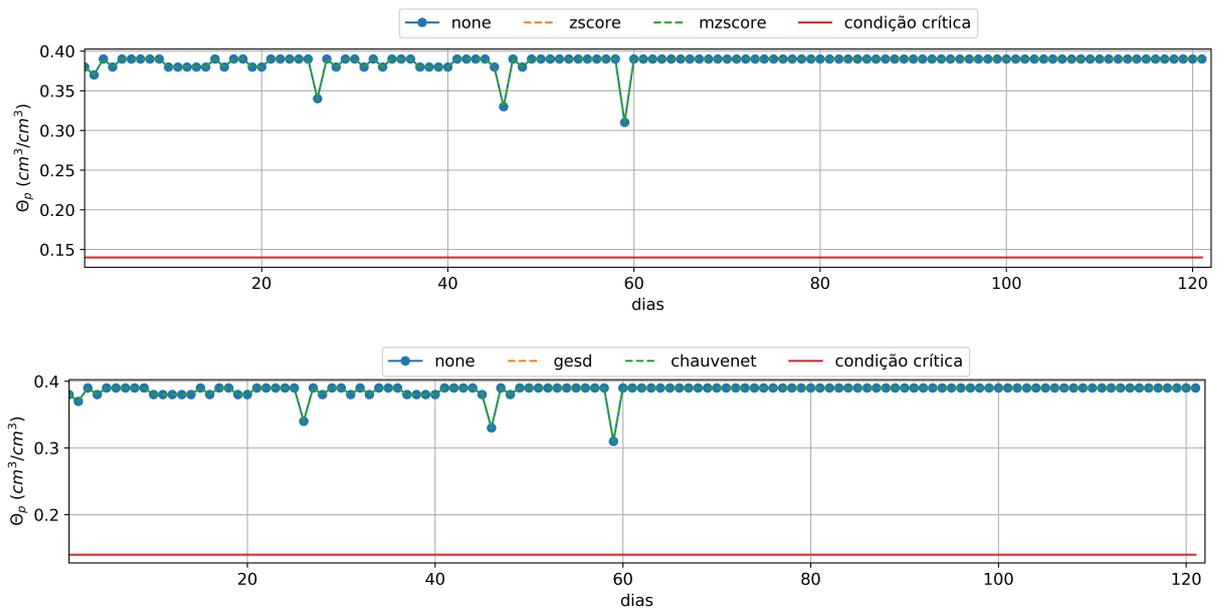


Figura 25 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição **Local** e fusionados pelo **Mínimo** — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

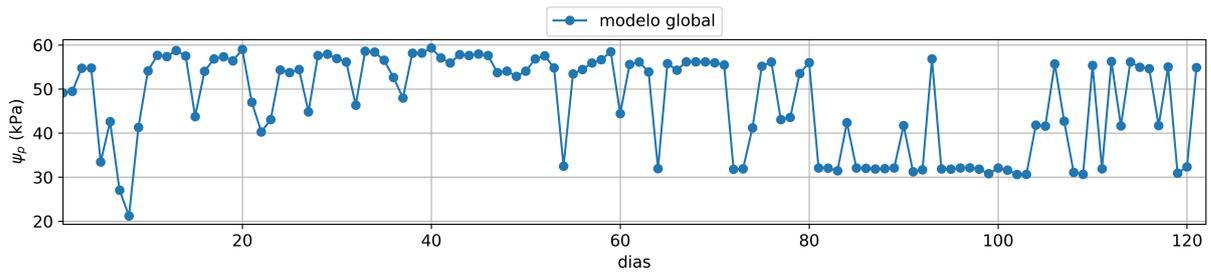


Figura 26 – Dados estimados de potencial mátrico ψ_p da abordagem de predição **Global** usados no manejo do Potencial Mátrico

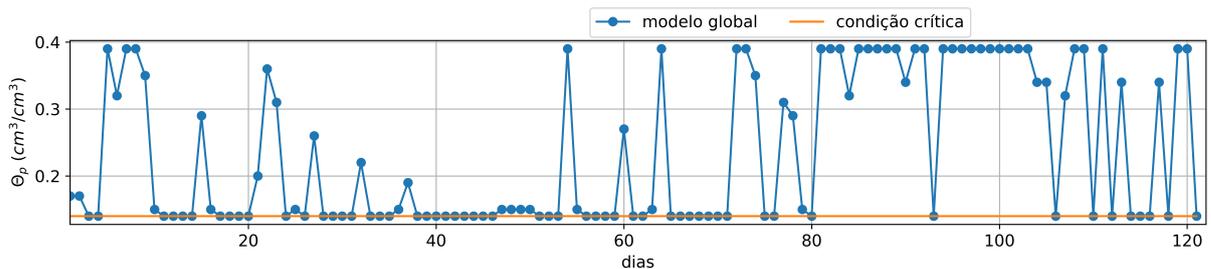


Figura 27 – Dados estimados de umidade do solo $\theta_p(\psi_p)$ da abordagem de predição **Global** — a irrigação ocorre quando $\theta_p \leq \theta_{cr}$ (umidade de condição crítica) no manejo de irrigação do Potencial Mátrico

Tabela 20 – Equivalência de desempenho entre o manejo do Potencial Mátrico que usa dados reais $\theta_c(\psi_c)$ e o mesmo que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição — Métrica: ITN média diferente em até ± 2 mm/dia.

Desemp. com Dados Reais $\theta_c(\psi_c)$			Desemp. com Dados Estimados $\theta_p(\psi_p)$		
Fusão + Rem. <i>Outlier</i>	Média (mm/dia)	Int. Conf. (90%)	Predição + Fusão + Rem. <i>Outlier</i>	Dif. Média	Dif. Int. Conf.
Mínimo + Chauvenet	43,5	(32,8; 54,2)	Global + Nenhuma + Nenhuma	-1,4	(-17,6; 14,9)
			Local + Máximo + Zscore	0,1	(-17,7; 18,0)
Mediana + Zscore	51,7	(41,2; 62,3)	Local + Média + Nenhuma	1,7	(-14,6; 18,0)
Mediana + MZscore	57,1	(47,5; 68,0)	Local + Máximo + Zscore	0,9	(-15,2; 17,1)

A Tabela 20 associa o manejo do Potencial Mátrico que usa $\theta_c(\psi_c)$ àquele que usa $\theta_p(\psi_p)$, pois, possuem desempenho equivalente, isto é, possuem uma ITN média diferente em até ± 2 mm/dia. A economia da ITN média dos manejos de desempenhos equivalentes estão na Tabela 21. Essa economia foi obtida comparando a ITN média do manejo do Potencial Mátrico que usa dados reais $\theta_c(\psi_c)$, a Fusão Mínimo e Mediana, e nenhuma técnica de remoção de *outliers* (Tabela 18). Por tanto, para um campo sem sensores de solo (Campo 1 da Figura 1), é

possível usar o manejo do Potencial Mátrico e os dados de umidade do solo estimados $\theta_p(\psi_p)$ pelas abordagens de predição global e local auxiliados pela Fusão Máximo e Média e as técnicas de remoção de *outliers* Zscore e MZscore.

Tabela 21 – Economia de ITN do manejo do Potencial Mátrico que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição de umidade do solo com desempenho equivalente ao mesmo manejo que usa dados reais $\theta_c(\psi_c)$

Desemp. com Dados Reais $\theta_c(\psi_c)$		Desemp. com Dados Estimados $\theta_p(\psi_p)$	
Fusão + Rem. <i>Outlier</i>	Média (%)	Predição + Fusão + Rem. <i>Outlier</i>	Média (%)
Mínimo + Chauvenet	-100,4	Global + Nenhuma + Nenhuma	-93,7
		Local + Máximo + Zscore	-100,6
Mediana + Zscore	19,8	Local + Média + Nenhuma	17,2
Mediana + MZscore	10,5	Local + Máximo + Zscore	9,0

5.4 Resultados da Avaliação do manejo do Balanço Hídrico

O manejo de irrigação do Balanço Hídrico (Eq. A.7) pode calcular a irrigação total necessária (ITN) apenas com a evapotranspiração¹ do cultivo ET_c (Eq. A.4) e a precipitação pluvial P . Assim, pode-se usar apenas dados climáticos e o coeficiente de cultivo k_c (Tabela 6) para planejar o Balanço Hídrico para um intervalo de n dias. No entanto, o cálculo da ITN torna-se mais preciso quando se incorpora a umidade do solo atual θ_a , pois, o armazenamento de água na zona radicular do cultivo pode contribuir para o suplemento de água.

O objetivo desta seção é avaliar o impacto das diferentes técnicas de fusão redundante e remoção de *outlier* implementadas pelo Serviço de Fusão do *framework* Smart&Green (Seção 3.3.3) no manejo do Balanço Hídrico que utiliza dados de umidade do solo reais θ_c (Seção 5.4.1) e estimados θ_p (Seção 5.4.2) pelas abordagens de predição global e local (Seção 3.2.2) baseadas no *Gradient Boosting Regression Tree* — o melhor método preditivo avaliado (Seção 4.2).

Considerando $n = 3$ dias, foram selecionados 26 intervalos no período definido (Seção 5.2), no qual ***o manejo do Balanço Hídrico que não usa θ_a (θ_c ou θ_p) forneceu uma ITN média de 32,3 mm/3 dias com intervalo de confiança de 90% variando de 24,1 a 40,6 mm/3 dias.*** Essa informação foi usada para calcular a economia de ITN do Balanço Hídrico que usa dados de umidade do solo reais θ_c ou de um modelo computacional θ_p .

¹ A evapotranspiração do cultivo mede o consumo de água do cultivo, de acordo com estágio de desenvolvimento do cultivo e condições meteorológicas (ANA, 2017)

5.4.1 Avaliação do manejo do Balanço Hídrico que usa dados reais θ_c de umidade do solo

A Tabela 22 relaciona as técnicas do Serviço de Fusão à ITN média provida pelo manejo do Balanço Hídrico que usa dados de umidade reais $\theta_c(\psi_c)$, isto é, deduzidos usando dados de potencial mátrico ψ_c (Eq. A.1) coletados pelos sensores de solo do Campo Experimental (4.1.1). Primeiramente, os dados ψ_c são tratados por uma técnica de remoção de *outliers* e, depois, agregados em um único valor ψ_c por uma técnica de fusão redundante. Logo, para analisar o impacto das técnicas de fusão redundante no desempenho do Balanço Hídrico que usa θ_c , deve-se considerar **nenhuma** influência das técnicas de remoção de *outliers*.

Tabela 22 – ITN média obtida pelo Balanço Hídrico usando dados reais $\theta_c(\psi_c)$ de umidade do solo processados pelo Serviço de Fusão e economia de ITN em comparação ao Balanço Hídrico que não usa umidade do solo

Técnica de Fusão Redundante	Técnica de Remoção de <i>Outlier</i>	Média (mm/3 dias)	Intervalo de Confiança (90%)	Economia Média (%)
Média	Nenhuma	3,1	(-2,6; 8,9)	90,4
	Zscore	3,5	(-2,3; 9,2)	89,3
	MZscore	3,5	(-2,3; 9,2)	89,3
	GESD	3,1	(-2,6; 8,9)	90,4
	Chauvenet	3,1	(-2,6; 8,9)	90,4
Mínimo	Nenhuma	7,6	(1,3; 13,9)	76,5
	Zscore	7,6	(1,3; 13,9)	76,5
	MZscore	7,6	(1,3; 13,9)	76,5
	GESD	7,6	(1,3; 13,9)	76,5
	Chauvenet	7,6	(1,3; 13,9)	76,5
Mediana	Nenhuma	3,1	(-2,6; 8,9)	90,4
	Zscore	4,4	(-1,4; 10,3)	86,3
	MZscore	4,0	(-1,8; 9,9)	87,6
	GESD	3,3	(-2,4; 9,1)	89,8
	Chauvenet	3,3	(-2,4; 9,1)	89,8
Máximo	Nenhuma	3,1	(-2,6; 8,9)	90,4
	Zscore	3,1	(-2,6; 8,9)	90,4
	MZscore	3,1	(-2,6; 8,9)	90,4
	GESD	3,1	(-2,6; 8,9)	90,4
	Chauvenet	3,1	(-2,6; 8,9)	90,4

Assim, verificamos na Tabela 22 que a Fusão Mínimo reduz a ITN média em 76,5% e as outras técnicas, em 90,4%. As Figuras 28, 29, 30 e 31 mostram os valores de potencial mátrico ψ_c fusionados. As Figuras 32, 33, 34, e 35 apresentam os valores de umidade do solo $\theta_c(\psi_c)$ calculada usando o modelo de Van Genuchten (Eq. A.1). A Fusão Mínimo seleciona o menor valor de ψ_c fornecidos e faz com que o Balanço Hídrico forneça a maior ITN média.

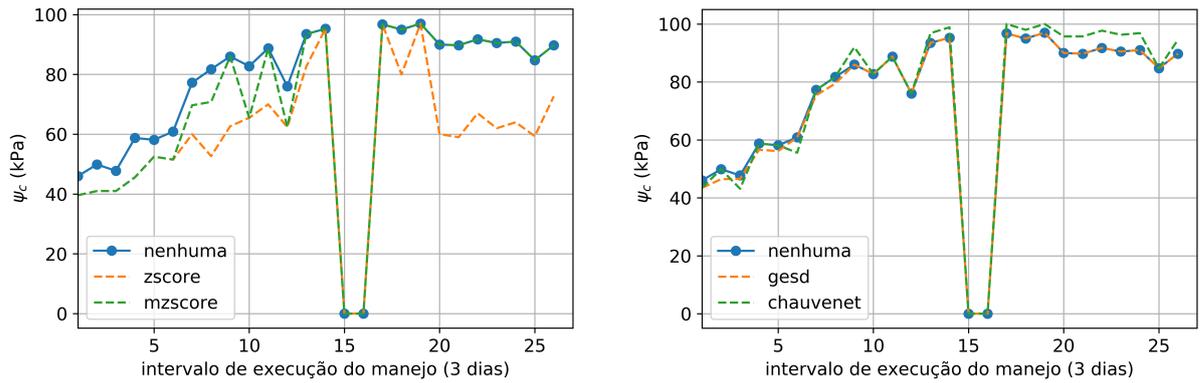


Figura 28 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da **Fusão Média**

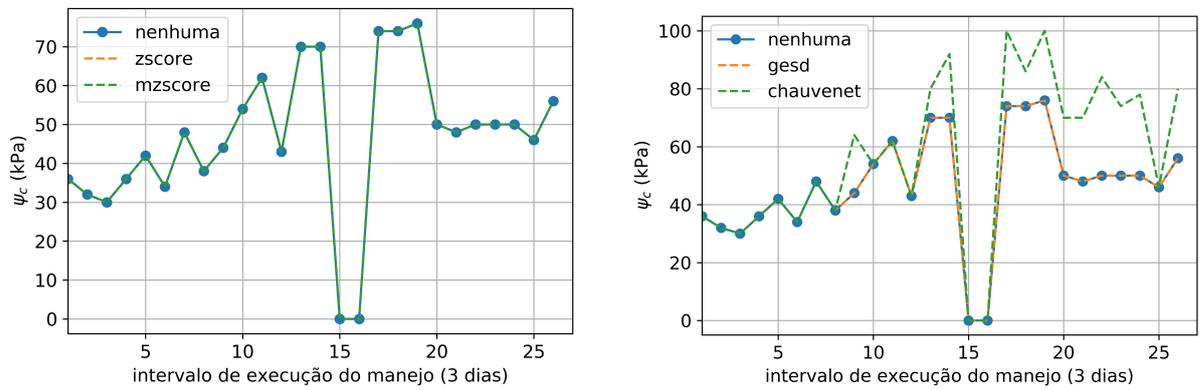


Figura 29 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da **Fusão Mínimo**

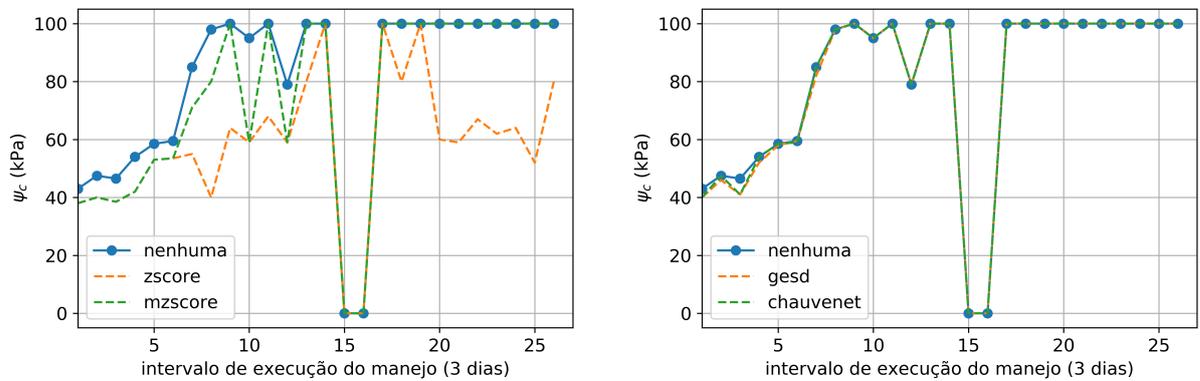


Figura 30 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da **Fusão Mediana**

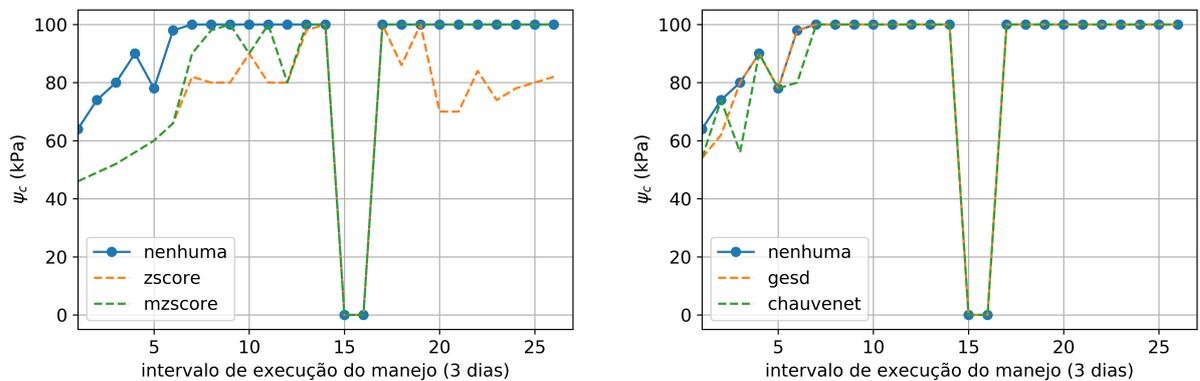


Figura 31 – Balanço Hídrico – Dados de potencial mátrico reais ψ_c da **Fusão Máximo**

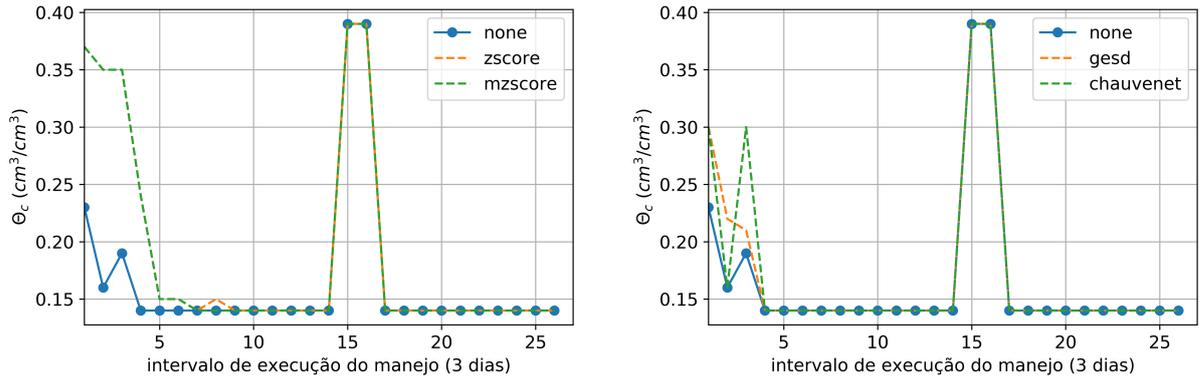


Figura 32 – Balanço Hídrico –Umidade do solo real $\theta_c(\psi_c)$ resultante da **Fusão Média**

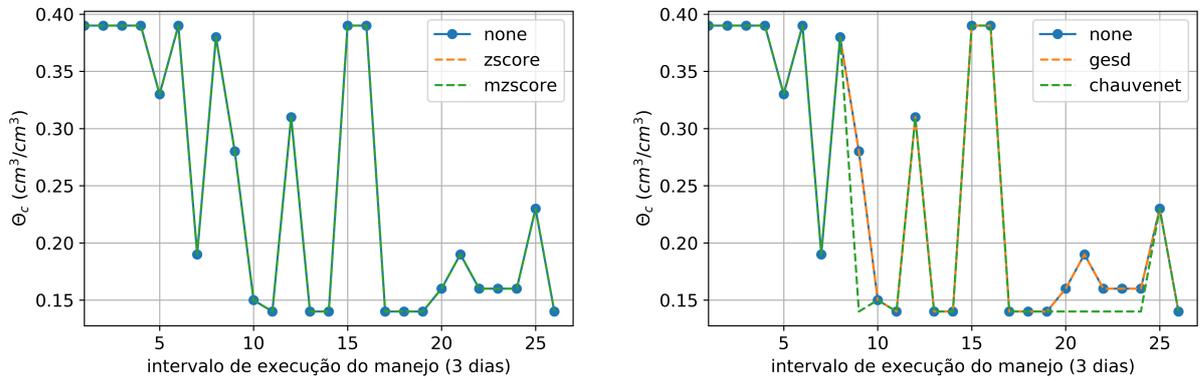


Figura 33 – Balanço Hídrico –Umidade do solo real $\theta_c(\psi_c)$ resultante da **Fusão Mínimo**

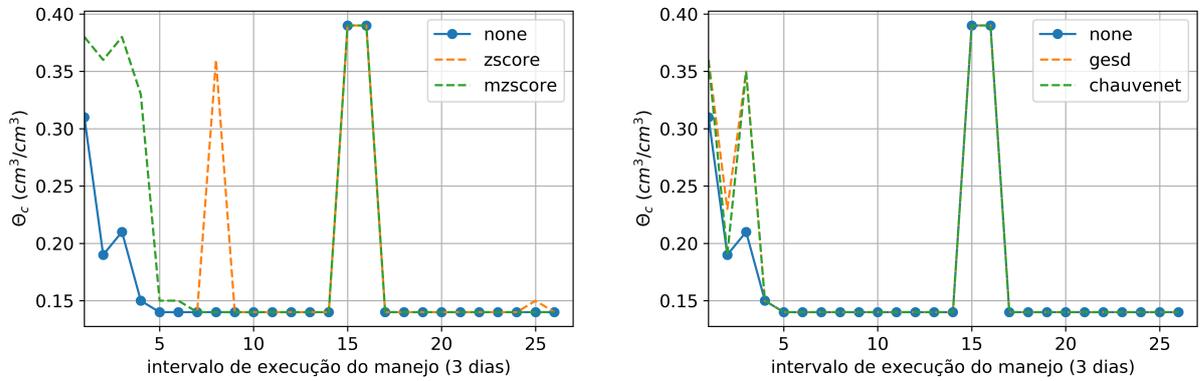


Figura 34 – Balanço Hídrico –Umidade do solo seal $\theta_c(\psi_c)$ resultante da **Fusão Mediana**

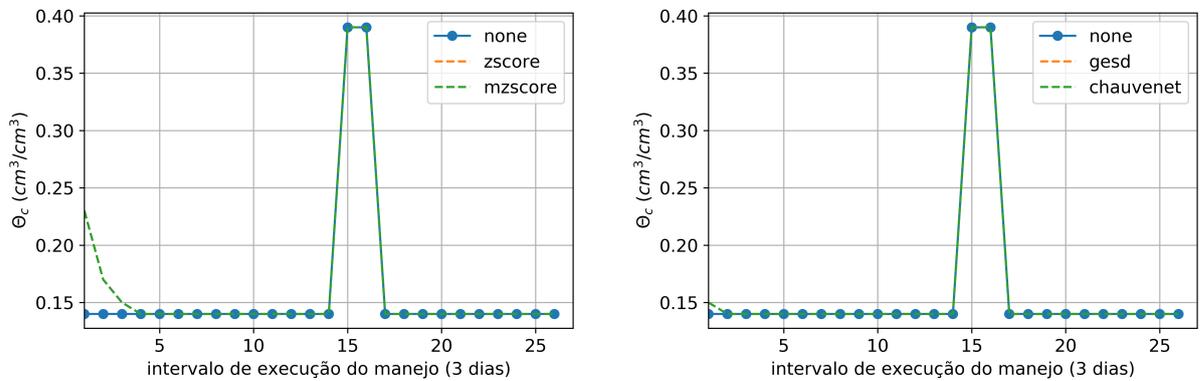


Figura 35 – Balanço Hídrico –Umidade do solo real $\theta_c(\psi_c)$ resultante da **Fusão Máximo**

As técnicas de remoção de *outlier* não tem efeito algum se os valores de ψ_c são fusionados com o Mínimo e Máximo no Balanço Hídrico que usa θ_c . As técnicas Zscore e MZscore diminuem os valores de ψ_c quando processam dados com a Fusão Média (Figura 28) e Mediana (Figura 30) e Máximo (Figura 31). Com isso, essas técnicas fazem com que o Balanço Hídrico aumente a ITN média, e conseqüentemente, diminuam a sua economia em 86,3% (Zscore) e 87,6% (MZscore), com a Fusão Mediana (Tabela 22). Quando os valores de ψ_c são fusionados com Média, Zscore e MZscore fazem com que Balanço Hídrico tenha o mesmo desempenho, economizando 89,3% da ITN média. GESD e Chauvenet também influenciam no aumento da ITN média do Balanço Hídrico apenas quando atuam em conjunto com a Fusão Mediana, contribuindo a economia de 89,8% da ITN.

Nesta avaliação, consideramos que poderá existir um *tradeoff* razoável entre economia de ITN e desenvolvimento saudável do cultivo quando o manejo do Balanço Hídrico usa apenas a Fusão Mínimo ou a Fusão Mediana com a remoção de *outliers* da Zscore e MZscore, visto que, essas técnicas não restringem tanto a ITN que o Balanço Hídrico tenta suprir a evapotranspiração do cultivo a cada três dias. Para comprovar esta afirmação, é necessário fornecer a ITN ao cultivo e fazer a biometria das plantas, isto é, medir a altura, a envergadura e número de folhas para verificar se ele se desenvolve de maneira saudável.

5.4.2 Avaliação do manejo do Balanço Hídrico que usa dados estimados θ_p de umidade do solo

A Tabela 23 relaciona o desempenho do Balanço Hídrico que usa dados de umidade do solo estimados $\theta_p(\psi_p)$ pelas abordagens de predição em termos de ITN média e sua economia quando comparamos ao Balanço Hídrico que não usa dados de umidade do solo. As técnicas de fusão redundante são usadas na abordagem de predição local, pois, são estimados nove valores de potencial mátrico ψ_p que precisam ser fusionados, para depois, calcular a umidade do solo θ_p por meio do modelo de Van Genuchten (Eq. A.1) e, enfim, ser usada pelo Balanço Hídrico (Eq. A.7). Para análise do impacto das técnicas de fusão redundante no desempenho do Balanço Hídrico que usa $\theta_p(\psi_p)$, deve-se verificar apenas o processamento de dados da fusão redundante com **nenhuma** técnica de remoção de *outlier*.

A Fusão Média, Mediana e Máximo provocaram a inibição da irrigação do Balanço Hídrico, com exceção da Fusão Mínima que obteve uma economia de 1,3% da ITN média (Tabela 23). Os valores de ψ_p fusionados por cada técnica são encontrados da Figura 36 até

Figura 39 e os valores de θ_p , da Figura 40 à Figura 43.

Tabela 23 – ITN média fornecida pelo Balanço Hídrico usando dados de umidade do solo estimados $\theta_p(\psi_p)$ pelos Modelos de Predição e processados pelo Serviço de Fusão. A economia de ITN é obtida comparando-se ao Balanço Hídrico que não usa umidade do solo

Predição	Técnica de Fusão	Técnica de Remoção de Outlier	Média (mm/3 dias)	Inter. de Conf. (90%)	Economia (%)
Global	Nenhuma	Nenhuma	14,1	(2,3; 25,6)	56,4
		Nenhuma	0	(0; 0)	100
		Zscore	3,3	(-0,8; 7,3)	90,0
Local	Média	MZscore	7,9	(0,4; 15,4)	75,6
		GESD	0	(0; 0)	100
		Chauvenet	3,3	(-0,8; 7,3)	90,0
		Nenhuma	32,0	(23,5; 40,6)	1,3
		Zscore	32,0	(23,5; 40,6)	1,3
Local	Mínimo	MZscore	32,0	(23,5; 40,6)	1,3
		GESD	32,0	(23,5; 40,6)	1,3
		Chauvenet	32,0	(23,5; 40,6)	1,3
		Nenhuma	0	(0; 0)	100
		Zscore	0	(0; 0)	100
Local	Mediana	MZscore	1,6	(-1,3; 4,6)	95,0
		GESD	0	(0; 0)	100
		Chauvenet	0	(0; 0)	100
		Nenhuma	0	(0; 0)	100
		Zscore	0	(0; 0)	100
Local	Máximo	MZscore	0	(0; 0)	100
		GESD	0	(0; 0)	100
		Chauvenet	0	(0; 0)	100

As técnicas de remoção de *outliers* não provocam a irrigação do Balanço Hídrico que usa a Fusão Máximo, nem alteração da ITN promovida pela Fusão Mínimo (Tabela 23). No entanto, a técnica MZscore em conjunto com a Fusão Mediana, faz com que o Balanço Hídrico aumente a ITN com uma economia de 95% em relação ao Balanço Hídrico que não usa dados de umidade do solo. Estas associações de fusão redundante e técnicas de remoção de outliers não são boas opções de tratamento de dados, pois, fornecem economias de ITN maiores que 90,4% de quando o Balanço Hídrico usa dados reais de umidade do solo (Tabela 22).

O processamento de dados das técnicas Zscore e Chauvenet com a Fusão Média faz com que o Balanço Hídrico calcule uma ITN média com uma economia de 90% da ITN diária, enquanto a técnica MZscore, uma economia de 75,6%. Isto porque, em geral, estas técnicas diminuem os valores de potencial mátrico ψ_p (Figuras 36, 38 e 39) de tal maneira que os valores de umidade do solo estimados θ_p (Figs. 40 e 42) colaboram para o aumento da ITN do Balanço Hídrico.

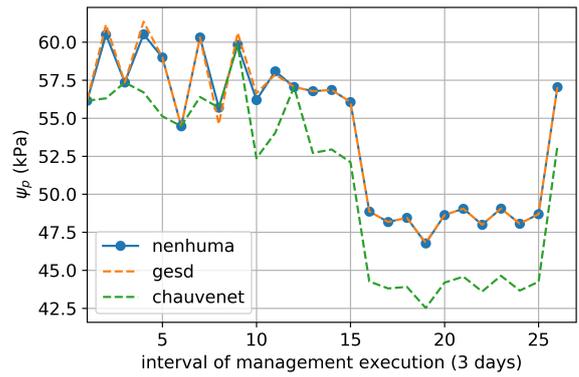
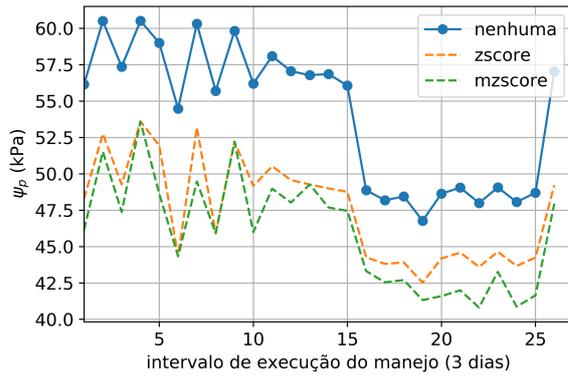


Figura 36 – Balanço Hídrico: ψ_p estimado pela abordagem **Local** e fusionado pela **Média**

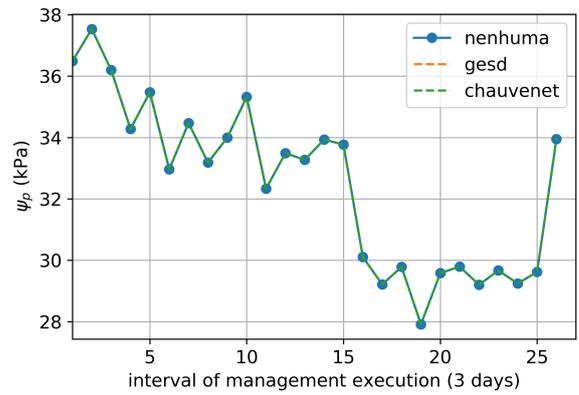
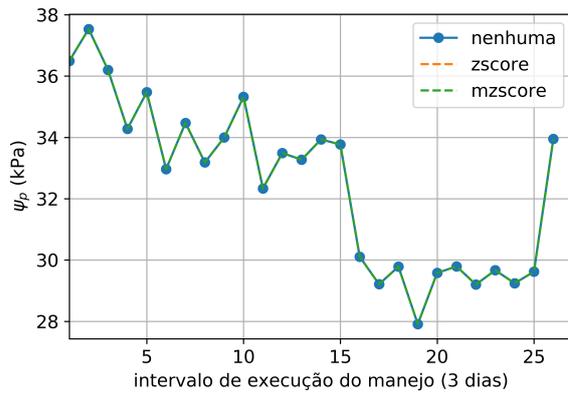


Figura 37 – Balanço Hídrico: ψ_p estimado pela abordagem **Local** e fusionado pelo **Mínimo**

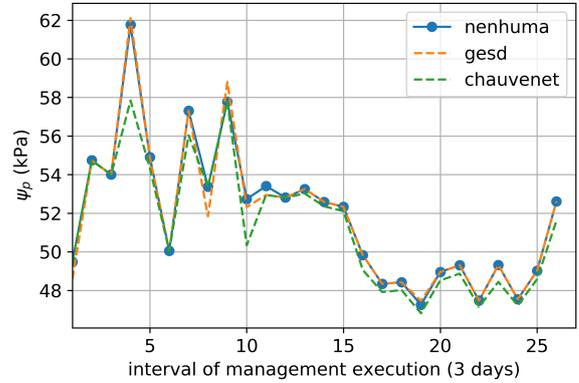
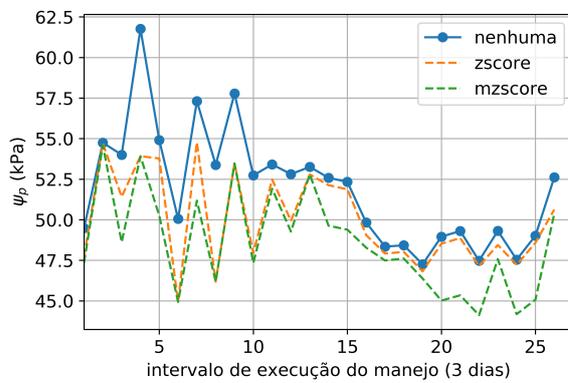


Figura 38 – Balanço Hídrico: ψ_p estimado pela abordagem **Local** e fusionado pela **Mediana**

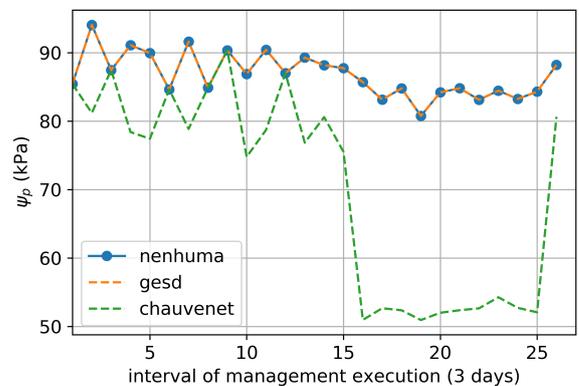
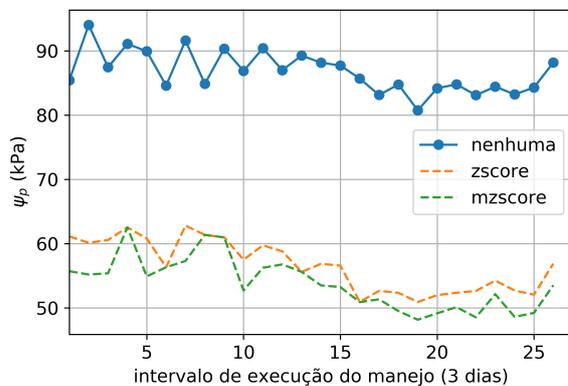


Figura 39 – Balanço Hídrico: ψ_p estimado pela abordagem **Local** e fusionado pelo **Máximo**

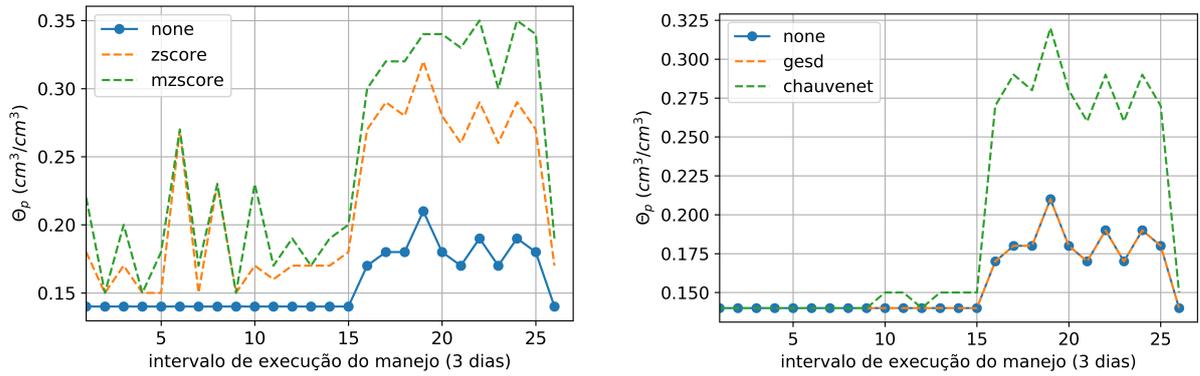


Figura 40 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem **Local** e fusão **Média**

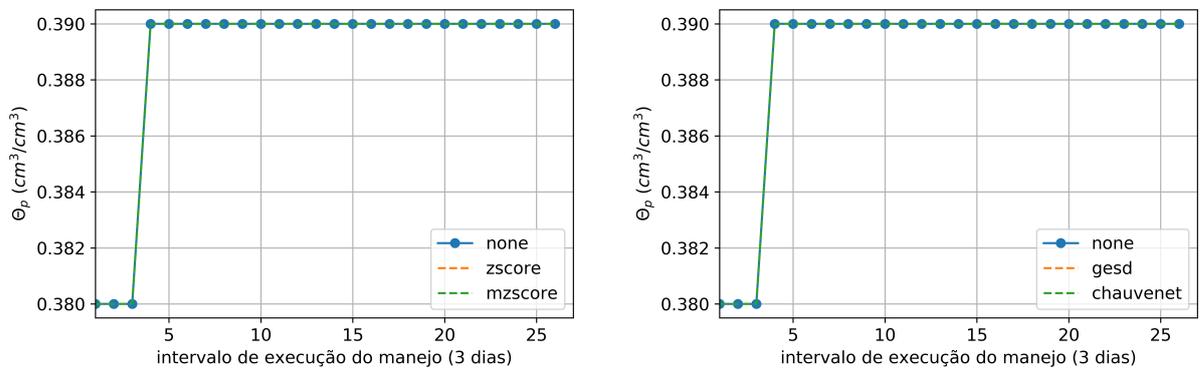


Figura 41 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem **Local** e fusão **Mínimo**

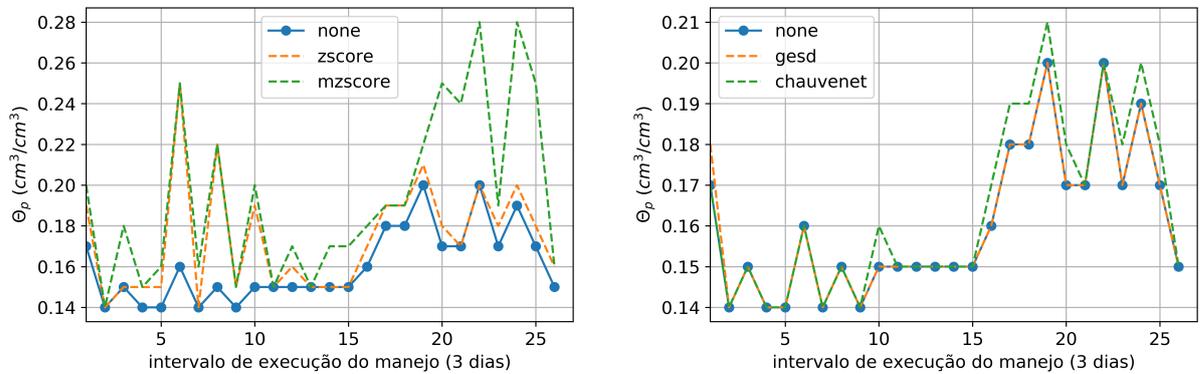


Figura 42 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem **Local** e fusão **Mediana**

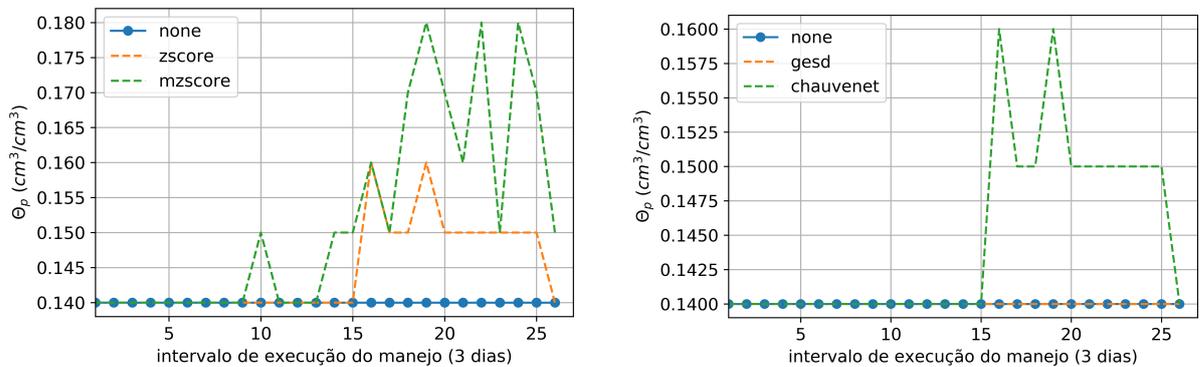


Figura 43 – Balanço Hídrico: θ_p estimada com dados ψ_p da abordagem **Local** e fusão **Máximo**

A Tabela 24 relaciona a equivalência entre os desempenhos do Balanço Hídrico com dados de umidade do solo estimados $\theta_p(\psi_p)$ e do Balanço Hídrico que usa dados reais θ_c . Nesta avaliação, considera-se desempenho equivalente quando a ITN média é diferente em ± 2 mm/3 dias. A economia da ITN média apresentada na Tabela 25 é calculada em relação ao Balanço Hídrico que não usa dados de umidade do solo. Portanto, para um campo sem sensores de umidade do solo (Campo 1 da Figura 1), é possível obter os desempenhos do Balanço Hídrico que usa dados de umidade do solo reais $\theta_c(\psi_c)$ processados apenas pela Fusão Mínimo e pela Fusão Mediana em conjunto com as técnicas GESD e Chauvenet.

Tabela 24 – Equivalência de Desempenho entre o manejo do Balanço Hídrico que usa dados reais $\theta_c(\psi_c)$ e o mesmo que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição — Métrica: Diferença de ITN de até ± 2 mm/dia

Desemp. com Dados Reais $\theta_c(\psi_c)$			Desemp. com Dados Estimados $\theta_p(\psi_p)$		
Fusão + Rem. <i>Outlier</i>	Média (mm/3 dias)	Int. Conf. (90%)	Predição + Fusão + Rem. <i>Outlier</i>	Dif. Média	Dif. Int. Conf.
Mediana + GESD ou Chauvenet	3,3	(-2,4; 9,1)	Local + Média + Zscore ou Chauvenet	-0,1	(-7,5; 7,3)
Mínimo + Nenhuma	7,6	(1,3; 13,9)	Local + Média + MZscore	0,3	(-11,0; 11,6)

Tabela 25 – Economia de ITN do manejo do Balanço Hídrico que usa dados estimados $\theta_p(\psi_p)$ pelos modelos de predição de umidade do solo com desempenho equivalente ao mesmo manejo que usa dados reais $\theta_p(\psi_p)$

Desemp. com Dados Reais $\theta_c(\psi_c)$		Desemp. com Dados Estimados $\theta_p(\psi_p)$	
Fusão + Rem. <i>Outlier</i>	Média (%)	Predição + Fusão + Rem. <i>Outlier</i>	Média (%)
Mediana + GESD ou Chauvenet	89,8	Local + Média + Zscore ou Chauvenet	90,0
Mediana + Zscore	76,5	Local + Média + MZscore	75,6

Embora, o Balanço Hídrico que usa dados $\theta_p(\psi_p)$ estimados pela abordagem de predição global não possua desempenho equivalente ao Balanço Hídrico que usa dados reais $\theta_c(\psi_c)$, seu uso é aconselhado, visto que, promove uma economia de ITN menor que 90,4%, de quando o Balanço Hídrico usa dados $\theta_c(\psi_c)$ (Tabela 23). Os valores de potencial mátrico ψ_p e umidade do solo θ_p estão, respectivamente, nas Figs. 44 e 45.

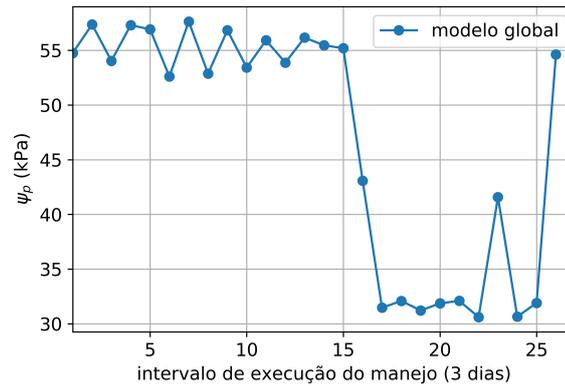


Figura 44 – Balanço Hídrico – Dados de potencial mátrico ψ_p estimado pela abordagem **Global**

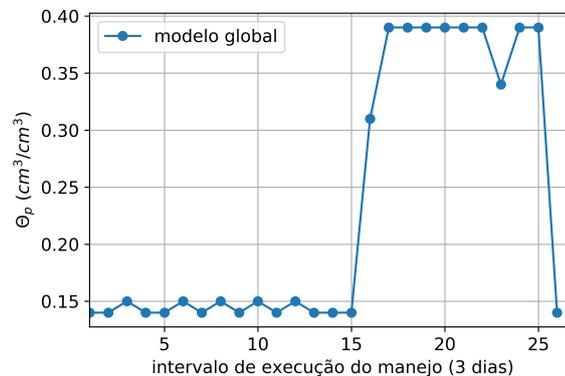


Figura 45 – Balanço Hídrico – Umidade do solo θ_p deduzida com dados ψ_p estimados pela abordagem **Global**

5.5 Síntese do Capítulo

Sobre a avaliação do manejo do Potencial Mátrico

A execução diária do manejo do Potencial Mátrico leva em consideração apenas a umidade θ deduzida a partir do potencial mátrico ψ (Eq. A.1) coletado da camada mais superficial do solo para decidir quando irrigar com um valor constante de ITN (Eq. A.3). A fusão redundante mínimo seleciona o menor valor da série de dados ψ , e, isto representa o ponto de monitoramento do campo com maior umidade do solo θ . Assim, se a fusão mínimo for usada o campo passará por dezenas de dias sem irrigação, situação que pode comprometer o desenvolvimento saudável do cultivo, embora, neste presente trabalho, não se pôde comprovar analisando biometria do cultivo, isto é, medindo a altura, envergadura e número de folhas do cultivo.

A fusão máximo seleciona o maior valor ψ , logo, a menor a umidade θ detectada no campo, e, conseqüentemente, haverá irrigação em quase todo período avaliado. Com a ITN superestimada, não haverá economia de água, situação que pode ser crítica em áreas com

escassez hídrica. Portanto, recomenda-se para o manejo do Potencial Mátrico que usa dados reais de umidade do solo $\theta_c(\psi_c)$, a fusão redundante média ou mediana, pois, promovem um aumento de 20 a 25,2% da ITN média resultante de dados de potencial mátrico reais ψ_c fusionados pelo mínimo e uma economia de 66,4% a 68,6%, em relação a fusão máximo.

Em geral, as técnicas de remoção de *outliers* otimizam a precisão do manejo do Potencial Mátrico que usa dados reais $\theta_c(\psi_c)$ diminuindo a ITN média quando executadas com a fusão redundante média, mediana e máximo. As técnicas Zscore e MZscore quando empregadas em conjunto com a fusão média ou mediana proporcionam as maiores reduções da ITN do que GESD e Chauvenet, isto é, entre 10,5% e 20,9% em relação a ITN fornecida pelo manejo do Potencial Mátrico sem nenhuma remoção de *outliers*.

O modelo de predição da abordagem local estima dados ψ_p que ao serem fusionados pela média ou máximo e tratado pela técnica Zscore, fazem com que o manejo do Potencial Mátrico obtenha um desempenho equivalente de quando o manejo usa dados reais ψ_c fusionados pela mediana e tratados pelas técnica Zscore ou MZscore.

Sobre a avaliação do manejo do Balanço Hídrico

O manejo do Balanço Hídrico necessariamente usa dados climáticos e do cultivo para calcular a ITN. No entanto, a ITN torna-se mais precisa quando são acrescentados ao Balanço Hídrico os dados de umidade do solo θ , pois, o solo funciona como um reservatório de água para o cultivo. Por isso, neste estudo, quando o Balanço Hídrico é executado a cada três dias e utiliza dados reais $\theta_c(\psi_c)$ de umidade do solo sem tratamento de *outliers* para a série de dados de potencial mátrico ψ_c , a ITN é reduzida em 76,5%, se os dados ψ_c são fusionados pelo mínimo, e em 90,4%, se os dados ψ_c são fusionados pela média, mediana ou máximo.

As técnicas de remoção de outliers aplicadas aos dados ψ_c promovem um aumento da ITN do manejo do Balanço Hídrico quando usadas em conjunto com a fusão média e mediana, no entanto, não tem efeito algum se empregadas com a fusão mínimo ou máximo. Por isso, em função do grau de economia de ITN, recomenda-se para o manejo do Balanço Hídrico que utiliza dados reais $\theta_c(\psi_c)$, as seguintes fusões redundantes e técnicas de remoção de *outliers*:

1. economia alta
 - a) de 89,3% – fusão média e técnicas Zscore ou MZscore;
 - b) de 89,8% – fusão mediana e técnicas GESD ou Chauvenet;
2. economia razoavelmente alta

- a) de 86,3% – fusão mediana e técnica Zscore;
 - b) de 87,6 % – fusão mediana e técnica MZscore;
3. economia média de 76,5% – apenas fusão mínimo.

Quanto aos modelos de predição de umidade do solo $\theta_p(\psi_p)$, a ITN do manejo do Balanço Hídrico é inibida quando os dados de potencial mátrico ψ_p estimados pela abordagem local são 1) fusionados pela mediana ou máximo e 2) pela média apenas quando tratados previamente pela técnica GESD ou nenhuma técnica de remoção de *outlier*. Assim, em termos de economia de ITN, recomenda-se para o manejo do Balanço Hídrico que usa dados $\theta_p(\psi_p)$ dos modelos de predição:

- 1. economia alta de 90,0% – abordagem de predição local, fusão média e técnicas Zscore ou Chauvenet;
- 2. economia média de 75,6% – abordagem de predição local, fusão média e técnica MZscore;
- 3. economia razoavelmente média de 56,4% – abordagem de predição global;
- 4. economia muito baixa de 1,3% – abordagem de predição local, fusão mínimo.

6 CONCLUSÃO

Neste capítulo de Conclusão, respondemos às três questões de pesquisa levantadas na Introdução desta tese, confirmamos suas respectivas hipóteses e discutimos como o *framework* Smart&Green ajuda na solução dos problemas relacionados (Seção 6.1). A Seção 6.2 é um compilado das publicações realizadas e a Seção 6.3 discute perspectivas e trabalhos futuros.

6.1 Respostas às Questões de Pesquisa (QP) e às respectivas hipóteses

6.1.1 QP#1 - confirmação da Hipótese #1

O *framework* Smart&Green contém objetos inteligentes virtuais/digitais IoT (BORGIA, 2014) que auxiliam nos serviços de execução dos manejos de irrigação do Turno de Rega, Balanço Hídrico e Potencial Mátrico (Apêndice A.3). O Smart&Green permite que seus usuários (proprietários rurais e agrônomos) configurem fontes de dados tais como a estação meteorológica brasileira mais próxima, sensores de umidade do solo (analógico/digital), diferentes tipos de cultivo, solo e sistema de irrigação presentes na infraestrutura de diferentes tipos de campos monitorados (Figura 1). Uma vez que as fontes de dados tenham sido configuradas, o Smart&Green escolhe/recomenda o manejo de irrigação mais adequado ao campo monitorado. Posteriormente, o Smart&Green sincroniza e fusiona os dados necessários e executa o manejo de irrigação em tempo real para o campo. O Smart&Green permite ainda a definição de critérios de remoção imediata de *outliers* para dados climáticos e de umidade do solo de acordo com os tipos de sensores e clima da região de um campo monitorado.

Portanto, o Smart&Green mitiga o Problema 1 (Seção 1.2.1) e a Hipótese#1 está confirmada.

6.1.2 QP#2 - confirmação parcial da Hipótese #2

O serviço de Fusão do Smart&Green executa em tempo real a fusão de dados multinível para sistemas de Internet das Coisas (TORRES, 2017; TORRES *et al.*, 2017) para os manejos de irrigação (Apêndice A.3), aplicando quando houver necessidade 1) os critérios aos dados do clima e umidade do solo e as 2) diferentes técnicas de remoção de *outliers* (Zscore, MZscore, GESD e Chauvenet) e 3) fusão redundante de dados (média, mínimo, mediana e máximo) aos dados de potencial mátrico ψ usados na dedução da umidade do solo $\theta(\psi)$. Este

serviço de fusão recomenda técnicas de pré-processamento de dados de umidade do solo segundo um estudo inédito (de acordo com nosso melhor conhecimento) feito nesta tese sobre o impacto das técnicas de remoção de *outliers* e fusão redundante de dados de umidade do solo $\theta(\psi)$ no desempenho dos manejos de irrigação do Potencial Mátrico e Balanço Hídrico. Os resultados mostram que para cada método de manejo de irrigação, existem técnicas de fusão redundante e de remoção de *outliers* que poderão favorecer ou não a otimização do *tradeoff* entre a economia de ITN e o desenvolvimento saudável do cultivo. No entanto, deve-se fazer a biometria do cultivo, isto é, medir a altura, envergadura e número de folhas das plantas à medida que se fornece a ITN de um manejo cujos dados foram pré-processados pelas técnicas recomendadas, e, então, comprovar se o desenvolvimento do cultivo. Portanto, o Smart&Green mitiga o Problema 2 (Seção 1.2.2) e a Hipótese #2 está confirmada parcialmente.

6.1.3 QP#3 - confirmação da Hipótese #3

Nossos resultados mostraram que a técnica de aprendizagem de máquina *Gradient Boosting with Regression Tree* (GBRT) superou as outras técnicas avaliadas, tanto na abordagem de predição local, quanto na abordagem de predição global. A abordagem local apresentou os melhores resultados do GBRT, embora os resultados do modelo de predição global sejam também aceitáveis. Desde modo, para campos sem sensores de umidade do solo (total como o Campo 1 da Figura 1 ou parcialmente) e com mesmo tipo de solo do Campo Experimental (EMBRAPA, 2019), o módulo de predição do Smart&Green estima a umidade do solo $\theta_p(\psi_p)$ incluída em manejos de irrigação como o Balanço Hídrico e Potencial Mátrico, mitigando o Problema 3 (Seção 1.2.3). Portanto, a Hipótese #3 também está confirmada.

6.2 Publicações

Contribuição central da tese

CAMPOS, N. G. S.; ROCHA, A. R.; GONDIM, R.; SILVA, T. L. Coelho da; GOMES, D. G. Smart & green: an internet-of-things framework for smart irrigation. *Sensors*, v. 20, n. 1, 2020. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/20/1/190>.

Artigo Tangencial à tese

CAMPOS, N. G. S.; ROCHA, A. R.; GOMES, D. G. Towards a better sensor data accu-

racy via quality of monitoring and semantic clustering. **International Journal of Communication Systems**, v. 32, n. 10, p. e3957, 2019. E3957 dac.3957. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.3957>.

Primeiro ano do doutorado

CAMPOS, N. G. S.; GOMES, D. G.; DELICATO, F. C.; NETO, A. J. V.; PIRMEZ, L.; SOUZA, J. N. de. Autonomic context-aware wireless sensor networks. **Journal of Sensors**, v. 2015, p. 1–14, 2015. Disponível em: <https://doi.org/10.1155/2015/621326>.

Core do framework Smart&Green

Disponível em: <https://github.com/nidiascampos/smartgreen/tree/master/framework>

Conjunto de dados do campo de caju experimental

Experimental Cashew Field Dataset. Disponível em: <http://smartgreen.great.ufc.br/publications>

6.3 Trabalhos Futuros

Seguem abaixo ideias para trabalhos futuros envolvendo a agricultura inteligente e o *framework* Smart&Green:

1. Especificação e implementação da comunicação do Smart&Green com nós sensores e atuadores para permitir a automatização de irrigação;
2. *Energy harvesting* (e.g. energia solar, eólica) nos nós sensores e atuadores para uma solução de IoT mais completa e sustentável (KALE *et al.*, 2017; SHARMA *et al.*, 2019a; SHARMA *et al.*, 2019b);
3. Inclusão de modelos computacionais de predição da evapotranspiração de referência ET_o (TORRES, 2017; CAMINHA *et al.*, 2017) para cada estação meteorológica brasileira (INMET, 2018). Assim, quando a conexão com a Internet falhar, o Smart&Green poderá estimar ET_o (Eq. A.5) de acordo com os manejos do Turno de Rega e do Balanço Hídrico;
4. Validação do Smart&Green com experimentos de campo;
5. Nova versão do Smart&Green com planejamento e análise de experimentos (*Design of Experiments*, DoE) em outros domínios da agricultura inteligente, e.g. gestão IoT de fertilizantes e pesticidas (TALAVERA *et al.*, 2017).

REFERÊNCIAS

- ABAYA, S.; VEGA, L. de; GARCIA, J.; MANIAUL, M.; REDONDO, C. A. A self-activating irrigation technology designed for a smart and futuristic farming. In: **2017 International Conference on Circuits, Devices and Systems (ICCDs)**. [S. l.: s. n.], 2017. p. 189–194.
- AMBLER, S. W. **Process patterns: building large-scale systems using object technology**. New York, NY, USA: Cambridge University Press, 1998. ISBN 0-521-64568-9.
- ANA. **Atlas irrigação: uso da água na agricultura irrigada**. 2017. [Nota: Agência Nacional de Águas - ANA; Acessado em 16 de abril de 2019]. Disponível em: <https://tinyurl.com/yapomxb7>.
- BALAMURUGAN, C.; SATHEESH, R. Development of raspberry pi and iot based monitoring and controlling devices for agriculture. **Journal of Social, Technological and Environmental Science**, v. 6, p. 207–215, 08 2017.
- BARNETT, V.; LEWIS, T. **Outliers in statistical data**. 2nd edition. ed. [S. l.]: John Wiley & Sons Ltd., 1978.
- BERNADI, A. C. C.; NAIME, J. M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: resultados de um novo olhar**. 2014. [Nota: livro online acessado em 12 de abril de 2019]. Disponível em: <http://tinyurl.com/y24u9qwa>.
- BHANU, B. B.; HUSSAIN, M. A.; ANDE, P. Monitoring of soil parameters for effective irrigation using wireless sensor networks. In: **2014 Sixth International Conference on Advanced Computing (ICoAC)**. [S. l.: s. n.], 2014. p. 211–215. ISSN 2377-6927.
- BORGIA, E. The internet of things vision: Key features, applications and open issues. **Computer Communications**, v. 54, p. 1 – 31, 2014. ISSN 0140-3664. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0140366414003168>.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, Oct 2001. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1010933404324>.
- BYISHIMO, A.; GARBA, A. A. Designing a farmer interface for smart irrigation in developing countries. In: **Proceedings of the 7th Annual Symposium on Computing for Development**. New York, NY, USA: Association for Computing Machinery, 2016. (ACM DEV '16). ISBN 9781450346498. Disponível em: <https://doi.org/10.1145/3001913.3006639>.
- CAETANO, F.; PITARMA, R.; REIS, P. Advanced system for garden irrigation management. **Advances in Intelligent Systems and Computing**, v. 353, p. 565–574, 01 2015.
- CAMINHA, H. D.; SILVA, T. L. C. da; ROCHA, A. R. da; LIMA, S. C. R. V. Estimating reference evapotranspiration using data mining prediction models and feature selection. **International Conference on Enterprise Information Systems**, n. 2, p. 201–213, 4 2017.
- CAMPOS, N. G. S.; GOMES, D. G.; DELICATO, F. C.; NETO, A. J. V.; PIRMEZ, L.; SOUZA, J. N. de. Autonomic context-aware wireless sensor networks. **Journal of Sensors**, v. 2015, p. 1–14, 2015. Disponível em: <https://doi.org/10.1155/2015/621326>.

- CAMPOS, N. G. S.; ROCHA, A. R.; GOMES, D. G. Towards a better sensor data accuracy via quality of monitoring and semantic clustering. **International Journal of Communication Systems**, v. 32, n. 10, p. e3957, 2019. E3957 dac.3957. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.3957>.
- CAMPOS, N. G. S.; ROCHA, A. R.; GONDIM, R.; SILVA, T. L. Coelho da; GOMES, D. G. Smart & green: an internet-of-things framework for smart irrigation. **Sensors**, v. 20, n. 1, 2020. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/20/1/190>.
- CAPRARO, F.; TOSETTI, S.; SERMAN, F. V. Supervisory control and data acquisition software for drip irrigation control in olive orchards: An experience in an arid region of argentina. **Acta Horticulturae**, v. 1057, p. 423–429, 10 2014.
- CARLESSO, R.; PETRY, M. T.; TROIS, C. The use of a meteorological station network to provide crop water requirement information for irrigation management. In: LI, D.; ZHAO, C. (Ed.). **Computer and Computing Technologies in Agriculture II, Volume 1**. Boston, MA: Springer US, 2009. p. 19–27. ISBN 978-1-4419-0209-2.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2016. (KDD '16), p. 785–794. ISBN 978-1-4503-4232-2. Disponível em: <http://doi.acm.org/10.1145/2939672.2939785>.
- COSTA, A. B. **Um sistema web para agricultura inteligente utilizando internet das coisas**. 2018. Trabalho de Conclusão de Curso (Bacharel em Engenharia da Computação)-Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, Brasil.
- CZESLA, S. **A collection of astronomy-related routines in Python**. 2013. Disponível em: <https://github.com/sczesla/PyAstronomy>. Acessado em 17 de outubro de 2019.
- ELMENREICH, W. **Sensor fusion in time-triggered systems**. 2002. Tese (Doutorado em technischen Wissenschaften)-Technischen Universität Wien, Viena, Austria.
- EMBRAPA. **Campo experimental do curu pertencente à embrapa agroindústria tropical**. 2019. Detalhes em: <http://www.cnpat.embrapa.br/conteudo52.php>. Acessado em 08 de abril de 2019.
- ENCODE. **Django rest framework**. 2011. Disponível em: <https://www.django-rest-framework.org/>. Acessado em 17 de outubro de 2019.
- FACEBOOK. **React native - a framework for building native apps using react**. 2018. Disponível em: <https://facebook.github.io/react-native/>. Acessado em 17 de outubro de 2019.
- FAO. **Crop evapotranspiration - Guidelines for computing crop water requirements**. 2014. [Nota: material online da Food and Agriculture Organization (FAO); Acessado em 11 de outubro de 2017]. Disponível em: <https://tinyurl.com/y7drx8vd>.
- FAO. **World agriculture: towards 2015/2030 - an fao perspective**. 2014.
- FERRANDEZ, J.; GARCÍA-CHAMIZO, J. M.; NIETO-HIDALGO, M.; MORA-MARTÍNEZ, J. Precision agriculture design method using a distributed computing architecture on internet of things context. **Sensors**, v. 18, p. 1731, 05 2018.

FIGUEROA, M.; POPE, C. Root system water consumption pattern identification on time series data. **Sensors (Switzerland)**, v. 17, 06 2017.

FLORES, K. O.; BUTASLAC, I. M.; GONZALES, J. E. M.; DUMLAO, S. M. G.; REYES, R. S. J. Precision agriculture monitoring system using wireless sensor network and raspberry pi local server. In: **2016 IEEE Region 10 Conference TENCN**. [S. l.: s. n.], 2016. p. 3018–3021. ISSN 2159-3450.

FOUNDATION, D. S. **Django - the web framework for perfectionists with deadlines**. 2013. Disponível em: <https://djangoproject.com>. Acessado em 17 de outubro de 2019.

FREEDMAN, D. A. **Statistical models: theory and practice**. 2. ed. [S. l.]: Cambridge University Press, 2009.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **Annals of Statistics**, v. 29, p. 1189–1232, 2000.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, v. 38, n. 4, p. 367 – 378, 2002. ISSN 0167-9473. Nonlinear Methods and Data Mining. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167947301000652>.

FRIZZONE, J. A. **Necessidade de água para irrigação (Lecture Notes)**. [S. l.]: Departamento de Engenharia de Biosistemas da Escola Superior de Agricultura Luiz de Queiroz (ESALQ) da Universidade de São Paulo (USP), 2017. Disponível em: <https://tinyurl.com/yc26lumq>. Acessado em 12 de março de 2018.

GENUCHTEN, M. T. V. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Science Society of America Journal**, v. 44, n. 5, p. 892–898, 1980. ISSN 0361-5995. Author Affiliation: Dep. of Soil and Environmental Sciences, University of California, Riverside, CA 92521, USA.

GOAP, A.; SHARMA, D. An iot based smart irrigation management system using machine learning and open source technologies. **Computers and Electronics in Agriculture**, v. 155, p. 41–49, 11 2018.

GOLDSTEIN, A.; FINK, L.; MEITIN, A.; BOHADANA, S.; LUTENBERG, O.; RAVID, G. Applying machine learning on sensor data for irrigation recommendations: Revealing the agronomist’s tacit knowledge. **Precision Agriculture**, v. 47, p. 1–24, 05 2017.

GOOGLE. **Firestore - a comprehensive app development platform**. 2016. Disponível em: <https://firebase.google.com/>. Acessado em 17 de outubro de 2019.

HALL, D. L.; LLINAS, J. An introduction to multisensor data fusion. **Proceedings of the IEEE**, v. 85, n. 1, p. 6–23, Jan 1997. ISSN 0018-9219.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: An update. **SIGKDD Explor. Newsl.**, ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, nov. 2009. ISSN 1931-0145. Disponível em: <http://doi.acm.org/10.1145/1656274.1656278>.

HAMOUDA, Y.; MSALLAM, M. Smart heterogeneous precision agriculture using wireless sensor network based on extended kalman filter. **Neural Computing and Applications**, 02 2018.

- HAMPANNAVAR, S.; YARAGATTI, U. R. Development of wsn system for precision agriculture. In: **2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)**. [S. l.: s. n.], 2015. p. 1–5.
- HEBLE, S.; KUMAR, A.; PRASAD, K. V. V. D.; SAMIRANA, S.; RAJALAKSHMI, P.; DESAI, U. B. A low power iot network for smart agriculture. **2018 IEEE 4th World Forum on Internet of Things (WF-IoT)**, p. 609–614, 2018.
- IBA, W.; LANGLEY, P. Induction of one-level decision trees. In: SLEEMAN, D.; EDWARDS, P. (Ed.). **Machine Learning Proceedings 1992**. San Francisco (CA): Morgan Kaufmann, 1992. p. 233 – 240. ISBN 978-1-55860-247-2. Disponível em: <http://www.sciencedirect.com/science/article/pii/B9781558602472500358>.
- INMET. **Estações automáticas do inmet (instituto nacional de metereologia)**. 2018. Disponível em: <https://tinyurl.com/yc47zzkk>.
- INVERTASE. **React native firebase - simple firebase integration for react native**. 2016. Disponível em: <https://rnfirebase.io/>. Acessado em 17 de outubro de 2019.
- ISSAD, H. A.; AOUDJIT, R.; RODRIGUES, J. J. A comprehensive review of data mining techniques in smart agriculture. **Engineering in Agriculture, Environment and Food**, v. 12, n. 4, p. 511 – 525, 2019. ISSN 1881-8366. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1881836619301533>.
- JAYARAMAN, P. P.; PALMER, D.; ZASLAVSKY, A.; GEORGAKOPOULOS, D. Do-it-yourself digital agriculture applications with semantically enhanced iot platform. In: **2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)**. [S. l.: s. n.], 2015. p. 1–6.
- KALE, D. S.; BHOMBE, D.; TULASKAR, D. P. Implementation of soil energy harvesting system for agriculture parameters monitoring and controlling using iot: A review. **International Research Journal of Engineering and Technology (IRJET)**, v. 4, 2017.
- KAMELIA, L.; RAMDHANI, M. A.; FAROQI, A.; RIFADIAPRIYANA, V. Implementation of automation system for humidity monitoring and irrigation system. **IOP Conference Series: Materials Science and Engineering**, IOP Publishing, v. 288, p. 012092, jan 2018. Disponível em: <https://doi.org/10.1088%2F1757-899x%2F288%2F1%2F012092>.
- KAMIENSKI, C.; SOININEN, J.-P.; TAUMBERGER, M.; TOSCANO, A.; CINOTTI, T.; DANTAS, R.; MAIA, R.; NETO, A.; FERREIRA, F. F. Smart water management platform: Iot-based precision irrigation for agriculture. **Sensors**, v. 19, p. 276, 01 2019.
- KAMILARIS, A.; GAO, F.; PRENAFETA-BOLDU, F. X.; ALI, M. I. Agri-iot: A semantic framework for internet of things-enabled smart farming applications. In: **2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)**. [S. l.: s. n.], 2016. p. 442–447.
- KANNAN, K. S.; THILAGAVATHI, G. Online farming based on embedded systems and wireless sensor networks. In: **2013 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)**. [S. l.: s. n.], 2013. p. 71–74.
- KARIMI, N.; ARABHOSSEINI, A.; KARIMI, M.; KIANMEHR, M. Web-based monitoring system using wireless sensor networks for traditional vineyards and grape drying buildings. **Computers and Electronics in Agriculture**, v. 144, p. 269–283, 01 2018.

LE, T. D.; TAN, D. H. Design and deploy a wireless sensor network for precision agriculture. In: **2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)**. [S. l.: s. n.], 2015. p. 294–299.

LIGHT, R. **Eclipse paho mqtt python client**. 2013. Disponível em: <https://pypi.org/project/paho-mqtt/>. Acessado em 17 de outubro de 2019.

LIGHT, R. A. Mosquitto: server and client implementation of the mqtt protocol. **Journal of Open Source Software**, The Open Journal, v. 2, n. 13, p. 265, 2017. Disponível em: <https://doi.org/10.21105/joss.00265>.

LUAN, Q.; FANG, X.; YE, C.; LIU, Y. An integrated service system for agricultural drought monitoring and forecasting and irrigation amount forecasting. In: **23rd International Conference on Geoinformatics, Geoinformatics 2015, Wuhan, China, June 19-21, 2015**. [S. n.], 2015. p. 1–7. Disponível em: <https://doi.org/10.1109/GEOINFORMATICS.2015.7378617>.

MAGALHÃES, R. P. **Speed prediction applied to dynamic traffic sensors and road networks**. 2018. Tese (Doutorado em Ciências da Computação)-Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, Ceará, Brasil.

MAIA, R. F.; NETTO, I.; TRAN, A. L. H. Precision agriculture using remote monitoring systems in brazil. In: **2017 IEEE Global Humanitarian Technology Conference (GHTC)**. [S. l.: s. n.], 2017. p. 1–6.

MARQUELLI, W. A.; SILVA, W. L. C.; SILVA, H. R. **Manejo da irrigação em hortaliças**. 5. ed. [S. l.]: EMBRAPA-SPI: EMBRAPA-CNPQ, 1996.

MAT, I.; KASSIM, M.; HARUN, I. A. N. Precision irrigation performance measurement using wireless sensor network. In: **International Conference on Ubiquitous and Future Networks, ICUFN**. [S. l.: s. n.], 2014. p. 154–157. ISBN 978-1-4799-3494-2.

MAT, I.; KASSIM, M. R. M.; HARUN, A. N. Precision agriculture applications using wireless moisture sensor network. In: **2015 IEEE 12th Malaysia International Conference on Communications (MICC)**. [S. l.: s. n.], 2015. p. 18–23.

MATH, R. K.; DHARWADKAR, N. V. A wireless sensor network based low cost and energy efficient frame work for precision agriculture. In: **2017 International Conference on Nascent Technologies in Engineering (ICNTE)**. [S. l.: s. n.], 2017. p. 1–6.

MAUGET, S.; LEIKER, G. The ogallala agro-climate tool. **Computers and Electronics in Agriculture**, v. 74, n. 1, p. 155 – 162, 2010. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169910001468>.

MESAS-CARRASCOSA, F.; SANTANO, D. V.; MERONO, J.; ORDEN, M. S. de la; GARCIA-FERRER, A. Open source hardware to monitor environmental parameters in precision agriculture. **Biosystems Engineering**, v. 137, p. 73 – 83, 2015. ISSN 1537-5110. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1537511015001208>.

MILLER, L.; VELLIDIS, G.; MOHAWESH, O.; COOLONG, T. Comparing a smartphone irrigation scheduling application with water balance and soil moisture-based irrigation methods: Part i—plasticulture-grown tomato. **HortTechnology**, v. 28, p. 354–361, 06 2018.

MIRANDA, F. R.; GOMES, A. R. M.; OLIVEIRA, C. H. C.; MONTENEGRO, F. M. L. B. A. A. T. Evapotranspiração e coeficientes de cultivo do coqueiro anão-verde na região litorânea do ceará. **Revista Ciência Agronômica**, v. 32, n. 2, 2007. Disponível em: <https://tinyurl.com/yd2s72lq>.

MOLIN, J. P.; AMARAL, L. R. do; COLACO, A. F. **Agricultura de precisão**. [S. l.]: Oficina de Textos, 2015.

NAKAMURA, E. F.; LOUREIRO, A. A. F.; FRERY, A. C. Information fusion for wireless sensor networks: Methods, models, and classifications. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 39, n. 3, set. 2007. ISSN 0360-0300. Disponível em: <http://doi.acm.org/10.1145/1267070.1267073>.

NAVARRO-HELLIN, H.; TORRES-SANCHEZ, R.; SOTO-VALLES, F.; ALBALADEJO-PEREZ, C.; LOPEZ-RIQUELME, J.; DOMINGO-MIGUEL, R. A wireless sensors architecture for efficient irrigation water management. **Agricultural Water Management**, v. 151, p. 64 – 74, 2015. ISSN 0378-3774. New proposals in the automation and remote control of water management in agriculture: agromotic systems. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378377414003424>.

NIST/SEMATECH. **E-handbook of statistical methods**. 2013. Seção 1.3.5.17. Detection of Outliers, Definição de Zscore e MZscore. Disponível em: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm>.

NIST/SEMATECH. **E-handbook of statistical methods**. 2013. Seção 1.3.5.17.3 Generalized ESD Test for Outliers. Disponível em: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h3.htm>.

OASIS. **Message queuing telemetry transport (mqtt)**. [S. l.]: Andrew Banks and Rahul Gupta, 2014. MQTT Version 3.1.1. (OASIS Standard). Disponível em: <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html>.

ORACLE. **Mysql community edition**. 2019. Disponível em: <https://www.mysql.com/products/community/>. Acessado em 17 de outubro de 2019.

PANG, Z.; CHEN, Q.; HAN, W.; ZHENG, L. Value-centric design of the internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion. **Information Systems Frontiers**, Kluwer Academic Publishers, Hingham, MA, USA, v. 17, n. 2, p. 289–319, abr. 2015. ISSN 1387-3326. Disponível em: <http://dx.doi.org/10.1007/s10796-012-9374-9>.

PATOKAR, A.; GOHOKAR, V. **Precision agriculture system design using wireless sensor network**. [S. l.: s. n.], 2018. 169-177 p. ISBN 978-981-10-5507-2.

PAVON-PULIDO, N.; LOPEZ-RIQUELME, J. A.; TORRES, R.; MORAIS, R.; PASTOR, J. A. New trends in precision agriculture: a novel cloud-based system for enabling data storage and agricultural task planning and automation. **Precision Agriculture**, v. 18, 08 2017.

POPOVIC, T.; LATINOVIC, N.; PESIC, A.; ZECEVIC, Z.; KRSTAJIC, B.; ĐUKANOVIC, S. Architecting an iot-enabled platform for precision agriculture and ecological monitoring: A case study. **Computers and Electronics in Agriculture**, v. 2017, p. 255–265, 06 2017.

- PRESSMAN, R. **Software engineering: a practitioner's approach**. 7. ed. New York, NY, USA: McGraw-Hill, Inc., 2010. ISBN 0073375977, 9780073375977.
- QUINLAN, R. J. Learning with continuous classes. In: **5th Australian Joint Conference on Artificial Intelligence**. Singapore: World Scientific, 1992. p. 343–348.
- RAJKUMAR, M. N.; ABINAYA, S.; KUMAR, V. V. Intelligent irrigation system — an iot based approach. In: **2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)**. [S. l.: s. n.], 2017. p. 1–5.
- REALM. **Realm: creative mobile apps in a fraction time**. 2014. Disponível em: <https://realm.io/>. Acessado em 17 de outubro de 2019.
- REITZ, K. **Requests: http for humans**. 2011. Disponível em: <https://pypi.org/project/requests/>. Acessado em 17 de outubro de 2019.
- RICHARDS, M. **Pyeto implements methods for estimating evapotranspiration**. 2015. Disponível em; <https://pyeto.readthedocs.io/en/latest/overview.html>. Acessado em 17 de outubro de 2019.
- RICHARDSON, L. **Beautiful soup: an screen-scraping library**. 2014. Disponível em: <https://pypi.org/project/beautifulsoup4/>. Acessado em 17 de outubro de 2019.
- ROCHA, A. R. **Um arcabouço de internet das coisas para agricultura inteligente**. 2017. Disponível em: <https://tinyurl.com/yda2m6dy>. [Nota: Projeto de Pesquisa].
- SAWANT, S.; DURBHA, S.; JAGARLAPUDI, A. Interoperable agro-meteorological observation and analysis platform for precision agriculture: A case study in citrus crop water requirement estimation. **Computers and Electronics in Agriculture**, v. 138, p. 175–187, 06 2017.
- SHARMA, H.; HAQUE, A.; JAFFERY, Z. Smart agriculture monitoring using energy harvesting internet of things (eh-iot). **World Scientific News - An International Scientific Journal**, 121 2019.
- SHARMA, H.; HAQUE, A.; JAFFERY, Z. A. Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring. **Ad Hoc Networks**, v. 94, p. 101966, 2019. ISSN 1570-8705. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1570870519300952>.
- SHELBY, Z.; HARTKE, K.; BORMANN, C. **The constrained application protocol (coap)**. [S. l.]: RFC Editor, 2014. RFC 7252. (Request for Comments, 7252). Available at: <https://rfc-editor.org/rfc/rfc7252.txt>.
- SILVA, J. P. **Uso de biocarvão e hidrogel na retenção hídrica e crescimento inicial de cajueiro precoce**. 2017. Trabalho de Conclusão de Curso (Bacharel em Engenharia Agrícola)-Centro de Ciências Agrárias, Universidade Federal do Ceará, Fortaleza, Brasil.
- SILVA NETO, M.; GOMES, D.; SOARES, J. . Credibility on crowdsensing data acquisition. **Journal of Communication and Information Systems**, v. 34, p. 248–269, 2019.
- SUNDMAEKER, H.; VERDOUW, C.; WOLFERT, S.; PEREZ-FREIRE, L. **Internet of food and farm 2020**. [S. l.: s. n.], 2016. 129-151 p.

TALAVERA, J. M.; TOBON, L. E.; GOMEZ, J. A.; CULMAN, M. A.; ARANDA, J. M.; PARRA, D. T.; QUIROZ, L. A.; HOYOS, A.; GARRETA, L. E. Review of iot applications in agro-industrial and environmental fields. **Computers and Electronics in Agriculture**, v. 142, p. 283 – 297, 2017. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169917304155>.

TESTEZLAF, R. **Irrigação: métodos, sistemas e aplicações**. 2017. [Nota: livro online disponível em: <https://tinyurl.com/y8vya3jr>. Acessado em 12 de abril de 2019].

TORRES, A. B. B. **Fusão de dados multinível para sistemas de internet das coisas em agricultura inteligente**. 2017. Dissertação (Mestrado em Engenharia de Teleinformática)-Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, Ceará, Brasil.

TORRES, A. B. B.; FILHO, J. A.; ROCHA, A. R. da; GONDIM, R. S.; SOUZA, J. N. de. Outlier detection methods and sensor data fusion for precision agriculture. **XXXVII Congresso da Sociedade Brasileira de Computação**, p. 928–937, 2017.

VAISHALI, S.; SURAJ, S.; VIGNESH, G.; DHIVYA, S.; UDHAYAKUMAR, S. Mobile integrated smart irrigation management and monitoring system using iot. In: **2017 International Conference on Communication and Signal Processing (ICCSP)**. [S. l.: s. n.], 2017. p. 2164–2167.

WANG, M.; PERERA, C.; JAYARAMAN, P. P.; ZHANG, M.; STRAZDINS, P.; SHYAMSUNDAR, R.; RANJAN, R. City data fusion: Sensor data fusion in the internet of things. **Int. J. Distrib. Syst. Technol.**, IGI Global, Hershey, PA, USA, v. 7, n. 1, p. 15–36, jan. 2016. ISSN 1947-3532. Disponível em: <http://dx.doi.org/10.4018/IJDST.2016010102>.

WANG, Y.; WITTEN, I. H. Induction of model trees for predicting continuous classes. In: **Poster papers of the 9th European Conference on Machine Learning**. [S. l.]: Springer, 1997.

WASILAK, M.; AMSUSS, C. **Aiocoap – the python coap library**. 2012. Disponível em: <https://github.com/chrysn/aiocoap#aiocoap----the-python-coap-library>. Acessado em 17 de outubro de 2019.

WOLFERT, S.; GE, L.; VERDOUW, C.; BOGAARDT, M.-J. Big data in smart farming – a review. **Agricultural Systems**, v. 153, p. 69 – 80, 2017. ISSN 0308-521X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0308521X16303754>.

APÊNDICE A – MANEJOS DA IRRIGAÇÃO

Neste apêndice, fornecemos detalhes dos dados climáticos, do cultivo e do sistema de irrigação (Seção A.1) e do solo (Seção A.2) usados nos diferentes tipos de manejos de irrigação estudados (Seção A.3).

A.1 Dados climáticos, do cultivo e do sistema de irrigação

Assumindo que as coordenadas geográficas da propriedade rural são definidas por latitude, longitude e altitude e exista uma estação meteorológica ms mais próxima à propriedade, definimos o **conjunto de dados climáticos** comum a todos os campos da propriedade por

$M_{ms} = \{T_{max}, T_{min}, RH_{max}, RH_{min}, R_n, u_2, P, PREC\}$, onde:

- T_{max} e T_{min} são a temperatura do ar máxima e mínima, respectivamente;
- RH_{max} e RH_{min} são a umidade relativa do ar máxima e mínima, respectivamente;
- R_n é o saldo de radiação solar;
- u_2 é a velocidade do vento;
- P é a pressão atmosférica;
- $PREC$ é a precipitação pluviométrica.

De acordo com a Figura 1, cada campo possui uma cultura/cultivo diferente. Logo, para cada campo existe um $K_{c, fase}$, isto é, um **coeficiente da cultura** sem restrição hídrica e em ótimas condições de desenvolvimento. O K_c varia de acordo com o estágio/fase de desenvolvimento de um tipo de cultivo. Cada estágio é caracterizado pelo número de dias após o plantio da cultura fazendo que o valor de K_c varie conforme a sua curva, como ilustra a Figura 46.

Cada campo da Figura 1 possui um sistema de irrigação caracterizado por dados constantes definidos por $DSI = \{P_{irri}, Ef\}$, onde a P_{irri} é **precipitação** e Ef é a **eficiência do sistema de irrigação** do Campo em questão. A P_{irri} é a quantidade de água por unidade de tempo que o sistema de irrigação proporciona ao cultivo, sendo definido através de testes específicos no sistema. A Ef é a relação entre (1) a quantidade de água aplicada ao cultivo e que ficou disponível no solo e (2) aquela quantidade de água que foi perdida pela evaporação ou pela deriva do vento ou ainda pela percolação profunda. Na irrigação por microaspersão, Ef varia entre 0,70 e 0,90 (FRIZZONE, 2017). Por exemplo, um sistema de irrigação por microaspersão é caracterizado pela $Ef = 81,85\%$ e $P_{irri} = 12,5mm/h$.

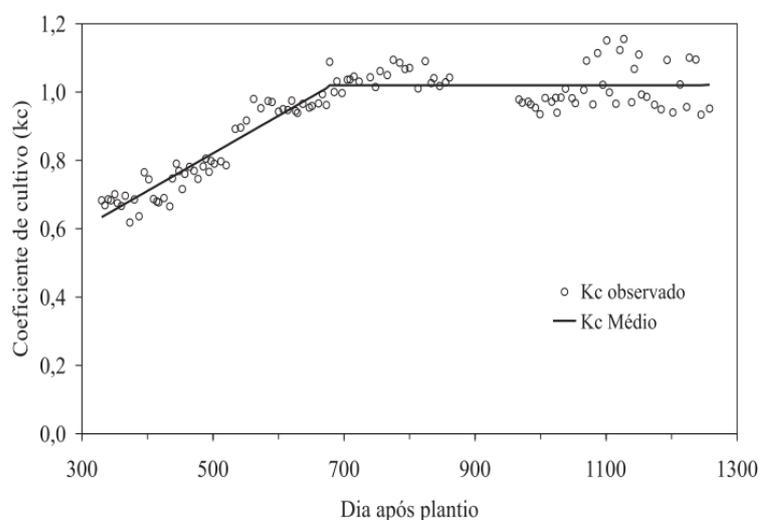


Figura 46 – Curva do coeficiente de cultivo (Kc) do coqueiro anão-verde. Fonte: (MIRANDA *et al.*, 2007)

A.2 Dados do solo

O solo é formado por materiais sólidos (minerais e orgânicos) e poros preenchidos por água e ar (vapor d'água, O_2 , CO_2 , N_2 , NH_3 , etc.). Logo, o solo é considerado um reservatório de água e de nutrientes de um cultivo. O solo, a água e o cultivo podem ser tratados como um sistema dinâmico contínuo, onde a água movimenta-se pelo solo até as raízes das plantas, passando por dentro delas e saindo para a atmosfera. Deste modo, a água modifica a energia potencial com que está retida no solo (FRIZZONE, 2017).

O **potencial mátrico** (ψ_m) refere-se a interação entre a matriz do solo (material mineral e sólido) e a água, incluindo forças associadas à absorção e capilaridade das plantas, resultando na retenção de água no solo. Portanto, podemos considerar ψ_m como uma função contínua da umidade do solo que existe apenas em solos não saturados. A pressão é negativa em solos não saturados ($\psi_m < 0$), visto que, o ar entra nos espaços porosos do solo enquanto a água é drenada. Após a chuva ou a irrigação, o solo pode se tornar saturado, isto é, quando a água ocupa todo espaço poroso e o $\psi_m = 0$ (ponto de saturação). O solo que está abaixo do nível do lençol freático é também considerado saturado, pois, nesse caso, a pressão será positiva e $\psi_m > 0$ (FRIZZONE, 2017).

Existem diversos tipos de solos classificados quanto à sua textura, como arenoso, areia franca, franco arenoso, argiloso, franca siltosa, entre outros. A estrutura do solo pode ser mantida ou melhorada pelo uso adequado das práticas agrícolas como também pode ser deteriorada com o manejo inadequado. Logo, cada tipo de solo possui uma **curva característica**

de umidade do solo, também chamada de **curva de retenção de água**. Isto porque a textura e estrutura dos diferentes tipos de solo afetam o tamanho e o número de poros contidos nele. Portanto, diferentes tipos de solos retêm a mesma quantidade de água com diferentes níveis de energia, conseqüentemente, existirão diferentes relações entre o teor de água e potencial mátrico ψ_m (FRIZZONE, 2017).

Existem vários modelos matemáticos para determinar a curva de retenção de água de um solo. Entre eles, o modelo de Genuchten (1980) que expressa o volume de água por volume do solo, θ , em função do potencial mátrico ψ_m :

$$\theta(\psi_m) = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha * |\psi_m|)^n]^{1 - \frac{1}{n}}}, \quad (\text{A.1})$$

na qual θ_r é volume de água residual, θ_s é o volume de água de saturação, α é o escalonador de ψ_m e n é um parâmetro relacionado ao formato da curva.

A Figura 47 mostra-nos a curva de retenção de água na profundidade de 15 cm de um terreno franco-arenoso. Nota-se a relação da umidade do solo θ , calculada em função do potencial mátrico ψ_m através da Equação A.1. Amostras de ψ_m são provenientes de leituras de tensiômetros e os demais parâmetros foram definidos após análises laboratoriais do solo (SILVA, 2017).

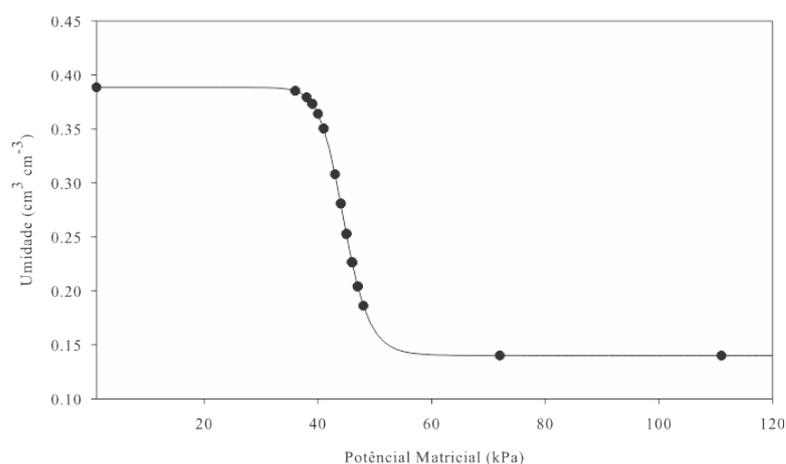


Figura 47 – Curva característica de solo franco-arenoso. Fonte: (SILVA, 2017)

Tensiômetros consistem em tubos de PVC cheios de água com uma cápsula porosa em uma das extremidades e um vacuômetro na outra extremidade. A parte porosa do tensiômetro é inserida no solo em uma profundidade desejada, enquanto o vacuômetro fica a certa altura da superfície do solo. Quando o solo está saturado, existirá um equilíbrio entre a água do tensiômetro e a água do solo fazendo que vacuômetro meça o potencial mátrico $\psi_m = 0$. Porém,

se o solo não está saturado, a água tenderá a se movimentar do tensiômetro para o solo, indicando que o potencial da água dentro do tensiômetro é maior que o potencial da água no solo, ou seja, o mesmo que $\psi_m < 0$ (FRIZZONE, 2017).

Por exemplo, o **Campo 2** da Figura 1 possui seis baterias de tensiômetros analógicos medindo ψ_m do solo nas camadas z_1 e z_2 . Para coletar seus dados, é necessário que um usuário dirija-se ao campo e anote o horário e o valor do ψ_m atual de cada tensiômetro. O **Campo 3** possui cinco baterias de tensiômetros digitais¹ que medem o ψ_m do solo nas camadas z_1 , z_2 e z_3 . Cada bateria está conectada a um nó sensor que coleta o potencial mátrico ψ_m das três camadas do solo enviando-os em um intervalo de tempo fixo por uma rede sem fio até o gateway (Seção 3.4).

Portanto, o **conjunto de dados variáveis do solo** é definido:

- para o Campo 2, como $BT_{c_2} = \{B_{1,c_2} = \{\psi_{m1,c_2,z_1}, \psi_{m1,c_2,z_2}\}, B_{2,c_2} = \{\psi_{m2,c_2,z_1}, \psi_{m2,c_2,z_2}\}, \dots, B_{6,c_2} = \{\psi_{m6,c_2,z_1}, \psi_{m6,c_2,z_2}\}\}$;
- para o Campo 3, como $BT_{c_3} = \{B_{1,c_3} = \{\psi_{m1,c_3,z_1}, \psi_{m1,c_3,z_2}, \psi_{m1,c_3,z_3}\}, B_{2,c_3} = \{\psi_{m2,c_3,z_1}, \psi_{m2,c_3,z_2}, \psi_{m2,c_3,z_3}\}, \dots, B_{5,c_3} = \{\psi_{m5,c_3,z_1}, \psi_{m5,c_3,z_2}, \psi_{m5,c_3,z_3}\}\}$.

O pré-processamento e fusão dos dados variáveis do solo são necessários para retirada de outliers antes de aplicá-los aos modelos matemáticos dos manejos de irrigação (Seção A.3).

Para cada campo, existem ainda **dados constantes do solo** tais como a umidade da capacidade de campo e umidade crítica. A **capacidade de campo (CC)** representa a condição do solo a uma profundidade z após a drenagem do excesso de água. A drenagem ocorre da superfície para outras camadas mais profundas do solo até a água atingir equilíbrio, ou seja, até o potencial mátrico ψ_m não mais varie. A **umidade da capacidade de campo** θ_{cc} pode ser determinada indiretamente em laboratório através da curva de retenção de água do solo (FRIZZONE, 2017).

A **umidade crítica** θ_{cr} , também denominada umidade ideal para irrigação, consiste no teor de água do solo a partir do qual o rendimento da cultura começa a ser reduzido havendo a possibilidade de diminuição da evapotranspiração (consumo de água pelo cultivo). θ_{cr} é necessária apenas para a camada de solo mais superficial que é monitorada por tensiômetro (FRIZZONE, 2017).

Assim, são definidos os dados constantes do solo (DCS) para:

- Campo 2, $DCS_{c_2} = \{\psi_{cr2}, \psi_{cc2,z_1}$ e $\psi_{cc2,z_2}\}$;
- Campo 3, $DCS_{c_3} = \{\psi_{cr3}, \psi_{cc3,z_1}, \psi_{cc3,z_2}, \psi_{cc3,z_3}\}$;

¹ Irrrometer Watermark 200SS: <https://tinyurl.com/ybtknd8r>

A.3 Métodos para manejo da irrigação

A irrigação faz com que o solo atinja sua condição de campo (CC) na qual reserva-se água o suficiente para a cultura em seu estágio de desenvolvimento. Existem três tipos de métodos usados no manejo da irrigação para determinar quando irrigar, a quantidade de água e o tipo de irrigação: (1) **Turno de Rega Calculado**, (2) **Balanço Hídrico** e (3) **Potencial Mátrico**.

A.3.1 Forma de Distribuição de Água

Eles são selecionados de acordo com a forma de distribuição de água no cultivo (FRIZZONE, 2017):

- para a distribuição de água **em rotação**, isto é, quando o usuário do sistema de irrigação recebe água em dias fixos, o método do **Balanço Hídrico** pode ser utilizado apoiando-se em um **Turno de Rega Fixo**;
- para a distribuição de água **por demanda**, a água está sempre disponível, ou pelo projeto de irrigação ou bombeada pelo próprio usuário sendo usada quando houver necessidade do cultivo. Portanto, os métodos do **Balanço Hídrico** e **Potencial Mátrico** são os mais indicados.

A.3.2 Método do Turno de Rega Calculado

O turno de rega (TR), isto é, a frequência das irrigações em dias, deve ser calculado previamente para cada estágio de desenvolvimento da planta a partir de :

$$TR(dias) = \frac{AFD(mm)}{ET_c(mmd^{-1})} \quad (A.2)$$

Sem considerar as variações da demanda de água do turno, esse método considera como valor constante a **água facilmente disponível (AFD)** do solo em cada estágio de desenvolvimento do cultivo (Equação A.3). Ele calcula também a **quantidade de água** usando a média mensal e histórica da **evapotranspiração da cultura** ET_c . Logo, esse tipo de manejo poderá provocar um déficit ou excesso de água no cultivo e deve ser utilizado apenas quando não houver dados/equipamentos (ψ_m , tensiômetros) necessários para implementar o método do Balanço Hídrico ou Potencial Mátrico (métodos mais eficientes) (MAROUELLI *et al.*, 1996).

A **água facilmente disponível (AFD)** ou o **déficit máximo de água** (D_{max}) permi-

tido no solo à uma profundidade z é dado por:

$$AFD = D_{max} = (\theta_{cc} - \theta_{cr}) \cdot z \quad (A.3)$$

No método de irrigação do TR, z representa a profundidade da raiz de acordo com o estágio/fase atual de desenvolvimento do cultivo.

A **evapotranspiração da cultura** (ET_c) é o suprimento de água necessário para os processos fisiológicos da planta em um determinado clima (ANA, 2017). Ela depende do monitoramento de variáveis climáticas e do cultivo a ser irrigado e definida por:

$$ET_c = k_c \cdot ET_o, \quad (A.4)$$

em que ET_o é a **evapotranspiração de referência** (ou potencial) nas condições climáticas atuais e K_c é o **coeficiente de cultura** sem restrição hídrica e em condições ótimas de desenvolvimento, descrita na Seção A.1.

Existem diversos métodos para calcular ET_o , sendo alguns mais específicos usados apenas em determinados climas, e outros universais que podem ser usados em qualquer clima. O método de Penman-Monteith (FAO, 2014a) consiste em um modelo matemático usado mundialmente para o cálculo universal de ET_o descrito pela seguinte equação:

$$ET_o = \frac{0,408 \cdot \Delta \cdot (R_n - G) + \frac{\gamma \cdot 900 \cdot u_2 \cdot (e_s - e_a)}{T + 273}}{\Delta + \gamma \cdot (1 + 0,34 \cdot u_2)}, \quad (A.5)$$

onde:

- R_n é a radiação líquida solar [$MJm^{-2}d^{-1}$];
- G é a densidade no fluxo de calor no solo [$MJm^{-2}d^{-1}$];
- T é a temperatura média [$^{\circ}C$] a uma altura z , dada por: $T = \frac{(T_{max} + T_{min})}{2}$;
- u_2 é a velocidade do vento a uma altura z [ms^{-1}];
- e_s é a pressão do vapor de saturação média [kPa], dada por $e_s = \frac{e^{\circ}(T_{max}) - e^{\circ}(T_{min})}{2}$;
- $e^{\circ}(T)$ é a pressão do vapor na temperatura do ar (T) [kPa], dada por $e^{\circ}(T) = 0,6108 \cdot \exp\left[\frac{17,27 \cdot T}{T + 237,3}\right]$;
- e_a é a pressão do vapor atual [kPa], dado por $e_a = \frac{e^{\circ}(T_{min}) \cdot \frac{RH_{max}}{100} + e^{\circ}(T_{max}) \cdot \frac{RH_{min}}{100}}{2}$, onde RH é a umidade relativa do ar;
- γ é o coeficiente psicométrica [$kPa^{\circ}C^{-1}$], dado por $\gamma = 0,665 \cdot 10^{-3} \cdot P$, onde P é a pressão atmosférica [kPa]

O modelo de Penman-Monteith requer um conjunto de dados coletados por estações meteorológicas definido por $DM_{em} = \{R_n, T_{max}, T_{min}, u_2, RH_{max}, RH_{min}, P, PREC\}$ (Seção A.1).

O **tempo de irrigação** calculado pelo método do Turno de Rega consiste na razão entre a **necessidade hídrica bruta (NHB)** do cultivo e a **precipitação efetiva** P_{irri} do sistema de microaspersão (Equação A.6). NHB consiste na diferença do consumo de água do cultivo definido por ET_c e a precipitação pluviométrica $PREC$.

$$t_{irri,TR} = \frac{NHB}{P_{irri}} = \frac{ET_c - PREC}{P_{irri}} \quad (A.6)$$

A.3.3 Método do Balanço Hídrico

O **método do balanço do hídrico** compara as entradas e saídas de água do solo considerando a ET_c diária, a precipitação efetiva P_{irri} e a variação de umidade do solo ΔA . Primeiro, define-se a **água facilmente disponível** (AFD, Equação A.3) para cada estágio de desenvolvimento do cultivo.

Segundo, calcula-se a **IRN** que é a quantidade de água necessária para suprir a evapotranspiração de uma cultura. Quando os dados climáticos M_{ms} (estação meteorológica ms , Seção A.1) e de umidade do solo BT_c (baterias de tensiômetros, Seção A.2) estão disponíveis, IRN é determinada por:

$$IRN = \sum_{i=1}^n (ET_{c,i} - P_i) - \Delta A \quad (A.7)$$

A IRN leva em consideração o número de dias decorridos desde a última irrigação, n . Assim, a evapotranspiração ET_c representa o consumo de água necessário da cultura no dia i o qual é suprido pela **precipitação** P em um dia i , pela variação do teor de água na zona radicular em um período considerado, $\Delta A = D_a$, ou seja, o **déficit atual de água do solo** D_a dado por:

$$D_a = (\theta_{cc} - \theta_a) \cdot z \quad (A.8)$$

Onde θ_{cc} é a **umidade da capacidade campo** (CC, Seção A.2) do solo a uma profundidade z (geralmente, representa a camada de solo mais superficial que é monitorada por um tensiômetro); θ_a é a **umidade atual** do solo a uma profundidade z . Se apenas os dados climáticos estão disponíveis, desconsideramos os dados de umidade do solo ($\Delta A = 0$).

Terceiro, compara-se a **irrigação real necessária** (IRN, Equação A.7) relativa aos dias desde a última irrigação com a AFD do estágio de desenvolvimento do cultivo. Se $IRN \leq AFD_{estagio}$, então, é o momento de irrigar com a quantidade de água da **irrigação total necessária (ITN)**, que consiste na quantidade de água provida totalmente pela irrigação para compensar a perda de água devido a evaporação, percolação e deriva do vento (FRIZZONE, 2017). ITN é definida por:

$$ITN = \frac{IRN}{Ef} \quad (A.9)$$

Onde Ef é a eficiência do sistema de irrigação presente em cada campo (Seção A.1).

A frequência do Balanço Hídrico pode ser variável ou fixo. Neste último caso, calcula-se o turno de rega por:

$$TR(dias) = \frac{IRN(mm)}{ET_c(mmd^{-1})} \quad (A.10)$$

onde, IRN é dada pela Equação A.7.

Conforme a Equação A.11, o tempo de irrigação $t_{irri,BH}$ é calculado pelo manejo do balanço hídrico através da razão entre ITN (Equação A.9) e a precipitação P_{irri} do sistema de microaspersão que fornece água ao cultivo.

$$t_{irri,BH} = \frac{ITN}{P_{irri}} \quad (A.11)$$

A.3.4 Método do Potencial Mátrico

O **método do Potencial Mátrico** determina o momento de irrigar monitorando o **potencial mátrico atual** (ψ_a) do solo da camada z onde existe a maior concentração de raízes do cultivo (geralmente na camada de solo mais superficial monitorada por tensiômetro). Vale ressaltar que o sistema de irrigação deve aumentar a **umidade atual** $\theta_a(\psi_a)$ até a umidade da capacidade de campo $\theta_{cc}(\psi_{cc})$ do solo para que o cultivo extraia a água pela sua zona radicular durante o processo de evapotranspiração (FRIZZONE, 2017).

Assim, o **momento de irrigar** é reconhecido quando a $\theta_a \leq \theta_{cr}$. A **quantidade de água** é dada pela ITN (Equação A.9) em função da IRN determinada da seguinte forma:

- Caso os dados sejam provenientes de apenas um tensiômetro, $IRN = AFD$ (Equação A.3);

- Se existir uma bateria de tensiômetros no mesmo ponto de monitoração, IRN é calculado por:

$$IRN = AFD_{13} + D_{a,23} = (\theta_{cc,13} - \theta_{cr,13})z_{13} + (\theta_{cc,23} - \theta_{a,23})z_{23} \quad (A.12)$$

Recomenda-se que no mínimo exista uma bateria de dois tensiômetros em um mesmo ponto de monitoramento para mensurar o grau de umidade nas camadas de solo a 13 e 23 da zona radicular do cultivo (FRIZZONE, 2017). Assim, o tensiômetro na camada mais superficial do solo determina o momento de irrigar ($\theta_a \leq \theta_{cr}$) e outro controla a quantidade de água com o primeiro pelo **déficit atual de água do solo** D_a (Equação A.8). O manejo do Potencial Mátrico deduz o tempo de irrigação $t_{irri,PM}$ com a Equação A.11, utilizada também pelo manejo do Balanço Hídrico (Seção A.3.3).

APÊNDICE B – PROBLEMA: PREDIÇÃO DE UMIDADE DO SOLO

A umidade do solo influencia na quantidade de água da irrigação fornecida a um cultivo como mostram as Equações A.3, A.7, A.12, pois, os manejos da irrigação consideram o nível de retenção de água no solo. Assim, a agricultura inteligente pode reduzir o consumo de água na irrigação avaliando se existe água disponível no solo. Tensiômetros (Seção A.2) são sensores de solo que detectam o potencial mátrico ψ usado na estimação da umidade (Equação A.1) em diferentes camadas do solo e pontos de monitoramento em um campo.

Seja $B = \{b_1, \dots, b_n\}$ o conjunto de tensiômetros monitorando a umidade do solo em um campo em n pontos diferentes. Um tensiômetro b_j coleta amostras $(\psi_{j,15}, \psi_{j,45}, \psi_{j,75})$, que são as leituras do potencial mátrico ψ a 15cm, 45cm e 75cm de profundidade, respectivamente, em um ponto j . Seja $W = \{w_1, \dots, w_n\}$ o conjunto da quantidade de água dada a um cultivo em n pontos diferentes. O manejo da irrigação calcula W usando as Equações A.12 e A.7, por exemplo.

Os dados meteorológicos são necessários visto que as condições climáticas afetam a necessidade de água do cultivo e a precipitação pode aumentar a umidade do solo. Assim, nós assumimos que exista uma estação meteorológica ms próximo do campo que forneça o conjunto de dados meteorológicos $M_{ms} = \{T_{min}, T_{max}, RH_{min}, RH_{max}, Rn, u_2, P, PREC\}$ da Seção A.1.

Diariamente, produz-se observações do campo e do clima. Denomina-se O a coleção de observações que consiste em (t, B, M_{ms}, W, k_c) registradas em um tempo t de um dia.

Os manejos de irrigação apresentados na Seção A.3 decidem iniciar a irrigação analisando a umidade atual θ_a da camada de solo mais superficial z monitorada por tensiômetros a 15 cm de profundidade neste trabalho. Assim, o potencial mátrico $\psi_{i,15}$ é verificado para se obter $\theta_a(\psi_{i,15})$ por meio do modelo de Van Genuchten (Equação A.1).

Seja $y(i, M_{ms}, w_i, k_c)$ a função que dado um ponto de monitoramento i , variáveis meteorológicas M_{ms} , a quantidade de água w_i dado a um ponto i e o coeficiente do cultivo k_c (mais detalhes na Seção A.1) retorne $\psi_{i,15}$ (o potencial mátrico a 15cm de profundidade). Então, $y(i, M_{ms}, w_i, k_c) = \psi_{i,15}$.

O problema de predição de umidade do solo é definido como o problema de encontrar uma função f acurada para predizer $y(i, M_{ms}, w_i, k_c)$ dada todas as observações registradas em O .

O problema de predição de umidade do solo deve encontrar uma função de predição

\hat{f} a partir de uma classe com todas as possíveis funções de predição H tal que:

$$\hat{f} = \operatorname{argmin} \Delta(f), \quad (\text{B.1})$$

Onde $f \in H$ e Δ é a função de perda que avalia a qualidade de uma função preditiva candidata f a partir das observações em O . Neste trabalho, nós usamos as funções de perda RMSE (Root Mean Squared Error) e MAE (Mean Absolute Error) definidas como:

$$\Delta_{RMSE}(f) = \sqrt{\frac{1}{|O'|} \sum_{(i, M_{ms}, w_i, k_c) \in O'} (f(i, M_{ms}, w_i, k_c) - y(i, M_{ms}, w_i, k_c))^2} \quad (\text{B.2})$$

$$\Delta_{MAE}(f) = \frac{1}{|O'|} \sum_{(i, M_{ms}, w_i, k_c) \in O'} |f(i, M_{ms}, w_i, k_c) - y(i, M_{ms}, w_i, k_c)| \quad (\text{B.3})$$

onde O' é o conjunto de observações usadas para avaliar a qualidade da predição e $f(i, M_{ms}, w_i, k_c)$ é a estimaco retornada pela funo f para $y(i, M_{ms}, w_i, k_c)$. Quanto menor valor gerado pelas funoes de perda, melhor ser o desempenho preditivo de f .

APÊNDICE C – BANCO DE DADOS DO SMART&GREEN

Objetivo: Realizar controle centralizado de dados de monitoramento climático, cultivo e solo presente em campos agrícolas de uma propriedade rural usados em aplicações de manejos da irrigação e predição de umidade do solo. O framework ainda oferece serviço de detecção e remoção de *outliers* que possibilita ao usuário a configuração de limites máximos e mínimos de variáveis climáticas e do solo.

C.1 Regras de Negócio Principais do Smart&Green

1. São tipos de usuários do Smart&Green: regular e especialista.
2. Uma propriedade rural tem um ou mais campos.
3. A propriedade rural é localizada pelas coordenadas geográficas latitude, longitude e altitude.
4. Existe uma estação meteorológica do INMET próxima a propriedade rural que fornece dados climáticos para o cálculo da evapotranspiração do cultivo.
5. Os dados climáticos são temperatura máxima do ar ($^{\circ}\text{C}$), temperatura mínima do ar ($^{\circ}\text{C}$), umidade relativa do ar máxima (%), umidade relativa do ar mínima (%), saldo de radiação solar ($\text{MJm}^{-2}\text{d}^{-1}$), velocidade do vento (ms^{-1}), pressão atmosférica (kPa), precipitação pluviométrica(mmd^{-1});
6. Cada campo de uma propriedade rural possui um cultivo, um sistema de irrigação e um solo.
7. Um cultivo possui um coeficiente (valor de 0 a 1,5) para fase inicial, média e final de desenvolvimento (dias) os quais compõem uma curva.
 - a) Usuários especialistas podem cadastrar a curva do coeficiente do cultivo em geral;
 - b) Usuários regulares podem configurar a fase de desenvolvimento do cultivo (dias) de um campo em sua propriedade rural. Para isso é necessário que a curva do coeficiente do cultivo já esteja cadastrada.
8. Um cultivo possui zero ou uma umidade da condição crítica do solo relacionada à fase de desenvolvimento do cultivo.
9. O sistema de irrigação de um campo possui uma precipitação efetiva (mm/h) e uma eficiência (varia entre 0 e 1).
10. O solo de um campo pode ter zero ou mais de um ponto de monitoramento.

11. Um ponto de monitoramento consiste em uma bateria ou conjunto de um ou mais tensiômetros que detectam o potencial mátrico (kPa) do solo em diferentes profundidades.
12. O potencial mátrico de uma camada de solo é coletado manualmente ou automaticamente.
 - a) No caso da coleta manual, os usuários devem inserir o valor e relacioná-lo com o tempo da coleta (dia/mês/ano hora:minuto).
 - b) No caso da coleta automática, o usuário deve configurar o tipo de protocolo de comunicação (CoAP, MQTT, HTTP), endereço ou URL do servidor e endereço do gateway do campo monitorado.
 - c) O servidor do protocolo escolhido deverá inserir os registros de potencial mátrico de uma profundidade relacionado ao momento da coleta (dia/mês/ano hora:minuto:segundo).
13. Cada camada de solo monitorada possui:
 - a) profundidade de instalação do tensiômetro (mm);
 - b) largura da camada de solo monitorada z (mm)
 - c) zero ou um valor (kPa) de umidade da condição de campo cadastrado por um usuário especialista.
 - d) uma técnica de remoção de *outliers* aplicada aos potencial mátrico da camada monitorada: (1) Zscore, (2) GESD, (3) MZscore, (4) Chauvenet.
 - e) uma umidade da capacidade de campo (θ_{cc} , Seção A.2).
 - f) um valor para as constantes usadas pelo modelo de Van Genutchen (Equação A.1) de estimação de umidade do solo:
 - i. conteúdo de água residual θ_r
 - ii. conteúdo de água de saturação θ_s
 - iii. sucção da entrada de ar alpha α
 - iv. distribuição do tamanho dos poros n
 - g) um método de coleta de dados do solo: (1) nenhum, (2) HTTP, (3) CoAP, (4) MQTT, (5) texto.
14. A camada de solo mais superficial possui zero ou uma umidade da condição crítica do solo relacionada à fase de desenvolvimento do cultivo.
15. As aplicações de manejos da irrigação e predição de umidade de umidade do solo podem consultar o armazenamento de registros de dados climáticos e do solo em intervalos de tempo (dia, hora, minuto) configurados pelos usuários.
16. Os usuários poderão inserir, remover e atualizar limites mínimos e máximos para as

variáveis climáticas e do solo.

- a) Dados meteorológicos: temperatura máxima do ar ($^{\circ}\text{C}$), temperatura mínima do ar ($^{\circ}\text{C}$), umidade relativa do ar máxima (%), umidade relativa do ar mínima (%), saldo de radiação solar ($\text{MJm}^{-2}\text{d}^{-1}$), velocidade do vento (ms^{-1}), pressão atmosférica (kPa), precipitação pluviométrica (mmd^{-1}).
- b) Dados do solo: potencial mátrico coletado por um tipo de sensor (analógico ou digital)

C.2 Diagrama de Entidade e Relacionamento

De acordo com o as regras de negócio do framework, modelamos o banco de dados como mostra o diagrama de entidade e relacionamento disponível em: <https://tinyurl.com/y94sk58h>).

C.3 Dicionário de dados

Disponível em: <https://tinyurl.com/vy7l3da>