



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

ANDREI BOSCO BEZERRA TORRES

FUSÃO DE DADOS MULTINÍVEL PARA SISTEMAS DE INTERNET DAS COISAS
EM AGRICULTURA INTELIGENTE

FORTALEZA

2017

ANDREI BOSCO BEZERRA TORRES

FUSÃO DE DADOS MULTINÍVEL PARA SISTEMAS DE INTERNET DAS COISAS EM
AGRICULTURA INTELIGENTE

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do 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 mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. José Neuman de Souza

Co-Orientadora: Prof.^a Dr.^a Atslands Rego da Rocha

FORTALEZA

2017

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)

- T643f Torres, Andrei Bosco Bezerra.
Fusão de Dados Multinível para Sistemas de Internet das Coisas em Agricultura Inteligente / Andrei Bosco Bezerra Torres. – 2017.
71 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2017.
Orientação: Prof. Dr. José Neuman de Souza.
Coorientação: Profa. Dra. Atslands Rego da Rocha.
1. Internte das Coisas. 2. Fusão de Dados. 3. Fusão Multinível. 4. Redes de Sensores Sem Fio. 5. Agricultura Inteligente. I. Título.

CDD 621.38

ANDREI BOSCO BEZERRA TORRES

FUSÃO DE DADOS MULTINÍVEL PARA SISTEMAS DE INTERNET DAS COISAS EM
AGRICULTURA INTELIGENTE

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do 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 mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 11 de Julho de 2017

BANCA EXAMINADORA

Prof. Dr. José Neuman de Souza (Orientador)
Universidade Federal do Ceará (UFC)

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

Prof. Dr. Giovanni Cordeiro Barroso
Universidade Federal do Ceará (UFC)

Prof. Dr. Kelvin Lopes Dias
Universidade Federal de Pernambuco (UFPE)

Aos meus pais, que me deram carinho, atenção e a oportunidade de ser quem eu sou. E à minha esposa, minha branquinha, minha vida, que trouxe cor e emoção ao meu mundo cinza e calculista. Joelma, nossa jornada apenas começou.

AGRADECIMENTOS

À minha família, em particular meus pais Lúcia e Bosco, que sempre me incentivaram a seguir meu próprio caminho.

À minha esposa Joelma, onde sua alegria e agitação trouxeram luz e equilíbrio para minha vida. Te amo minha branquinha!

Ao meu irmão João Neto, por ser meu amigo e companheiro, apesar de nossos gênios e estilos serem tão opostos.

Ao meu orientador, professor Dr. José Neuman, e minha coorientadora, professora Dr.^a Atslands Rocha, por seus conselhos, paciência e atenção, principalmente por terem me recebido de bom grado, apesar de eu ter uma formação tão eclética.

Ao pesquisador Dr. Rubens Gondim, pelo seu apoio nesta pesquisa, que foi primordial, e por estar sempre disponível seja para conversar ou para acompanhar nos experimentos em campo.

À professora Ticiania Linhares, por conseguir encontrar tempo em seu calendário tão apertado para me guiar no mundo de análises estatísticas e do WEKA.

Ao colega José Adriano, pelo seu auxílio na criação do hardware dos protótipos, e por sua pronta disponibilidade sempre que solicitado.

E à Embrapa por ter disponibilizado material, espaço e pessoal, viabilizando a implementação e a realização dos experimentos desta pesquisa.

“I must not fear. Fear is the mind-killer. Fear is the little-death that brings total obliteration. I will face my fear. I will permit it to pass over me and through me. And when it has gone past I will turn the inner eye to see its path. Where the fear has gone there will be nothing. Only I will remain.”

(Frank Herbert)

RESUMO

A utilização de Redes de Sensores Sem Fio para detecção de fenômenos e monitoramento de ambientes não é um conceito novo, com estudos iniciados na década de 1980, mas ele tem ganhado força pela expansão da Internet das Coisas. A Internet das Coisas trata de capacitar os objetos ao nosso redor para sensoriar, identificar e analisar o mundo. Para torná-la viável em larga escala, é interessante que os objetos/sensores sejam de baixo custo, e isso implica uma série de limitações: bateria limitada, baixa capacidade de processamento e armazenamento, baixa acurácia, dentre outros. Nesse contexto, técnicas de fusão de dados podem ser utilizadas para mitigar algumas das limitações citadas e viabilizar a adoção de sensores de baixo custo. A proposta desta dissertação é uma arquitetura de fusão de dados multinível para Internet das Coisas chamada Hydra, que visa melhorar a acurácia dos sensores, identificar eventos e possibilitar tomadas de decisões baseadas em regras definidas pelo usuário. A arquitetura Hydra é dividida em três níveis: baixo nível, com foco na fusão de dados dos sensores, incluindo a identificação e remoção de *outliers*; médio nível, que lida com eventos e decisões com base nos dados sensorizados e em regras definidas pelo usuário; e alto nível, que trata da fusão de decisões com base em múltiplas aplicações. Em parceria com a Embrapa, foram desenvolvidas duas aplicações no âmbito da Agricultura Inteligente com foco no monitoramento de cultivos experimentais de caju anão-precoce e de coqueiros. A primeira aplicação visou determinar a necessidade de irrigação com base na umidade do solo e a segunda estimou a evapotranspiração para determinar o tempo adequado de irrigação. Com base nos dados sensorizados e resultantes das aplicações, foram realizados três experimentos, em que o primeiro focou na avaliação de métodos de detecção e remoção de *outliers*, o segundo focou na análise dos dados resultantes das aplicações e o terceiro explorou a utilização de métodos de aprendizado de máquina para criar um novo modelo de evapotranspiração utilizando como base o método Hargreaves-Samani e os dados sensorizados. Os resultados mostraram que o melhor método de identificação e remoção de *outliers* foi uma combinação do método ESD (*Extreme Studentized Deviate*) e do Filtro de Kalman ponderado resiliente a *outliers* (WRKF), com o qual as aplicações desenvolvidas obtiveram um bom resultado no monitoramento dos cultivos e nas decisões obtidas, além do método de aprendizado de máquina SVM (*Support Machine Vector*) quadrático ter sido capaz de gerar um modelo de evapotranspiração com valores próximos ao modelo de referência de evapotranspiração (Penman-Monteith).

Palavras-chave: Internet das Coisas. Fusão de Dados. Fusão Multinível. Rede de Sensores sem Fio. Agricultura Inteligente.

ABSTRACT

The usage of Wireless Sensor Networks (WSN) to detect and monitor phenomena isn't a new concept, with studies dating back to 1980, but it has gained momentum with the expansion of Internet of Things (IoT). Internet of Things aims to enable day to day objects to sense, identify and analyze our world. For IoT to be viable, it is recommended for the objects/sensors to be low-cost, and that implies a series of limitations: low battery, low processing and storage capabilities, low accuracy, etc. In this context, data fusion techniques can be used to mitigate some of these limitations and make the adoption of low-cost sensors viable. This dissertation proposes a multilevel data fusion architecture for IoT called Hydra, that aims to improve sensor accuracy, identify events and make decisions based on user-defined rules. The Hydra architecture is composed of three level: low level, focused in sensor data fusion, including outlier identification and removal; medium level, dealing with events and decision making based on sensed data and user defined rules; and high level, that handles decision fusion based on multiple applications. In partnership with Embrapa, two applications were developed in the scope of Intelligent Agriculture focusing on the monitoring of experimental cultures of precocious-dwarf cashew and coconut trees. The first application aimed to determine the need for irrigation based on soil moisture and the second estimated the crop evapotranspiration to determine the adequate irrigation time. Three experiments were performed based on data provided by the sensors and by the applications, in which the first focused on the evaluation of methods of detection and removal of outliers, the second focused on the analysis of the data resulting from the applications and the third explored the use of machine learning to create a new evapotranspiration model based on the Hargreaves-Samani method and the sensed data. Results showed that the best method to identify and remove outliers was a combination of the ESD method (Extreme Studentized Deviate) and WRKF filter (weighted outlier-robust Kalman filter), with which the developed applications obtained a good result in monitoring the crops and in the decisions obtained. And also the SVM (Support Machine Vector) quadratic machine learning model was able to generate an evapotranspiration model that resulted in values close to the evapotranspiration reference model (Penman-Monteith).

Keywords: Internet of Things. Data Fusion. Multilevel Fusion. Wireless Sensor Network. Intelligent Agriculture.

LISTA DE ILUSTRAÇÕES

Figura 1 – Relacionamento entre os termos ligados à fusão de dados (NAKAMURA <i>et al.</i> , 2007) [tradução do autor].	22
Figura 2 – Níveis de fusão com base no relacionamento entre as fontes de dados (ELMENREICH, 2002 apud NAKAMURA <i>et al.</i> , 2007) [tradução do autor]. . .	23
Figura 3 – Níveis de fusão com base na abstração da entrada e saída dos dados (DASARATHY, 1997).	24
Figura 4 – Visão geral da arquitetura de fusão multinível	29
Figura 5 – Detalhes das camadas da arquitetura Hydra	30
Figura 6 – Exemplo de decisão composta da camada de alto nível	31
Figura 7 – Níveis da aplicação de umidade de solo	34
Figura 8 – Níveis da aplicação de evapotranspiração	39
Figura 9 – Nó sensor tipo A (sem módulo de rádio e sensores)	43
Figura 10 – Nó sensor tipo B (sem módulo de rádio e sensores	44
Figura 11 – Nó sorvedouro	45
Figura 12 – Croqui de trecho do campo experimental de caju anão-precoce	46
Figura 13 – Croqui de trecho do campo experimental de coqueiros	46
Figura 14 – Exemplo de estrutura de mensagem	48
Figura 15 – Outliers e dado fusionado (15cm de profundidade): Critério de Perice e Z-Score Modificado	52
Figura 16 – Outliers e dados fusionados (45cm de profundidade): Boxplot Ajustado e ESD	52
Figura 17 – Outliers e dados fusionados (45cm de profundidade): Critério de Peirce e Z-Score Modificado	52
Figura 18 – Dados filtrados dos sensores a 15cm de profundidade (a)	54
Figura 19 – Dados filtrados dos sensores a 15cm de profundidade (b)	54
Figura 20 – Dados filtrados dos sensores a 45cm de profundidade (a)	54
Figura 21 – Dados filtrados dos sensores a 45cm de profundidade (b)	54
Figura 22 – Dados filtrados dos sensores a 75cm de profundidade (a)	55
Figura 23 – Dados filtrados dos sensores a 75cm de profundidade (b)	55
Figura 24 – Resultados da decisão de início de irrigação	56
Figura 25 – Resultados da decisão de tempo de irrigação	57
Figura 26 – WEKA: árvore de decisão gerada pelo algoritmo M5P	59

Figura 27 – WEKA: visualização do ETo previsto (y) e o ETo real (x)	61
Figura 28 – MATLAB: gráfico de valores reais e estimados	62
Figura 29 – MATLAB: gráfico de precisão da predição	62
Figura 30 – Comparação entre os métodos de cálculo de ETo	63

LISTA DE TABELAS

Tabela 1 – Trabalhos relacionados (sumário)	28
Tabela 2 – Consumo energético dos nós sensores	44
Tabela 3 – Critérios de remoção de <i>outliers</i>	47
Tabela 4 – Coleta 01: lista completa de atributos	49
Tabela 5 – Coleta 02: Lista completa de atributos	50
Tabela 6 – Total de <i>outliers</i> detectados	53
Tabela 7 – WEKA: regras criadas pelo método M5P	60
Tabela 8 – Resultados dos modelos de regressão linear do MATLAB	63

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application programming interface</i> - Interface de Programação de Aplicativos
DAI-DAO	<i>Data Input - Data Output</i>
DAI-FEO	<i>Data Input - Feature Output</i>
DEI-DEO	<i>Decision Input - Decision Output</i>
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
ESD	<i>Generalized Extreme Studentized Deviate</i>
ETo	Evapotranspiração de referência
FEI-DEO	<i>Feature Input - Decision Output</i>
FEI-FEO	<i>Feature Input - Feature Output</i>
INMET	Instituto Nacional de Meteorologia
IoT	<i>Internet Of Things</i> - Internet das Coisas
kPa	Quilopascal
MAE	<i>Mean Absolute Error</i> - Erro absoluto médio
MQTT	<i>Message Queuing Telemetry Transport</i>
MSE	<i>Mean Square Error</i> - Erro quadrático médio
OSUS	<i>Open Standard for Unattended Sensors</i> - Padrão Aberto de Sensores Não Acompanhados
R ²	Coefficiente de determinação
RMSE	<i>Root Mean Square Error</i> - Raiz do erro quadrático médio
RSSF	Rede de Sensores Sem Fio
SoC	<i>Systems on a Chip</i> - Sistemas em um chip
SVM	<i>Support Machine Vector</i> - Máquina de Suporte de Vetores
WRKF	<i>Weighted outlier-robust Kalman filter</i> - Filtro de Kalman ponderado robusto a outliers

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Motivação	17
1.2	Hipótese e Objetivo Central	18
1.3	Estrutura da Dissertação	18
2	REVISÃO DA LITERATURA	19
2.1	Contextualização	19
2.1.1	<i>Internet das Coisas</i>	19
2.1.2	<i>Fusão de Dados</i>	21
2.1.2.1	<i>Classificação</i>	22
2.1.3	<i>Agricultura Inteligente</i>	23
2.2	Trabalhos Relacionados	25
2.2.1	<i>Metodologia da Revisão Sistemática</i>	25
2.2.2	<i>Trabalhos Selecionados</i>	26
2.3	Conclusão do Capítulo	27
3	ARQUITETURA DE FUSÃO MULTINÍVEL	29
3.1	Visão geral	29
3.2	Conclusão do Capítulo	32
4	INSTANCIAMENTO DA PROPOSTA EM AGRICULTURA INTELIGENTE	33
4.1	Aplicações	33
4.1.1	<i>Umidade do solo</i>	33
4.1.1.1	<i>Conversão dos dados dos sensores de tensão de água no solo</i>	33
4.1.1.2	<i>Filtragem</i>	35
4.1.1.3	<i>Fusão dos sensores de umidade</i>	36
4.1.1.4	<i>Decisão sobre irrigação</i>	37
4.1.2	<i>Evapotranspiração</i>	37
4.1.2.1	<i>Dados da estação meteorológica</i>	38
4.1.2.2	<i>Cálculo de evapotranspiração</i>	38
4.1.2.3	<i>Decisão sobre tempo de irrigação</i>	40
4.1.3	<i>Decisão de alto nível</i>	41
4.1.4	<i>Gerência e interface com usuário</i>	42

4.2	Conclusão do Capítulo	42
5	EXPERIMENTOS E ANÁLISE DE RESULTADOS	43
5.1	Material	43
5.1.1	<i>Ambiente</i>	43
5.1.2	<i>Protótipo de hardware</i>	43
5.1.3	<i>Software</i>	46
5.1.3.1	<i>Filtragem: camada de nível baixo</i>	47
5.1.3.2	<i>Filtragem: camada de nível médio</i>	47
5.1.3.3	<i>Interface com usuário: camada de gerência</i>	48
5.1.4	<i>Comunicação</i>	48
5.2	Planejamento dos Experimentos	49
5.2.1	<i>Coleta de dados</i>	49
5.2.1.1	<i>Primeira Coleta</i>	49
5.2.1.2	<i>Segunda coleta</i>	50
5.3	Discussão do Experimento I: detecção e remoção de outliers	51
5.3.1	<i>Algoritmos de identificação de outliers</i>	51
5.3.2	<i>Filtragem de sinal</i>	53
5.3.3	<i>Conclusão do Experimento I</i>	55
5.4	Discussão do Experimento II: aplicações de evapotranspiração e umidade do solo	56
5.4.1	<i>Aplicação de Umidade de Solo</i>	56
5.4.2	<i>Aplicação de Evapotranspiração</i>	56
5.4.3	<i>Validação dos Dados</i>	57
5.4.4	<i>Conclusão do Experimento II</i>	61
5.5	Discussão do Experimento III: novo modelo de evapotranspiração	61
5.5.1	<i>Conclusão do Experimento III</i>	63
5.6	Conclusão do Capítulo	64
6	CONCLUSÕES E TRABALHOS FUTUROS	65
6.1	Artigos Publicados	66
6.2	Capítulo de Livro Publicado	66
6.3	Dificuldades Encontradas	66
6.4	Trabalhos Futuros	67

REFERÊNCIAS 68

1 INTRODUÇÃO

Este projeto de mestrado propõe uma Arquitetura de Fusão de Dados para sistemas de Internet das Coisas para Agricultura Inteligente, almejando viabilizar a utilização de sensores de baixo custo, detecção de eventos e automatização de decisões. Este capítulo está dividido da seguinte forma: a Seção 2.1 explica o contexto em que esta pesquisa foi realizada, abordando conceitos sobre Internet das Coisas, fusão de dados e agricultura inteligente. A Seção 1.1 trata da motivação para o desenvolvimento desta pesquisa. A Seção 1.2 apresenta a hipótese inicial e as questões de pesquisa que guiaram o desenvolvimento das atividades. E na Seção 1.3 é apresentada a estrutura dos capítulos desta dissertação.

1.1 Motivação

O agronegócio tem grande importância na economia brasileira, representando quase 23% do Produto Interno Bruto (MOLIN, 2017) e, com ajuda da tecnologia, foi possível aumentar a produtividade de 1.990 quilos por hectare para 3.900kg no últimos 30 anos (G1 Economia, 2017). Porém, a utilização de equipamentos para monitoramento de cultivo são caros, de difícil acesso, implementação e manutenção para pequenos e médios produtores. Além disso, produtores de estados como o Ceará convivem com longos períodos de estiagem, dificultando e encarecendo ainda mais o processo da produção agrícola.

O Nordeste está entrando no sexto ano de estiagem, e a previsão é de que os reservatórios não tenham uma recuperação hídrica significativa, resultando em um grande risco de esgotamento de diversas represas e açudes (ANDRADE, 2017; PINA, 2016). Segundo Braga (2017), doze cidades só possuem reserva suficiente para chegar ao final do mês de julho deste ano, enquanto os reservatórios de outros vinte e quatro municípios devem atingir o volume morto em outubro. Torna-se essencial utilizar os recursos hídricos de maneira planejada e controlada.

Para resolver a questão do controle e monitoramento agrícola, a *Internet Of Things* - Internet das Coisas (IoT) pode ser uma abordagem viável. Por exemplo, uma rede de sensores e atuadores sem fio pode ser utilizada para monitorar cultivos (e determinar o momento de irrigação, por exemplo), enquanto objetos inteligentes podem informar as condições de armazenamento da colheita. Tais informações podem ser fusionadas para obter uma visão geral e tomar decisões relativas ao cultivo.

Nesse contexto, a adoção da agricultura inteligente (também conhecida como agri-

cultura de precisão), em conjunto com um sistema IoT, permitiria um melhor gerenciamento do cultivo, resultando em retornos econômicos e melhoria na sustentabilidade do ambiente.

1.2 Hipótese e Objetivo Central

Partimos da hipótese de que o uso de uma arquitetura de fusão de dados multinível pode permitir o uso de sensores de baixo custo em agricultura inteligente, propiciando acurácia, detecção de eventos e tomada de decisão. Sendo assim, com base nessa hipótese, o objetivo central desta dissertação é propor tal arquitetura e demonstrar sua aplicação no domínio da agricultura inteligente, possibilitando tomadas de decisão em tempo real com base em dados providos por um sistema de Internet das Coisas, com infraestrutura de Rede de Sensores Sem Fio (RSSF).

1.3 Estrutura da Dissertação

Os capítulos restantes desta dissertação estão organizadas da seguinte maneira:

- **Capítulo 2 - Revisão da Literatura:** apresenta uma revisão da literatura sobre os conceitos principais abordados nesta dissertação, e uma revisão sistemática de trabalhos similares que abordam arquiteturas de fusão de dados para uso com IoT e RSSF.
- **Capítulo 3 - Arquitetura de Fusão Multinível:** aborda a proposta desta dissertação: a arquitetura de fusão de dados multinível Hydra.
- **Capítulo 4 - Instanciação da proposta em agricultura inteligente:** apresenta a instanciação da arquitetura Hydra no domínio de agricultura inteligente e as aplicações desenvolvidas.
- **Capítulo 5 - Experimentos e Análise de Resultados:** detalha o design experimental realizado, com os procedimentos seguidos, descrição do ambiente, implementação, coleta de dados e apresenta a análise e validação da Hydra.
- **Capítulo 6 - Conclusões e Trabalhos Futuros:** traz a conclusão desta dissertação, objetivos alcançados, artigos publicados, contribuições e trabalhos futuros.

2 REVISÃO DA LITERATURA

Neste capítulo são apresentadas uma breve explicação sobre os conceitos principais abordados nesta dissertação e uma revisão sistemática de trabalhos similares que abordam arquiteturas de fusão de dados para IoT.

2.1 Contextualização

2.1.1 *Internet das Coisas*

Uma Rede de Sensores Sem Fio (RSSF) consiste de uma rede formada por nós, que, cooperativamente, são capazes de detectar fenômenos físicos do ambiente, tais como luz, pressão ou umidade (YINBIAO *et al.*, 2014). Os estudos neste tema foram iniciados em meados da década de 1980, mas só ganharam força no início do século XXI, com os avanços nas tecnologias de miniaturização, de rádio e de processadores, permitindo a criação de *Systems on a Chip* - Sistemas em um chip (SoC) baratos e eficientes.

Paralelamente ao desenvolvimento da RSSF, em 1991 o termo IoT foi cunhado por Ashton (2009) em uma apresentação do laboratório *Auto-ID Center* do Instituto de Tecnologia de Massachusetts (MIT) para a *Procter and Gamble (P&G)*, que se refere a dar aos computadores a capacidade de obter informações independentemente dos humanos, ou seja, a utilização de tecnologias de sensoriamento e de comunicação para observar, identificar e analisar o mundo. Com esse conceito, fica claro que as RSSF têm um papel crucial na IoT, funcionando como uma ponte entre o mundo real e o digital por meio das informações do ambiente e dos objetos (ATZORI *et al.*, 2010). Dentre as potenciais áreas para utilização de IoT usando RSSF como infraestrutura, pode-se citar (ATZORI *et al.*, 2010; SINGH *et al.*, 2014):

- agricultura;
- casas/edifícios inteligentes;
- cidades inteligentes;
- assistência médica (*e-health*);
- gerenciamento inteligente de produtos/inventários;
- segurança e vigilância;
- militar.

Através das tecnologias da IoT, tornou-se viável utilizar um grande número de

sensores de baixo custo pela fração do valor de um sensor especializado de alto custo (SANCHEZ *et al.*, 2013). Apesar dos sensores especializados terem um alto grau de acurácia, seus custos iniciais e de manutenção podem inviabilizar uma aplicação em larga escala. Já os sensores de baixo custo não possuem o mesmo grau de acurácia, mas, através de técnicas de fusão de dados, a utilização de uma grande quantidade de sensores pode possibilitar a obtenção de medições mais acuradas. Porém, para que sejam de baixo custo, os nós que compõem as RSSF, usadas como infraestrutura da IoT, possuem várias limitações que precisam ser superadas, dentre elas: (ABDELGAWAD; BAYOUMI, 2012; RAWAT *et al.*, 2014)

1. recursos limitados;
2. vida útil limitada (normalmente utilizam baterias ou pilhas como fonte de energia);
3. transmissão de dados (transmissões em excesso resultam em alto consumo de energia e possíveis problemas de colisão);
4. detecção de falhas/imprecisão de sensores.

Os protocolos para a comunicação entre os componentes da IoT devem lidar com as limitações acima citadas, e alguns protocolos foram criados exatamente para lidar com tais fatores: CoAP (SHELBY *et al.*, 2014), MQTT (OASIS STANDARD, 2015), WAMP (OBERSTEIN; GOEDDE, 2017), dentre outros. Nesta pesquisa, o protocolo MQTT foi adotado para a realização dos experimentos. O protocolo *Message Queuing Telemetry Transport* (MQTT) tem sido largamente adotado por diversas empresas (Amazon Web Services, 2016; Microsoft, 2017; IBM, 2017), seja com foco no uso na IoT ou apenas como um protocolo de comunicação. Um exemplo fora do contexto de IoT é a adoção do MQTT pelo Facebook como protocolo de comunicação de seu sistema de *instant messaging* (ZHANG, 2011). E dentro do contexto de IoT, a Amazon e a IBM adotaram MQTT como protocolo padrão em suas plataformas IoT (AWS IoT e IBM Watson IoT, respectivamente). Segundo Skerrett (2015), diretor de marketing da fundação Eclipse, “me parece que o MQTT se tornou o padrão a ser suportado por qualquer provedor sério de soluções para IoT”¹.

O presente trabalho foca nos problemas 3 e 4 acima citados que, por meio da arquitetura de fusão multinível proposta, objetiva reduzir a quantidade de dados a ser transmitida, aumentar o grau de confiabilidade dos dados, detectar *outliers* e permitir a tomada de decisão com base nos dados coletados. Para reduzir o consumo de energia dos nós sensores, também foram realizadas modificações no *hardware* (remoção de componentes) para viabilizar o uso

¹ Tradução livre do autor.

dos nós por maiores períodos de tempo. Como estudo de caso, trabalhamos no ambiente de Agricultura Inteligente com a implementação de uma estrutura de sensores heterogêneos de baixo custo para monitoramento de cultivos experimentais.

2.1.2 Fusão de Dados

O termo **fusão de dados** , apesar de amplamente utilizado, não possui um significado definitivo, variando de acordo com o contexto em que é aplicado. Segundo Nakamura *et al.* (2007), diversos termos (fusão de dados, fusão de sensores e fusão de informação) são utilizados para descrever algum aspecto de fusão. Ainda assim, os termos fusão de dados e fusão de informação são aceitos e utilizados de maneira intercambiável. Tem-se a seguir algumas definições formais:

U.S. DEPARTMENT OF DEFENSE (1991): “processo multinível e multifacetado que lida com automatização da detecção, associação, correlação, estimativa e combinação de dados e informações de múltiplas fontes.”

Waltz e Llinas (1990): “a combinação de dados de múltiplos sensores e de informação relacionada provida por bancos de dados associados, para alcançar uma maior acurácia e inferências específicas do que poderia ser obtida ao utilizar um único sensor”

Wald (1999) “fusão de dados é um *framework* formal no qual são expressos meios e ferramentas para a união de dados originados de múltiplas fontes. Almeja obter informações de maior qualidade; a definição exata de ‘maior qualidade’ vai depender da aplicação.”

Boström *et al.* (2007): “o estudo de métodos eficientes para transformar de forma automática ou semi-automática informações de diferentes fontes e diferentes pontos no tempo em uma representação que forneça suporte eficaz para tomada de decisão humana ou automatizada”.

A última definição, proposta por Boström *et al.* (2007), foi criada após um estudo dentre as diversas definições existentes, seus contextos e focos. Seu objetivo foi remover ambiguidades, como o uso de termos vagos como “qualidade”, e ser o mais claro e objetivo possível. Para esta pesquisa, essa será a definição adotada.

Também é importante deixar claro a distinção entre fusão de dados, fusão de sensores e agregação de dados. Na Figura 1 é apresentado o relacionamento entres os termos. Fusão de dados e fusão de informação são termos intercambiáveis, que lidam com a fusão de dados provenientes de qualquer fonte, inclusive de sensores. Agregação de dados tem como foco a redução de dados de qualquer tipo de fonte, inclusive de sensores. Fusão de sensores (ou

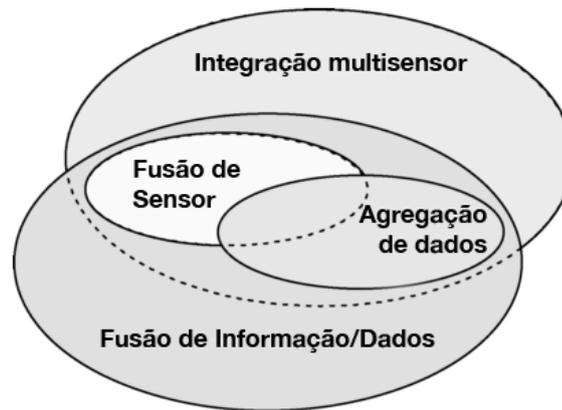


Figura 1 – Relacionamento entre os termos ligados à fusão de dados (NAKAMURA *et al.*, 2007) [tradução do autor].

fusão multissensorial) é um subgrupo que lida com a união de fontes sensoriais. Já a integração multissensorial se aplica à utilização de dados sensoriais e de fontes associadas (como bancos de dados) para realizar inferências e interagir com o ambiente. Portanto, fusão de sensores é contida na interseção entre integração multissensorial e fusão de dados.

2.1.2.1 Classificação

Fusão de dados pode ser classificada com base em diversos aspectos: relacionamento entre as fontes de dados, níveis de abstração dos dados, níveis de abstração da entrada e saída dos dados, dentre outros.

Na fusão com base no relacionamento entre as fontes de dados não existe dependência direta entre os sensores. Esse relacionamento pode ser dividido em: **complementar**, em que os sensores fornecem informações de trechos diferentes de um mesmo cenário; **redundante**, em que os sensores fornecem o mesmo tipo de informação de um mesmo cenário; e **cooperativo**, em que as informações de sensores diferentes podem ser unidas para gerar uma nova informação (veja Figura 2).

O conceito de fusão com base nos níveis de abstração dos dados, apresentado por Iyengar *et al.* (2001) e complementado por Nakamura *et al.* (2007), é dividido em quatro camadas de abstração: Baixo Nível, Médio Nível, Alto Nível e Multinível. A camada de baixo nível (também chamada de camada de dados ou de sinal) trata dos dados puros, que podem ser combinados para gerar valores mais precisos. A camada de médio nível (ou camada de características), representa as características/atributos de uma entidade, que podem ser fusionados para obter outras características. A camada de alto nível (ou camada de decisão) representa a

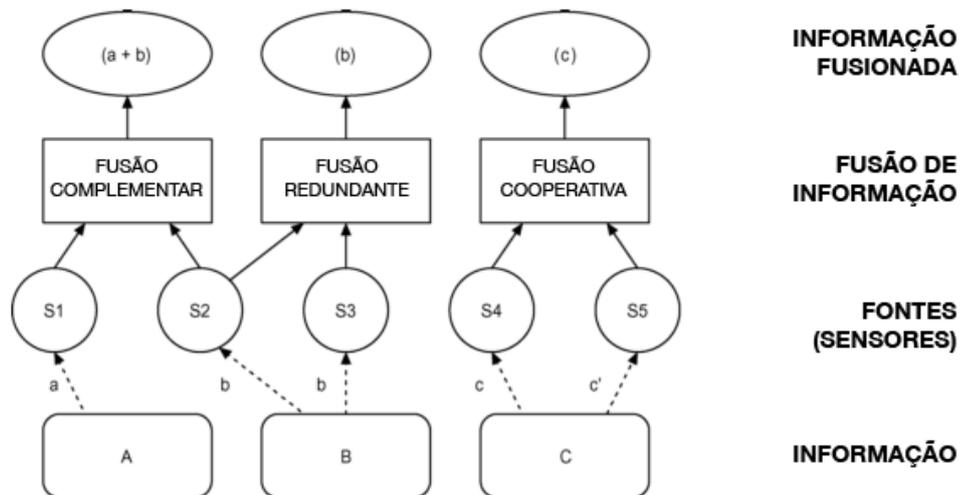


Figura 2 – Níveis de fusão com base no relacionamento entre as fontes de dados (ELMENREICH, 2002 apud NAKAMURA *et al.*, 2007) [tradução do autor].

união de decisões para a tomada de novas decisões com maior grau de confiança. Já a camada multinível representa a possibilidade de unir dados de diferentes níveis de abstração, como unir dados de uma medição com uma característica para gerar uma decisão.

A classificação com base nos níveis de abstração da entrada e saída de dados (Figura 3) pode ser considerada uma extensão da classificação com base nos níveis de abstração dos dados, mas com maior granularidade. Ela expande os conceitos de dados, característica e decisão (equivalentes aos níveis baixo, médio e alto) em cinco categorias (DASARATHY, 1997):

- *Data Input - Data Output (DAI-DAO)*: A fusão deste nível lida com dados puros como entrada e saída, possivelmente resultando em valores mais confiáveis e acurados.
- *Data Input - Feature Output (DAI-FEO)*: A fusão deste nível utiliza dados puros para apreender características ou atributos de uma entidade.
- *Feature Input - Feature Output (FEI-FEO)*: A fusão com entrada e saída de características permite o refinamento de uma característica ou a criação de uma nova.
- *Feature Input - Decision Output (FEI-DEO)*: Este nível de fusão utiliza um grupo de características para gerar uma decisão.
- *Decision Input - Decision Output (DEI-DEO)*: E neste nível é realizada a fusão de decisões para aumentar a acurácia de uma decisão ou gerar uma nova.

2.1.3 Agricultura Inteligente

A Agricultura Inteligente (*Smart Agriculture* ou *Smart Farming*), também conhecida como Agricultura de Precisão (*Precision Agriculture*), surgiu em meados dos anos 1980 com

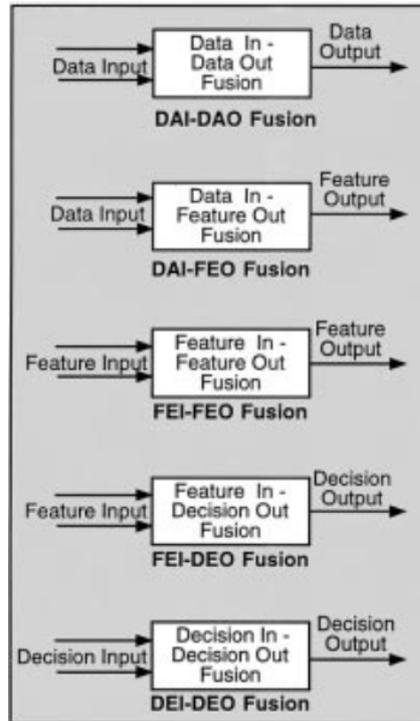


Figura 3 – Níveis de fusão com base na abstração da entrada e saída dos dados (DASARATHY, 1997).

o objetivo de fazer a coisa certa, na quantidade certa e no tempo certo, utilizando práticas que otimizem a eficiência e reduzam o impacto ambiental, tendo como base tecnologias de monitoramento e análise para auxiliar e embasar as decisões (GEBBERS; ADAMCHUK, 2010; LEONARD, 2016). O Ministério da Agricultura, Pecuária e Abastecimento (Mapa) definiu Agricultura de Precisão como “um sistema de gerenciamento agrícola baseada na variação espacial e temporal da unidade produtiva e visa ao aumento de retorno econômico, à sustentabilidade e à minimização do efeito ao ambiente” (BRASIL, 2012 apud BERNARDI *et al.*, 2014). Como os termos são intercambiáveis, a expressão Agricultura Inteligente será utilizada a partir deste ponto.

A Empresa Brasileira de Pesquisa Agropecuária (Embrapa) indica o uso da tecnologia como o caminho para tornar o crescimento da agricultura economicamente viável e sustentável à frente dos desafios enfrentados pelo Brasil: uso racional da água, necessidade de aumento da produção, mudanças climáticas, eficiência da produção, etc (BERNARDI *et al.*, 2014).

É estimado que em meados de 2050, a produção de alimentos deve aumentar em 70% para suportar a demanda da população, que deve alcançar 9,4 bilhões de pessoas (MELO; TREINISH, 2012). Nesse contexto, a aplicação de tecnologias de IoT auxiliam no gerenciamento e monitoramento dos cultivos, além da otimização do uso de recursos como água, pesticidas e

fertilizantes. Um dos recursos que permitem um uso mais efetivo da água é o condicionador de solo (ex.: hidrogel), que pode melhorar a capacidade de retenção de umidade pelo solo, porém é necessário, no mínimo, um monitoramento contínuo do solo para garantir sua eficácia.

2.2 Trabalhos Relacionados

2.2.1 Metodologia da Revisão Sistemática

Foi realizada uma revisão sistemática para encontrar os trabalhos mais relevantes na área de fusão de dados, que tivessem como proposta um modelo ou arquitetura de fusão multissensorial ou multinível. A sistemática utilizada teve a seguinte estrutura:

- **String de busca:**

("data fusion"OR "information fusion") AND ("multilevel"OR "multisensor")

- **Fontes de dados:** IEEE Digital Library, Science@Direct, Scopus;

- **Período:** 2014 a 2017;

- **Critérios:**

- Inclusão: arquitetura de fusão; agricultura inteligente;
- Exclusão: *survey* ou *review*; fora do escopo; editorial; documento não disponível;

- **Checklist de verificação de qualidade:** (nota de corte: 1.0)

- Questões:

- * Apresenta uma arquitetura ou modelo de fusão de dados?
- * O modelo é agnóstico quanto ao domínio?
- * A arquitetura é voltada para agricultura inteligente?
- * Utiliza a classificação de Dasarathy?

- Respostas:

- * Sim (1.0);
- * Parcialmente (0.5);
- * Não (0.0);

- **Filtragem:**

- Artigos obtidos pela *string* de busca: 433 (somente 27 duplicados);
- Artigos selecionados após aplicação dos critérios: 14;
- Artigos remanescentes após avaliação seguindo o *checklist*: 4.

Além dos dois artigos selecionados na revisão acima descrita, um artigo nacional foi

encontrado pelo *Google Scholar* e adicionado como um trabalho relacionado.

2.2.2 *Trabalhos Seleccionados*

Callegaro *et al.* (2014) “propõe uma arquitetura para a fusão de informação voltada para sensores de baixo custo em RSSF”, em que a arquitetura é dividida em três camadas: fusão local (ocorre no próprio sensor que coleta os dados), fusão de baixo nível (ocorre no sensor coordenador da rede ou posteriormente *offline*) e gerenciamento (interface com o usuário e fusões de médio e alto nível). Os autores se limitam a especificar o funcionamento apenas da camada de baixo nível, focando na análise de algoritmos de detecção de *outliers* para o aumento de acurácia dos dados.

Wichit (2014) apresenta uma arquitetura de fusão de dados multisensorial, com foco no reconhecimento de comportamento humano utilizando um algoritmo de fusão baseada em lógica difusa (*fuzzy*). Os autores não especificam uma estrutura de camadas para a arquitetura, apenas um fluxo para o reconhecimento de atividade humana e aplicação de inferência com lógica difusa.

Wang *et al.* (2015) propõem uma estrutura de fusão de dados em dois níveis baseada em estimação do estado, cujo foco é a melhoria da acurácia dos dados. No primeiro nível, são agrupados sensores que possuam a mesma estrutura de dados e é realizada a fusão de seus dados. O segundo nível recebe os dados fusionados dos grupos do primeiro nível e realiza sua fusão utilizando um algoritmo de interseção de covariância. Os autores focam no aumento da acurácia dos dados, sem abordar a interpretação dos dados, criação de regras e tomadas de decisão.

De Paola *et al.* (2016) propõem um “sistema de fusão de dados sensível ao contexto, auto-otimizado e adaptativo, baseado em uma arquitetura de três níveis”². A camada inferior é responsável pelo sensoriamento, gerando dados brutos. A camada intermediária realiza a fusão de dados, em que tenta integrar informações disponíveis no contexto. E a camada superior busca alcançar um equilíbrio entre performance do sistema e custos de execução (como consumo energético). Mas não foi especificado se alguma das camadas realiza fusão de decisões e nem como seria a interface de criação de regras e gerência. Como prova de conceito, utilizaram o cenário de detecção de atividade em um ambiente inteligente.

Bish *et al.* (2016) descrevem uma “arquitetura de fusão que combina fusão multimodal e multissensorial dentro do arcabouço do *Open Standard for Unattended Sensors* - Padrão

² Tradução do autor

Aberto de Sensores Não Acompanhados (OSUS)''². O objetivo da arquitetura dos autores é ser um sistema modular, *plug and play*, e que métodos de fusão que venham a existir no futuro sejam facilmente integrados. Essa proposta foca em como a estrutura proposta pelos autores funciona no arcabouço OSUS, onde a fusão multissensorial é uma das camadas do modelo, sem entrar em detalhes sobre o funcionamento dessa camada.

A Tabela 1 apresenta um sumário dos trabalhos citados acima e da arquitetura Hydra proposta nesta dissertação. O diferencial da arquitetura Hydra, em comparação às citadas acima, é o fato dela ser multinível, baseada na organização de abstração dos dados de entrada e saída proposta por Dasarathy (1997), e a fusão de dados ser feita não apenas para redução de dados a serem transmitidos, mas também para identificação de eventos e tomadas de decisão, além da possibilidade de fusão de decisões de múltiplas aplicações que estejam executando na mesma plataforma.

2.3 Conclusão do Capítulo

Neste capítulo foram apresentados os conceitos abordados nesta pesquisa, a metodologia de busca dos trabalhos relacionados, bem como foram abordados o foco, pontos fortes e fracos de cada um destes. Por fim, foi apresentada uma comparação entre os trabalhos selecionados e o que diferencia este projeto.

Tabela 1 – Trabalhos relacionados (sumário)

Trabalho	Estrutura	Estudo de Caso	Fusão de decisões	Camada de gerência/inter-face	Classificação adotada
Esta proposta (Hydra)	hierárquica (3 níveis)	agricultura inteligente	sim	sim	Dasarathy (1997)
Callegaro <i>et al.</i> (2014)	hierárquica (3 níveis)	agricultura inteligente	não	sim	Dasarathy (1997)
Wichit (2014)	não-hierárquica	reconhecimento de atividade	não	não	não especificam
Wang <i>et al.</i> (2015)	hierárquica (2 níveis)	vigilância	não	não	não especificam
De Paola <i>et al.</i> (2016)	hierárquica (3 níveis)	ambiente inteligente	não	não	não especificam
Bish <i>et al.</i> (2016)	não-hierárquica	–	não	não	não especificam

3 ARQUITETURA DE FUSÃO MULTINÍVEL

3.1 Visão geral

Neste trabalho é proposta a arquitetura Hydra, uma arquitetura de fusão de dados multinível que usa como base as classificações apresentadas por Dasarathy (1997) e Iyengar *et al.* (2001), sendo composta por três camadas de fusão (baixo nível, médio nível e alto nível) e uma camada de gerência (Figura 4).

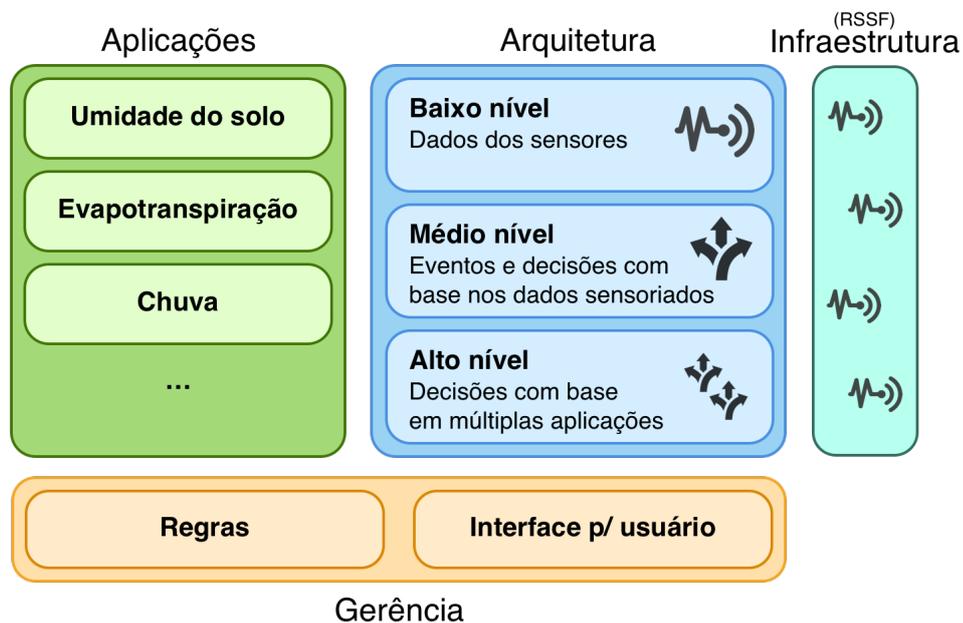


Figura 4 – Visão geral da arquitetura de fusão multinível

A camada de baixo nível (Figura 5) é responsável por receber os dados brutos dos sensores, executar uma filtragem inicial, de acordo com o escopo de funcionamento nominal de cada sensor, e gerar dados filtrados. Nela também é possível realizar a fusão de dados brutos para obter uma *feature* (característica, ou dado interpretado). Eis alguns exemplos que ilustram a camada de baixo nível no domínio de agricultura inteligente:

- **Fusão DAI-DAO:** Dados brutos de sensores resultam em um novo dado bruto. Ex.: múltiplos sensores de temperatura geram números inteiros de 12 bits, que são fusionados e o valor pode ser convertido em Celsius ou Fahrenheit;
- **Fusão DAI-FEO:** Dados interpretados são obtidos a partir de dados brutos de sensores, sejam eles iguais (fusão redundante) ou diferentes (fusão cooperativa). Ex.: sensor de tensão de água no solo gera número inteiro de 10 bits que representa a resistência do solo à corrente elétrica (Ohms), e, em conjunto com os dados do sensor de temperatura de solo

(°C), um valor interpretado é gerado (kPa).

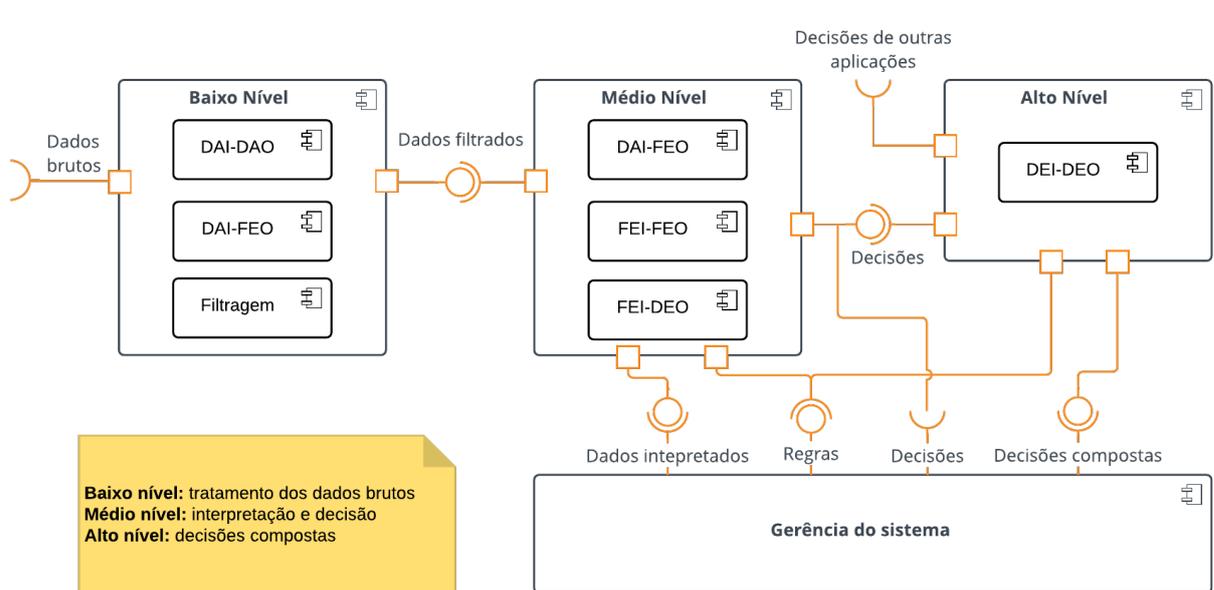


Figura 5 – Detalhes das camadas da arquitetura Hydra

Os dados resultantes da camada de baixo nível alimentam a camada de médio nível (Figura 5), que por sua vez é responsável por realizar a fusão de sensores complementares¹ e, a partir de uma base de regras definidas pelo usuário/aplicação, realizar decisões. Eis uma breve explicação de cada tipo de fusão desta camada:

- **Fusão DAI-FEO:** Dados interpretados são obtidos a partir de dados brutos filtrados de sensores. Ex.: dados de múltiplos sensores (chuva, temperatura, radiação) de uma estação meteorológica são combinados para obter um novo tipo de informação, como a evapotranspiração (ET_o).
- **Fusão FEI-FEO:** Dados interpretados são combinados para obter novos dados. Ex.: dados interpretados de múltiplos sensores são fusionados (utilizando média, mediana ou média móvel, por exemplo) para gerar um dado interpretado mais acurado;
- **Fusão FEO-DEO:** Dados interpretados são utilizados para chegar a uma decisão. Ex.: os dados interpretados fusionados de umidade de solo são analisados de acordo com parâmetros e o resultado é uma decisão se o solo precisa ser irrigado ou não. Essa decisão também pode ser mais complexa, incluindo dados de outros sensores: “caso a umidade de solo ultrapasse o limite e tenha chovido”.

Os dados e decisões gerados pela camada de médio nível são enviados para a camada de gerência e para a camada de alto nível, quando necessário (Figura 5). A camada

¹ Ver subseção 2.1.2.1 para explicação sobre fusão de sensores complementares

de alto nível trata da fusão de decisões (DEI-DEO) realizadas por múltiplas aplicações, para executar uma decisão composta mais complexa. Ex.: utilizar as decisões sobre a necessidade de irrigação do cultivo (gerado pela aplicação de umidade de solo) e sobre o tempo de irrigação (gerado pela aplicação de evapotranspiração) para decidir quais aspersores ativar e por quanto tempo (Figura 6).

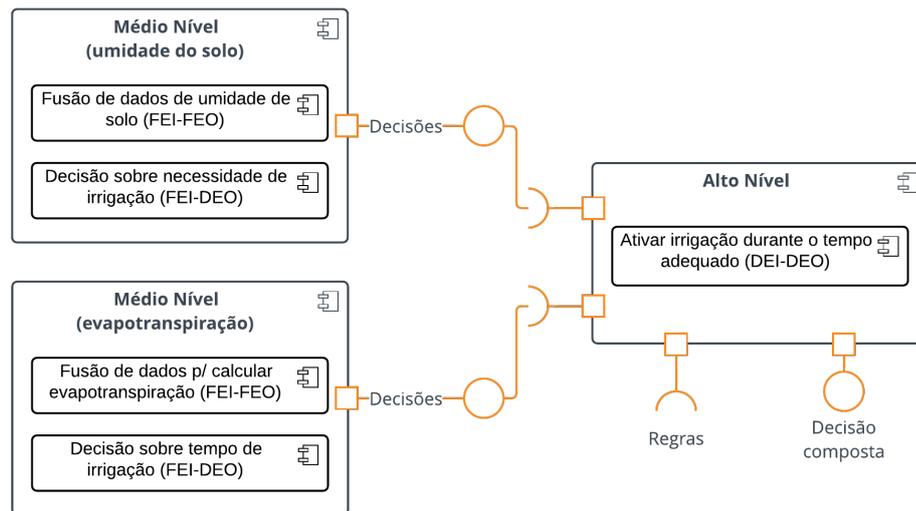


Figura 6 – Exemplo de decisão composta da camada de alto nível

E acima de todos os níveis existe a camada de gerência, que recebe dados das camadas de médio e alto nível. Ela é responsável pela interface com o usuário, e nela são definidas as regras que irão gerir as decisões das camadas inferiores, as prioridades das regras e o tempo de execução das aplicações. Por meio dessa interface, o usuário pode acompanhar os dados gerados e os resultados das decisões.

Resumindo, no caso de um cultivo com diversos sensores de umidade de solo, seriam executados os seguintes passos:

- **baixo nível:**

- receber da RSSF dados brutos dos sensores;
- interpretar os dados;
- filtrar os dados de acordo com o funcionamento nominal dos sensores;
- enviar os dados para a camada de médio nível;

- **médio nível:**

- fusionar os dados interpretados de sensores iguais em um valor único;
- utilizar as regras definidas para realizar uma decisão;
- enviar os dados interpretados e as decisões para a camada de gerência;

- enviar as decisões para a camada de alto nível;
- **alto nível:**
 - receber decisões de múltiplas aplicações;
 - utilizar as regras definidas para realizar uma decisão composta;
 - enviar as decisões compostas para a camada de gerência.

A arquitetura Hydra apresentada é agnóstica quanto ao domínio e à infraestrutura, podendo ser aplicada de maneira distribuída (ex.: a camada de baixo nível pode ser executada completamente em um nó sensor e a de médio nível em um líder de *cluster*) ou de maneira concentrada (ex.: os nós sensores enviam os dados brutos para um nó sorvedouro e as camadas de baixo e médio nível são executadas no sorvedouro ou até mesmo em um serviço na nuvem).

3.2 Conclusão do Capítulo

Neste capítulo, foi apresentada a arquitetura de fusão de dados Hydra, baseada nos modelos de Dasarathy (1997) e Iyengar *et al.* (2001), cujo propósito é ser agnóstica quanto ao domínio e à infraestrutura, realizando fusão multinível e entre as decisões das aplicações. Foram apresentadas as estruturas geral e detalhada das camadas da arquitetura, além da exemplificação da execução de cada camada. No próximo capítulo é abordado um exemplo de instanciação da arquitetura proposta, no âmbito da agricultura inteligente.

4 INSTANCIACÃO DA PROPOSTA EM AGRICULTURA INTELIGENTE

Utilizando a arquitetura Hydra foram desenvolvidas duas aplicações com foco em agricultura inteligente: monitoramento de umidade do solo e estimativa de evapotranspiração. As aplicações compartilham os recursos da infraestrutura de sensores. Como a arquitetura foi especificada de modo a ser agnóstica quanto ao domínio, ela também pode ser aplicada em diferentes contextos e com outras múltiplas aplicações.

4.1 Aplicações

A primeira aplicação desenvolvida visa determinar se um cultivo requer irrigação utilizando a umidade do solo como referência e a segunda realiza um cálculo de evapotranspiração estimado, que será utilizado para determinar o tempo adequado de irrigação. Ambas foram desenvolvidas tendo em vista as necessidades reais de cultivos experimentais de coco e caju anão-precoce, realizados pela Embrapa, em um campo experimental em Paraipaba - CE, auxiliando no monitoramento de experimentos de condicionadores de solo.

4.1.1 Umidade do solo

Esta primeira aplicação requer dados de tensão de água no solo a partir de diversas profundidades para monitorar a absorção da água pelo solo e determinar seu nível de umidade. Um fator que influencia a absorção da água pelo solo é a própria temperatura do solo, sendo necessário também coletar esta informação. Na Figura 7, é detalhada a estrutura da aplicação de acordo com a arquitetura Hydra apresentada no Capítulo 3.

4.1.1.1 Conversão dos dados dos sensores de tensão de água no solo

Os sensores de tensão de água no solo (que indicam a umidade do solo) geram dados brutos em *Ohms*, resultantes da resistência elétrica do solo. Este valor precisa ser interpretado para uma unidade que seja útil para os cálculos agrícolas, Quilopascal (kPa). A fórmula recomendada para esta conversão também requer os dados do sensor da temperatura do solo (SHOCK *et al.*, 1998):

$$umidadeSolo = \frac{3.213 * \frac{dadoBruto}{1000} + 4.093}{1 - 0.009733 * \frac{dadoBruto}{1000} - 0.01205 * temperaturaSolo} \quad (4.1)$$

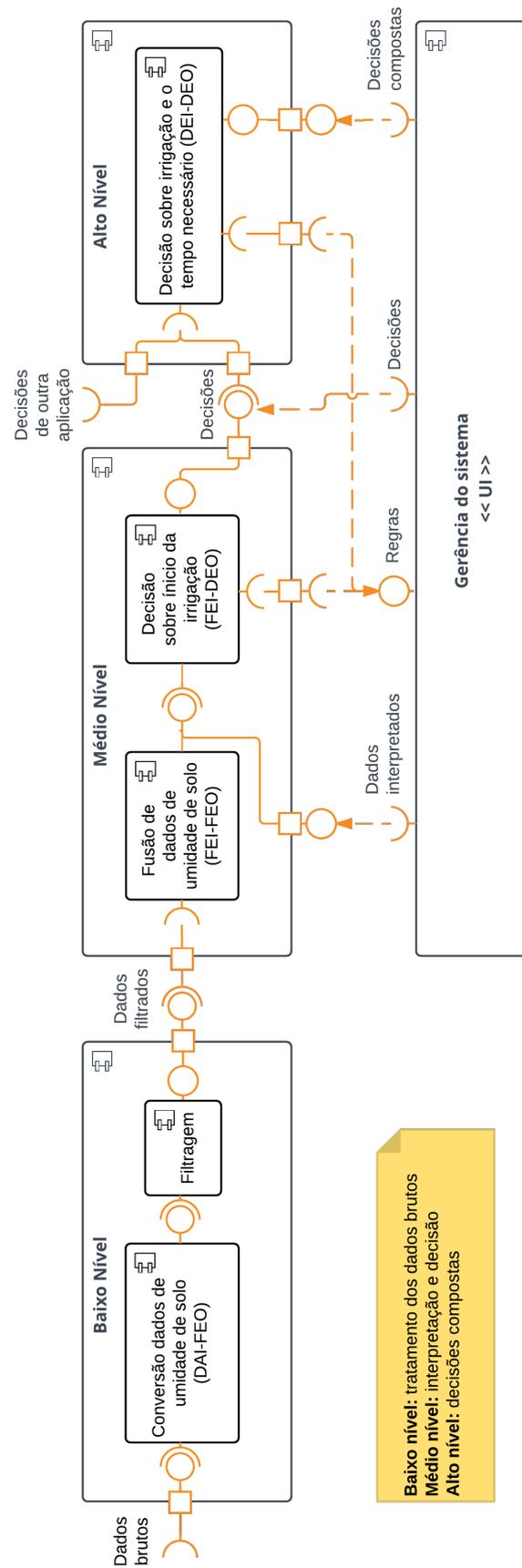


Figura 7 – Níveis da aplicação de umidade de solo

Em que:

umidadeSolo é o resultado do cálculo que indica umidade do solo (kPa);

dadoBruto é o valor bruto provido pelo sensor de tensão de água no solo(Ohm);

temperaturaSolo é a temperatura do solo (°C).

O resultado final é o algoritmo 1:

Algoritmo 1: Algoritmo de conversão dos dados dos sensores de tensão de água no solo

Entrada: Dados brutos dos sensores de tensão de água e de temperatura

Saída: Dados interpretados sobre umidade

início

Coletar dados de tensão de água;

Coletar dados de temperatura;

$umidadeSolo(dadoBruto, temperaturaSolo)$;

fim

Desta maneira, temos uma fusão DAI-FEO, em que dados brutos de dois tipos de sensores são fusionados e o resultado é uma característica de fácil interpretação e análise.

4.1.1.2 Filtragem

Mesmo com a conversão dos dados é necessário garantir que os dados estejam dentro de parâmetros normais de funcionamento. Uma simples filtragem pela especificação de limites mínimos e máximos de cada sensor evita que dados atípicos (*outliers*) causem decisões errôneas no nível médio da arquitetura. *Outliers* são dados que desviam do valor esperado e podem ser causados por falhas na leitura ou erros no experimento. Portanto é importante que os *outliers* sejam identificados adequadamente, pois do contrário, informações importantes podem acabar sendo descartadas erroneamente (SEMATECH, 2003). Essa etapa de filtragem também pode ser utilizada para identificar sensores problemáticos que necessitem de reparo ou substituição.

O algoritmo 2 é um exemplo de filtragem básica, em que os limites mínimo e máximo foram definidos com valores adequados para o clima da região e, caso a temperatura fuja dos valores definidos pelo usuário, ela será removida e o sensor pode ser marcado como potencialmente defeituoso.

Outro exemplo de filtragem de identificação e remoção de *outliers* é apresentado no algoritmo 3, em que um método é utilizado para comparar os dados oriundos de diversos

sensores, realizar uma comparação e identificar o dado que desvia do padrão.

Algoritmo 2: Algoritmo de filtragem de temperatura de solo

Entrada: Dados de temperatura de solo

Saída: Dados filtrados de temperatura de solo

início

Obter dados de temperatura de solo;

tempMax = 32 °C ;

tempMin = 20 °C ;

se tempSolo > tempMax *OU* tempSolo < tempMin **então**

 Remover dado;

 Marcar sensor como potencialmente defeituoso;

fim

fim

Algoritmo 3: Algoritmo de filtragem dos dados de umidade do solo

Entrada: Dados de umidade do solo

Saída: Dados filtrados de umidade do solo

início

Obter dados de umidade do solo de N sensores;

Aplicar método de detecção de outliers (ex.: Método de Chauvenet);

se *Outlier detectado* **então**

 Remover dado;

 Marcar sensor como potencialmente defeituoso;

fim

fim

4.1.1.3 Fusão dos sensores de umidade

Para aumentar a acurácia dos dados, é possível utilizar diversos sensores e realizar uma fusão para obter um dado mais acurado (algoritmo 4). Um cuidado a ser tomado é que, mesmo com a filtragem realizada na camada de baixo nível, um sensor pode gerar *outliers* cujos valores estejam dentro do escopo, seja por falha no sensor ou na comunicação. Uma solução é o uso de algoritmos de detecção e remoção de *outliers*, que comparam os dados enviados por todos os sensores e verificam se algum deles possui um alto desvio padrão, por exemplo. Após a remoção do *outlier*, existem diversas técnicas para obter um valor fusionado, como média

simples, média ponderada e algoritmos mais robustos, como filtro de Kalman.

Algoritmo 4: Algoritmo de fusão de dados de umidade

Entrada: Dados de umidades de solo de múltiplas fontes

Saída: Dado fusionado de umidade de solo

início

Obter dados de temperatura de solo de diversas fontes;

Aplicar algoritmo de identificação de outliers;

se *Outliers forem detectados* **então**

 | Marcar sensor como potencialmente defeituoso;

 | Remover outliers detectados;

fim

Fusionar dados remanescentes;

fim

Neste caso é realizada uma fusão FEI-FEO, em que dados interpretados de múltiplos sensores de umidade são fusionados para gerar um valor mais acurado. Os dados gerados nesta etapa serão utilizados pela etapa de decisão e pela camada de gerência, onde o usuário pode acompanhar o status do cultivo e dos sensores.

4.1.1.4 Decisão sobre irrigação

A decisão sobre a irrigação é feita baseada em regras definidas pelo usuário na camada de gerência e, a partir delas, os dados de umidade e chuva serão avaliados para chegar a uma decisão: irrigar ou não o cultivo (algoritmo 5). Sendo assim, temos uma fusão FEI-DEO, com dados interpretados que geram uma decisão, que pode resultar na ação de um atuador, como a ativação de aspersores ou o envio de mensagem para o usuário.

4.1.2 Evapotranspiração

A segunda aplicação trata da determinação de uma variável chamada de Evapotranspiração de referência (ET_o) (ALLEN *et al.*, 1998), que indica a taxa de evaporação de água pelo solo e de transpiração pelas plantas. Existem diversas técnicas de estimativa de ET_o, desde as mais complexas e precisas, como a Penman-Monteith (ALLEN *et al.*, 1998), que dependem de diversas variáveis disponíveis apenas por meio de uma estação meteorológica, até as mais

Algoritmo 5: Algoritmo de decisão de irrigação

Entrada: Regras de decisão e dados dos sensores de umidade e chuva

Saída: Decisão sobre irrigação

início

Obter regras relacionadas à irrigação;

Obter dados de umidade e chuva;

umidadeSoloMaxima = 60 kPa;

statusChuvaMaxima = 1 (valor booleano);

se *umidadeSolo* >= 60 **E** *statusChuva* <= *statusChuvaMaxima* **então**

retorna *irrigar*;

fim

fim

simples, como a Thornthwaite (THORNTHWAITE, 1948), que depende apenas da temperatura média do ar. E com base na ETo e em outras propriedades, que variam de acordo com o cultivo, é possível determinar a quantidade de água necessária para a irrigação.

Esta aplicação utiliza os dados de sensores instalados no campo experimental de Paraipaba e de sensores de uma estação meteorológica de Itaitinga. Na Figura 8, são mostrados os níveis da aplicação na arquitetura Hydra.

4.1.2.1 Dados da estação meteorológica

Representando a camada de baixo nível, a estação meteorológica recebe os dados brutos de diversos sensores e os converte para dados brutos em unidades físicas (saldo de radiação solar em $\text{MJ m}^{-2} \text{d}^{-1}$, por exemplo). Estes dados passam pelo mesmo tipo de filtragem básica realizada na aplicação de umidade de solo antes de serem utilizados pela camada de médio nível.

4.1.2.2 Cálculo de evapotranspiração

O cálculo de evapotranspiração mais preciso é baseado no método Penman-Monteith:

$$ET_{oPM} = \frac{0,408 * \Delta * (Rn - G) + \frac{\gamma * 900 * U_2 * (e_s - e_a)}{T + 273}}{\Delta + \gamma * (1 + 0,34 * U_2)}, \quad (4.2)$$

Em que:

ET_{oPM} é a evapotranspiração de referência pelo método Penman-Monteith;

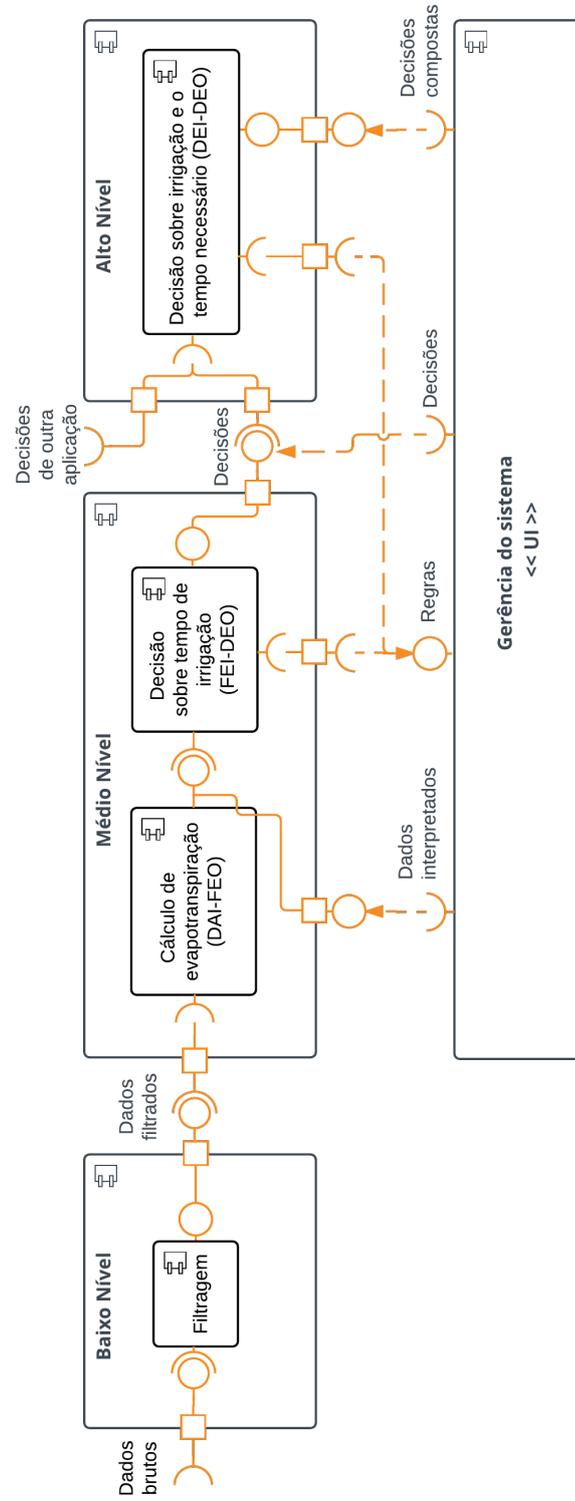


Figura 8 – Níveis da aplicação de evapotranspiração

Δ é a declividade da curva de pressão de vapor em relação à temperatura ($\text{kPa}^\circ\text{C}^{-1}$)

Rn é o saldo de radiação diário ($\text{MJ m}^{-2} \text{d}^{-1}$);

G é o fluxo total diário de calor no solo ($\text{MJ m}^{-2} \text{d}^{-1}$);

γ é o coeficiente psicrométrico ($\text{kPa}^\circ\text{C}^{-1}$);

U₂ é a velocidade do vento a 2 m de altura (m s^{-1});

es é a pressão de saturação de vapor (kPa);

ea é a pressão atual de vapor (kPa);

T é a temperatura média do ar ($^\circ\text{C}$).

que utiliza diversas variáveis meteorológicas e de cultivo, porém esses valores dependem de uma estação com diversos sensores. Uma alternativa viável para nossa região é o método de Hargreaves-Samani (SAMANI, 2000):

$$ET_{oHS} = 0,0135 * K * Ra * \sqrt{(T_{max} - T_{min})} * (T_{med} + 17,8), \quad (4.3)$$

Em que:

ET_{oHS} é evapotranspiração de referência pelo método Hargreaves-Samani (mm/dia);

K é um coeficiente que depende se a cidade é costeira ou continental (adimensional);

Ra é radiação solar no topo da atmosfera, expressa em equivalente de evaporação (mm/dia);

T_{max} é a temperatura máxima ($^\circ\text{C}$);

T_{min} é a temperatura mínima ($^\circ\text{C}$);

T_{med} é a temperatura média ($^\circ\text{C}$);

que depende apenas de dados de um sensor de temperatura. No entanto, este método é considerado menos preciso e requer ajustes a fim de evitar super-estimativas.

Nesta etapa é realizada uma fusão do tipo FEI-FEO, em que a fusão dos dados interpretados pela estação meteorológica gera um novo dado, o valor da evapotranspiração, que será utilizado na etapa de decisão.

4.1.2.3 Decisão sobre tempo de irrigação

Baseado em Allen *et al.* (1998), a decisão sobre quanto de água deve ser utilizada na irrigação depende dos seguintes fatores:

- Evapotranspiração (valor calculado pela aplicação);
- Kc (coeficiente estático, de acordo com o tipo de cultivo);
- Eficiência do sistema de irrigação (valor estático fornecido pelo usuário).

Com base nesses dados é calculado por quanto tempo o cultivo deve ser irrigado (algoritmo 6).

Algoritmo 6: Algoritmo de decisão do tempo de irrigação

Entrada: Regras de decisão (coeficiente K_c e eficiência da irrigação) e valor de evapotranspiração

Saída: Tempo de irrigação

início

Obter regras relacionadas ao tempo de irrigação;

Obter valor de evapotranspiração;

Obter valor de umidade de solo;

Calcular tempo de irrigação;

retorna *Tempo de irrigação*;

fim

4.1.3 Decisão de alto nível

A camada de alto nível recebe as decisões efetuadas pela camada de médio nível e permite uma nova tomada de decisão, baseada nas regras criadas na camada de gerência. O algoritmo 7 apresenta um exemplo de nova decisão baseada nas decisões obtidas pelas duas aplicações apresentadas acima.

Algoritmo 7: Algoritmo de decisão composta (alto nível)

Entrada: Regras de decisão e decisões de médio nível

Saída: Nova decisão sobre irrigação dos cultivos

início

Obter regras relacionadas à fusão de decisões;

Obter decisões das camadas de médio nível;

Decidir quais cultivos devem ser irrigados e por quanto tempo;

retorna *Comandos para aspersores*;

fim

4.1.4 Gerência e interface com usuário

A camada de gerência é responsável pela interação com o usuário, na qual devem ser definidas as regras de decisão das camadas dos níveis médio e alto, a prioridade de cada regra e o tempo de execução das aplicações.

4.2 Conclusão do Capítulo

Neste capítulo, foi apresentada uma instanciação da proposta no domínio de agricultura inteligente, detalhando as aplicações desenvolvidas, o funcionamento de cada camada e algoritmos de exemplos. No próximo capítulo será abordada a implementação das aplicações, do sistema IoT e os experimentos realizados.

5 EXPERIMENTOS E ANÁLISE DE RESULTADOS

Neste capítulo é detalhado o design experimental realizado, apresentando os materiais (ambiente, hardware, comunicação, software) e métodos de execução e análise dos três experimentos realizados.

5.1 Material

5.1.1 Ambiente

O ambiente de testes foi o campo experimental da Embrapa em Paraipaba - CE. No local são realizados ensaios de técnicas de manutenção, cultivo e irrigação. Um dos ensaios é a análise de condicionadores de solo (ex: hidrogel e biocarvão) em cultivos de coco e caju anão-precoce, que visam melhorar a retenção de água no solo e assim reduzir o estresse do cultivo em períodos de estiagem.

5.1.2 Protótipo de hardware

Dois tipos de nós sensores foram desenvolvidos para o estudo de caso. O primeiro tipo (ver Figura 9) foi criado para verificar a viabilidade do uso do sensor Irrometer Watermark 200SS¹ em conjunto com Arduino. O segundo (ver Figura 10) foi desenvolvido para realizar a coleta de dados climáticos (temperatura do solo e chuva), também utilizando Arduino. Ambos os nós sensores utilizam pilhas AA de 1.5v como fonte de energia, e, para alcançar uma vida útil maior, o Arduino de alguns dos nós foi modificado, do qual foram removidos o controlador de tensão e os LEDs de status.

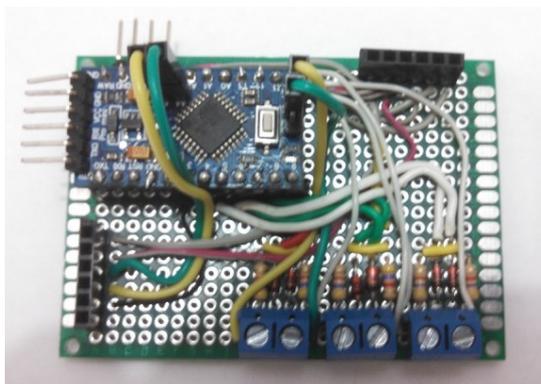


Figura 9 – Nó sensor tipo A (sem módulo de rádio e sensores)

¹ Irrometer Watermark 200SS: <<http://www.irrometer.com/sensors.html#wm>>

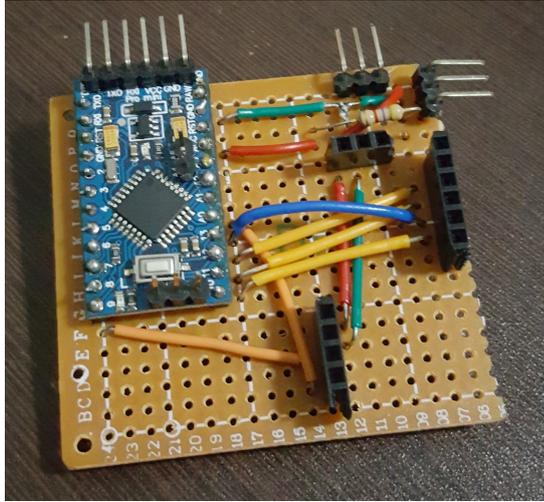


Figura 10 – Nó sensor tipo B (sem módulo de rádio e sensores)

A plataforma Arduino foi escolhida como base para o desenvolvimento do protótipo de nó sensor por se tratar de um *hardware* de código aberto (sob a licença *Creative Commons Attribution Share-Alike*²), de fácil acesso, baixo custo e que permite rápida prototipação.

Antes da instalação dos protótipos em campo, foi realizada uma medição do consumo energético dos nós sensores utilizando um medidor de tensão e corrente USB para determinar a economia alcançada com a modificação do Arduino (ver Tabela 2). Os Arduinos não modificados e modificados foram comparados nos seguintes cenários: ocioso (sem nenhum módulo de rádio acoplado), ocioso (com módulo de rádio acoplado) e durante a transmissão de mensagem. Em modo ocioso, a redução no consumo de energia chegou a 99,67%, e a 32,69% durante a transmissão. Como os nós sensores passam a maior parte do tempo em hibernação, tal redução aumentará consideravelmente a vida útil deles em campo.

Tabela 2 – Consumo energético dos nós sensores

Atividade	Arduino Pro Mini	Arduino Pro Mini (modificado)	Redução de consumo
Ocioso (sem módulos)	3,05 mA	>0,01 mA	99,67%
Ocioso (com módulos)	11,50 mA	2,20 mA	80,87%
Transmitindo	26,00 mA	17,50 mA	32,69%

Fonte – O autor. Medição realizada com testador USB de tensão e corrente.

O nó sorvedouro (ver Figura 11) foi desenvolvido tendo como base um Raspberry Pi Zero, que é responsável por coletar os dados gerados pelos nós sensores. Apenas este nó foi ligado diretamente na rede elétrica.

² CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/deed.pt_BR>



Figura 11 – Nó sorvedouro

Eis as especificações completas de todos os nós:

- Protótipo de nó sensor tipo A:
 - Arduino Pro Mini;
 - 3x Sensor de Umidade de Solo Irrrometer Watermark 200SS;
 - nRF24L01+ com antena externa e amplificador de potência.
- Protótipo de nó sensor tipo B:
 - Arduino Pro Mini;
 - Sensor de chuva (*wetness*);
 - Sensor de temperatura à prova d'água;
 - nRF24L01+ com antena externa e amplificador de potência.
- Protótipo de nó sorvedouro:
 - Raspberry Pi Zero;
 - nRF24L01+ com antena externa e amplificador de potência;
 - Módulo Wi-fi USB.

No cultivo de caju foram utilizados nós sensores tipo A (4 unidades), instalados de acordo com a Figura 12. Os sensores ficaram de ~11m a ~22m de distância do nó sorvedouro, estando todos os nós instalados em campo aberto, com obstáculos inerentes a um cultivo (folhas, galhos, troncos, etc). As siglas de T1 a T4 representam as plantas cujo solo foi tratado com biocarvão, enquanto as siglas de T5 a T8 representam os tratamentos de hidrogel e a sigla T9 representa as testemunhas, ou seja, o solo não recebeu qualquer tipo de tratamento.

Já no cultivo de coqueiros, foram utilizados nós sensores tipo A (4 unidades) e tipo B (1 unidade), instalados de acordo com a Figura 13. Os nós sensores ficaram de 48m a 98m de distância do nó sorvedouro. Todos os nós sensores ficaram em campo aberto e o nó sorvedouro foi instalado em uma casa de apoio existente no terreno.

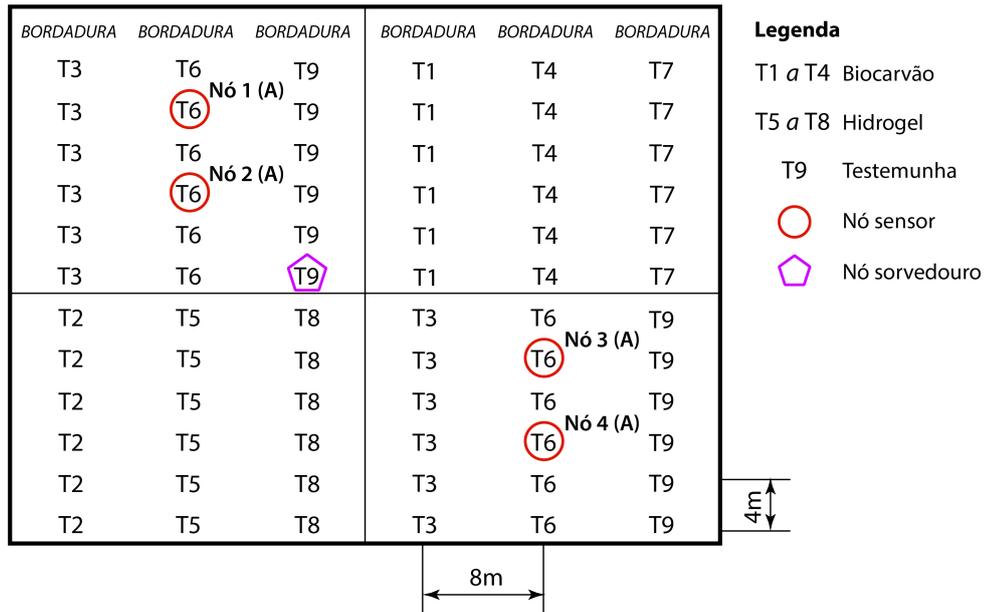


Figura 12 – Croqui de trecho do campo experimental de caju anão-precoce

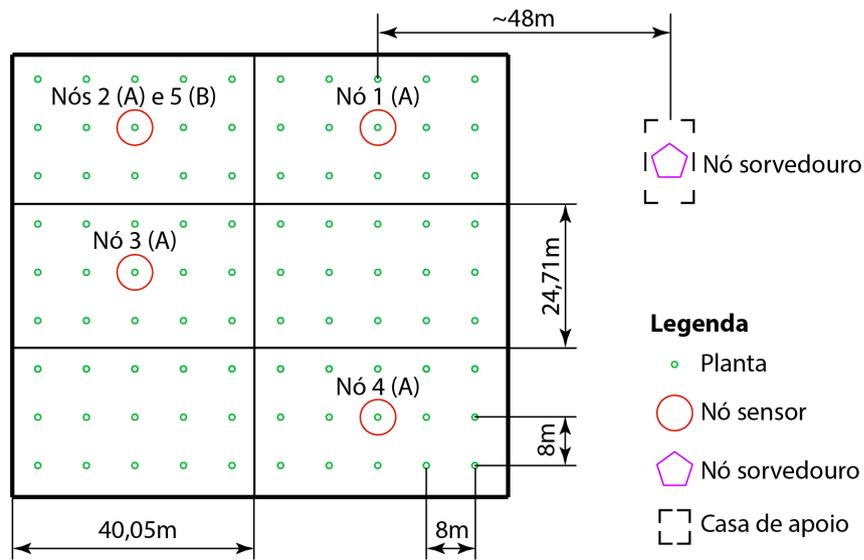


Figura 13 – Croqui de trecho do campo experimental de coqueiros

5.1.3 Software

O software de controle dos nós sensores foi escrito em C++, utilizando bibliotecas desenvolvidas para a plataforma Arduino. O software do nó sorvedouro, responsável pela coleta dos dados, gravação no banco de dados e comunicação com a internet foi escrito em Python. Já as aplicações de umidade do solo e de evapotranspiração (apresentadas na seção 4.1) foram escritas em MATLAB.

As camadas de fusão da arquitetura Hydra foram executadas posteriormente, no ambiente do MATLAB, utilizando os dados coletados entre abril e maio de 2017. Na subseção 5.2.1 é explicado como foi realizada a coleta.

Nas próximas subseções, 5.1.3.1 e 5.1.3.2, são apresentados os parâmetros e algoritmos utilizados para realizar a filtragem dos dados das camadas de nível baixo e médio. E na subseção 5.1.3.3 é abordado o que foi implementado da camada de gerência.

5.1.3.1 Filtragem: camada de nível baixo

Conforme apresentado no capítulo anterior, antes dos dados serem enviados para a camada de médio nível é necessário identificar e remover *outliers*, ou seja, dados que fogem dos parâmetros normais de funcionamento dos sensores e que podem afetar a acurácia dos resultados. Alguns dos critérios de remoção são apresentados na Tabela 3, sendo determinados pelos parâmetros normais de funcionamento de cada tipo de sensor.

Tabela 3 – Critérios de remoção de *outliers*

Atributo	Critérios
todos	valores inválidos (ex.: -1)
umidade de solo	< 0 ou > 200 kPa
temperatura do solo	< 20 ou > 32 °C
radiação	< 0 kJ m ⁻²

Fonte – O autor

O objetivo da identificação e remoção de *outliers* na camada de baixo nível é reduzir o ruído da leitura dos sensores, resultando em uma maior acurácia dos dados gerados.

5.1.3.2 Filtragem: camada de nível médio

Mesmo após a filtragem na camada de nível baixo, é possível que ainda existam *outliers*, mas dentro do escopo de valores normais de funcionamento do sensor, seja ele causado por falhas ou por ruído na leitura do sensor.

Na aplicação de umidade de solo é realizada a fusão de múltiplos sensores de umidade de solo, possibilitando a comparação dos valores entre os sensores por meio de métodos de identificação de *outliers*. Os seguintes métodos foram avaliados:

- Método de Chauvenet;
- Critério de Peirce;
- Z-Score;
- Z-Score Modificado;
- Boxplot Ajustado;

- ESD (*Generalized Extreme Studentized Deviate*).

Dentre os métodos citados acima, o que apresentou melhores resultados foi o ESD, que identificou o maior número de *outliers* na base de dados utilizada (obtida em uma coleta preliminar de testes). Na seção 5.3 é apresentado como foi realizada a análise dos dados e a escolha pelo método ESD em conjunto com um filtro de Kalman ponderado.

5.1.3.3 Interface com usuário: camada de gerência

A camada de gerência e interface com usuário foi implementada utilizando a plataforma Thingspeak³. O nó sorvedouro se comunica com a internet e publica os dados coletados na plataforma Thingspeak utilizando o protocolo MQTT, permitindo a visualização em tempo real dos dados. No momento, as regras não são definidas pela interface do Thingspeak, mas por parâmetros definidos antes de executar as aplicações.

5.1.4 Comunicação

A comunicação entre os nós ocorre por uma rede não-IP através dos módulos de rádio nRF24L01+ e utilizando a bibliotecas nRF24⁴ e nRF24Network⁵. Na Figura 14, é apresentado um exemplo de estrutura de mensagem transmitida pelos nós (neste caso é a mensagem transmitida pelo nó sensor tipo B). Após a transmissão de cada mensagem, para economizar energia, o nó sensor entra em hibernação durante uma hora, em que ao acordar reinicia o ciclo de realizar medições e enviar dados.

Cabeçalho (8 bytes)				Mensagem (13 bytes)				
Origem (2 bytes)	Destino (2 bytes)	ID (2 bytes)	Tipo (2 bytes)	Frag. (2 bytes)	Bateria (4 bytes)	Temper. (4 bytes)	Chuva An. (4 bytes)	Chuva Di. (1 byte)

Mensagem máxima: 24 bytes (sem fragmentação)
120 bytes (com fragmentação)

Figura 14 – Exemplo de estrutura de mensagem

Já a comunicação do nó sorvedouro com a internet foi por meio de comunicação Wi-Fi, utilizando o protocolo MQTT como protocolo de mensagens. O protocolo MQTT foi escolhido para uso na comunicação do nó sorvedouro com a internet por adotar o padrão

³ Thingspeak: <<http://www.thingspeak.com/>>

⁴ nRF24: <<https://github.com/nRF24/RF24>>

⁵ nRF24Network: <<https://github.com/nRF24/RF24Network>>

Publisher/Subscriber, onde os dados ficam centralizados em um *gateway* chamado de *broker*, que gerencia os dados e requisições dos clientes⁶, ou seja, os dados gerados pela infraestrutura de RSSF são enviados, por meio do nó sorvedouro, para um *broker* na nuvem, tornando os dados acessíveis por qualquer serviço ou dispositivo na internet.

5.2 Planejamento dos Experimentos

O experimento I teve como objetivo avaliar diversos algoritmos de identificação de *outliers* e filtros de suavização de sinal. O experimento II visou avaliar os dados obtidos pelas camadas de nível médio das aplicações de evapotranspiração e de umidade de solo. E o experimento III avaliou a viabilidade de criação de um novo modelo de evapotranspiração com os dados sensoriados.

5.2.1 Coleta de dados

Foram realizadas duas coletas distintas em dois períodos: *31 de janeiro a 07 de fevereiro de 2017*, no cultivo de caju anão-precoce, e de *25 de abril a 29 de maio de 2017*, no cultivo de coqueiros.

5.2.1.1 Primeira Coleta

Na primeira coleta, os nós sensores enviavam os dados coletados a cada 1 hora ao nó sorvedouro, que os armazenava em um banco de dados MongoDB. A lista completa dos atributos segue na Tabela 4.

Os dados brutos contêm algumas lacunas nos dados, provenientes de falhas na comunicação ou no próprio hardware. Da quantidade original de **168 tuplas**, após o pré-processamento dos dados, o valor final é de **151 tuplas**.

Tabela 4 – Coleta 01: lista completa de atributos

Nó	Atributo	Unidade
nós 1 a 4 (tipo A)	umidade de solo a 15 cm de profundidade	Ohm
	umidade de solo a 45 cm de profundidade	Ohm
	umidade de solo a 75 cm de profundidade	Ohm

Fonte – O autor

⁶ Para mais informações sobre o funcionamento do protocolo MQTT: <<http://www.hivemq.com/mqtt-essentials/>>

5.2.1.2 Segunda coleta

Na segunda coleta, os nós sensores enviavam os dados coletados a cada 30 minutos, o nó sorvedouro os armazenava em um banco de dados MongoDB e, a cada hora, os enviava para o serviço Thingspeak via Wi-Fi. Os dados meteorológicos foram coletados da estação automática do Instituto Nacional de Meteorologia (INMET), localizada em Itapipoca, Ceará. Esta estação foi escolhida por ser a mais próxima da fazenda experimental de Paraipaba (aproximadamente 46 quilômetros). Os dados da estação são gerados a cada hora e disponibilizados automaticamente por meio do site do serviço ou por bibliotecas que acessam os dados diretamente, que foram armazenados em arquivos CSV. A lista completa dos atributos segue na Tabela 5.

Os dados brutos contêm algumas lacunas nos dados, provenientes de falhas na comunicação ou no próprio hardware. Para facilitar a comparação e análise, os dados foram padronizados em intervalos de sessenta minutos, resultando em um total de **667 tuplas**. Após o pré-processamento, o uso de filtros e a remoção de *outliers*, o valor final é de **340 tuplas**.

Tabela 5 – Coleta 02: Lista completa de atributos

Nó	Atributo	Unidade
nós 1 a 4 (tipo A)	umidade de solo a 15 cm de profundidade	Ohm
	umidade de solo a 45 cm de profundidade	Ohm
	umidade de solo a 75 cm de profundidade	Ohm
nó 5 (tipo B)	chuva	-
	temperatura do solo a 45cm de profundidade	Celsius
estação meteorologica de itapipoca	temperatura instantanea (Celsius)	Celsius
	temperatura máxima (Celsius)	Celsius
	temperatura mínima (Celsius)	Celsius
	umidade relativa do ar instantanea (%)	%
	umidade relativa do ar máxima (%)	%
	umidade relativa do ar mínima (%)	%
	ponto de orvalho instantaneo (Celsius)	Celsius
	ponto de orvalho máximo (Celsius)	Celsius
	ponto de orvalho mínimo (Celsius)	Celsius
	pressão atmosférica (kPa)	kPa
	pressão atmosférica máxima (kPa)	kPa
	pressão atmosférica mínima (kPa)	kPa
	velocidade do vento (m/s)	m/s
	direção do vento (graus)	m/s
	rajada de vento (m/s)	m/s
	radiação solar (kJ/m ²)	kJ/m ²
precipitação (mm)	mm	

Fonte – O autor

5.3 Discussão do Experimento I: detecção e remoção de *outliers*

Os dados da coleta 01 foram utilizados para avaliar diversos métodos de identificação de *outliers* com objetivo de melhorar a acurácia (subseção 5.3.1). Os seguintes métodos foram avaliados:

- Método de Chauvenet;
- Critério de Peirce;
- Z-Score;
- Z-Score Modificado;
- Boxplot Ajustado;
- *Generalized Extreme Studentized Deviate* (ESD).

Além desses algoritmos, também avaliamos o uso de filtros para suavizar o ruído inerente da leitura de sensores e remover possíveis *outliers* remanescentes (subseção 5.3.2). Os filtros avaliados foram:

- Filtro de Kalman;
- *Weighted outlier-robust Kalman filter* - Filtro de Kalman ponderado robusto a outliers (WRKF);
- Savitzky-Golay;
- *Robust locally weighted scatterplot smoothing* (RLOESS e RLOWESS);
- Espaço-escala;

5.3.1 Algoritmos de identificação de *outliers*

Na Figura 15, são exibidos *outliers* detectados pelo Critério de Peirce e pelo Z-Score Modificado. Os outros métodos não estão presentes pois não detectaram qualquer *outlier* (lista completa de resultados na Tabela 6). O Critério de Peirce detectou alguns picos oriundos do nó 2, mas falhou em detectar os valores mais altos, provavelmente porque o nó 3 parou de funcionar e ele teve apenas 3 valores para realizar a comparação. O Z-Score Modificado falhou em detectar os picos iniciais (menores), mas continuou a detectar *outliers* mesmo após a morte do nó 3.

Na profundidade de 45cm, tivemos resultados mais interessantes devido à clara presença de *outliers* oriundos do nó 3. As Figuras 16 e 17 apresentam os *outliers* detectados e, ainda assim, alguns métodos foram incapazes de detectar qualquer dado. É importante deixar claro que devido aos valores extremos dos *outliers* nos dias 31 de janeiro e 01 de fevereiro, o eixo

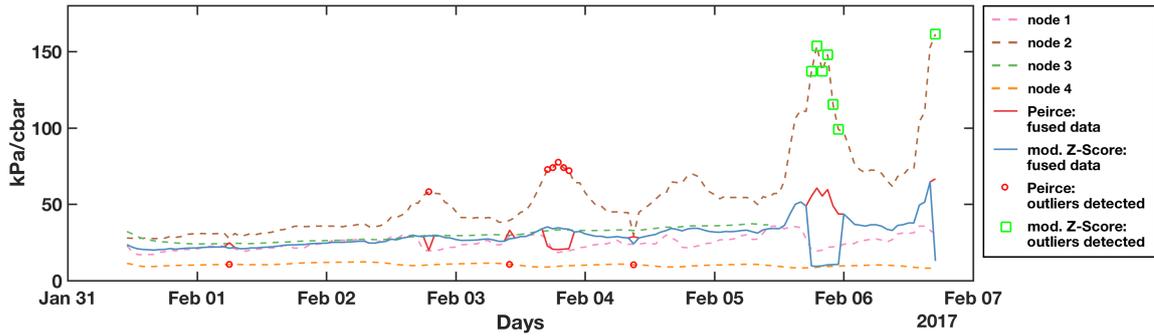


Figura 15 – Outliers e dado fusionado (15cm de profundidade): Critério de Peirce e Z-Score Modificado

Y precisou ser representado na escala logarítmica para que os dados ficassem legíveis. Eis os resultados dos métodos que detectaram *outliers*:

- Boxplot Ajustado: falhou em detectar os *outliers* de 31 de janeiro e 01 de fevereiro. Em 05 de fevereiro começou a detectar *outliers* do nó 4;
- ESD: detectou com sucesso os *outliers* de 31 de janeiro e 01 de fevereiro e também detectou os mesmos *outliers* que o Boxplot Ajustado;
- Critério de Peirce e Z-score modificado: estes métodos foram os mais sensíveis, detectando *outliers* em cinco dos sete dias monitorados, acabando por remover dados válidos.

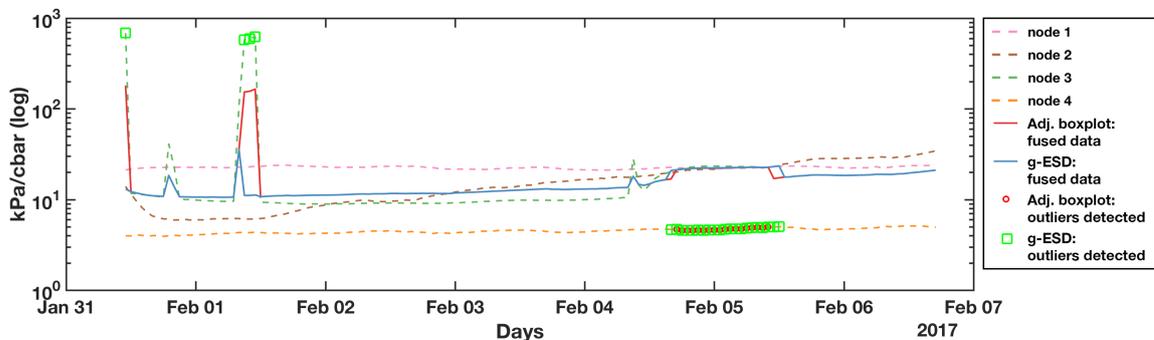


Figura 16 – Outliers e dados fusionados (45cm de profundidade): Boxplot Ajustado e ESD

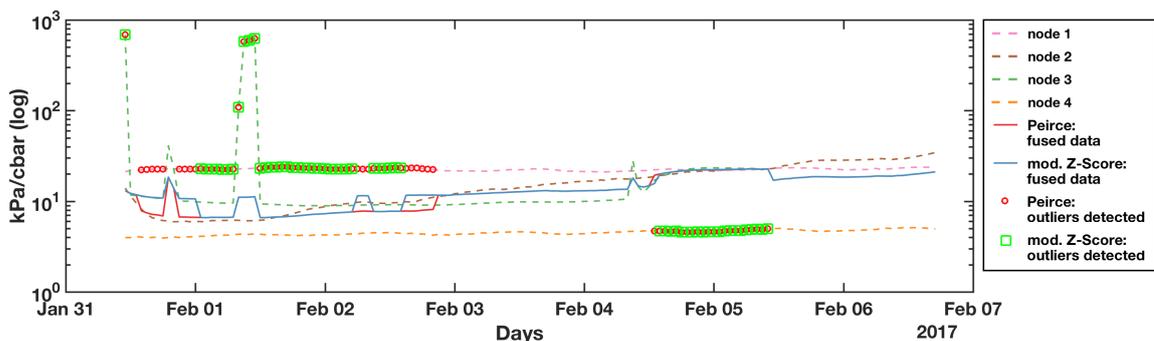


Figura 17 – Outliers e dados fusionados (45cm de profundidade): Critério de Peirce e Z-Score Modificado

E na profundidade de 75cm, todos os métodos funcionaram corretamente e não detectaram nenhum *outlier*, já que os dados coletados pelos sensores não apresentaram grandes variações.

Desta maneira, com base nos dados apresentados, decidimos adotar o método ESD por ter sido capaz de detectar os *outliers* óbvios e por não ser tão sensível ao ponto de remover dados relevantes. Na Tabela 6, é apresentado um resumo da performance de cada método. Agora cabe aos filtros suavizarem o ruído e qualquer *outlier* remanescente.

Tabela 6 – Total de *outliers* detectados

	15cm	45cm	75cm	Comentários
Boxplot Ajustado	0	18	0	não removeu <i>outliers</i> óbvios
Chauvenet	0	0	0	não detectou <i>outliers</i>
ESD	0	25	0	método adotado
Crítério de Peirce	9	76	0	sensível demais
Z-score	0	0	0	nenhum <i>outlier</i> detectado
Z-score Modificado	7	63	0	sensível demais

5.3.2 Filtragem de sinal

Os dados obtidos após a remoção de *outliers* pelo método ESD agora foram filtrados para remover o ruído inerente da leitura de sinais de sensores. Todos os filtros testados suavizaram os dados, mas os resultados variaram quanto à remoção de picos nos dados.

Nos dados fusionados de 15cm de profundidade é notável que os filtros RLOESS e RLOWESS suavizaram demais os dados após 5 de fevereiro (Figura 18). O filtro de Kalman apresentou um resultado mais próximo do dado original, porém ele foi influenciado pelo pico no final do dia 5 de fevereiro. No geral, o filtro WRKF apresentou o melhor resultado. Na Figura 19, são apresentados os dados dos filtros Savitzky-Golay e espaço-escala, que removeram o ruído, mas seguiram fielmente o formato dos dados oriundos do ESD.

Nos dados de 45cm profundidade, apesar do ESD ter removido os *outliers* mais óbvios, ainda restaram alguns dados anômalos. Nas Figuras 20 e 21, é exibido como os filtros de Kalman, Savitzky-Golay e espaço-escala são afetados pelos picos presentes nos dias 31 de janeiro e 01 de fevereiro. RLOESS e RLOWESS não foram afetados por tais picos, mas apenas o RLOWESS foi resiliente aos picos presentes em 05 de fevereiro. Novamente, o filtro que apresentou melhor resultado foi o WRKF.

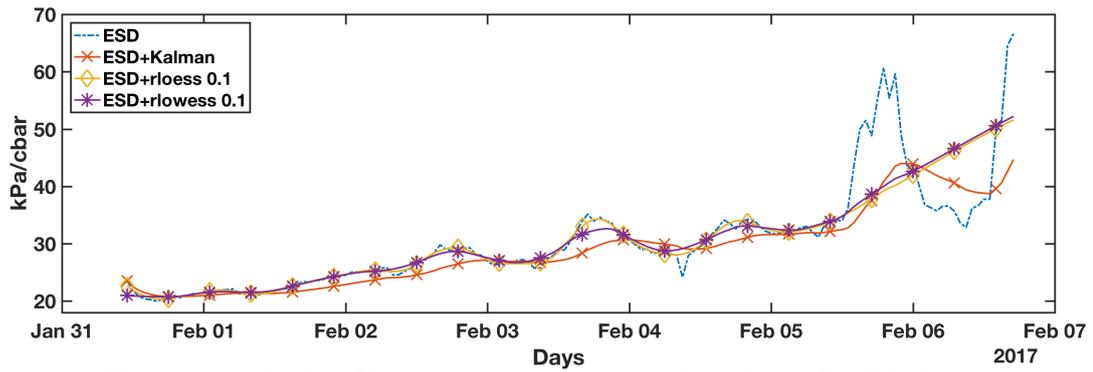


Figura 18 – Dados filtrados dos sensores a 15cm de profundidade (a)

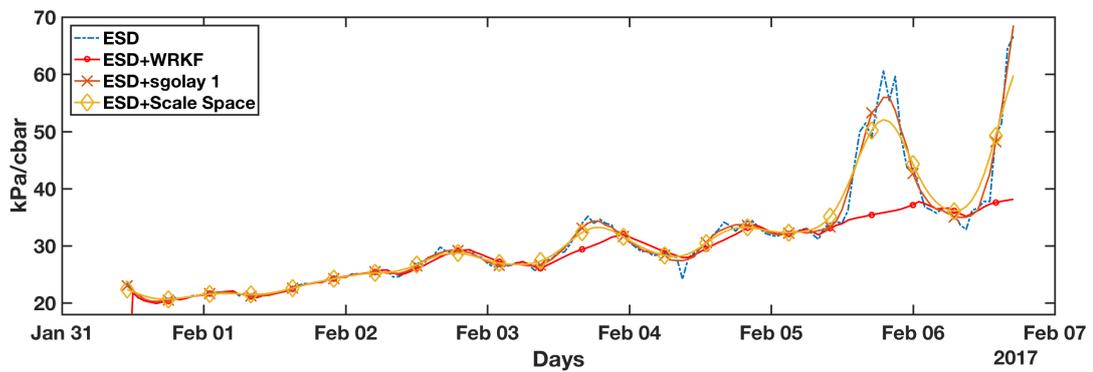


Figura 19 – Dados filtrados dos sensores a 15cm de profundidade (b)

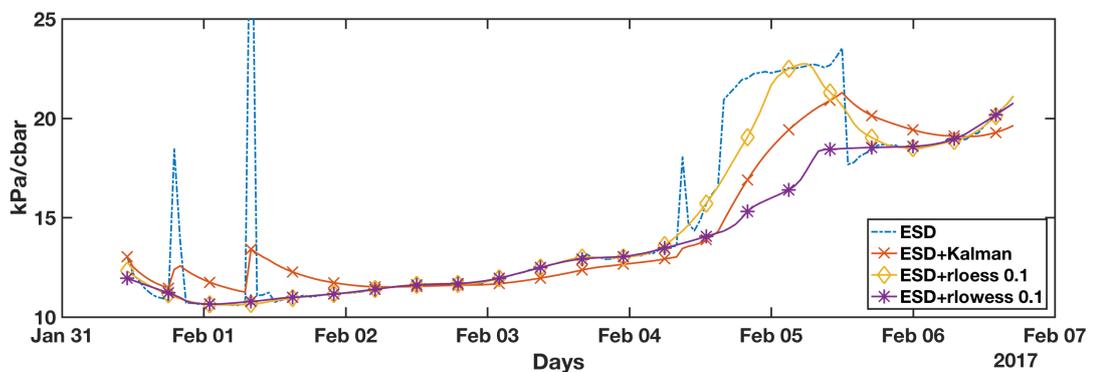


Figura 20 – Dados filtrados dos sensores a 45cm de profundidade (a)

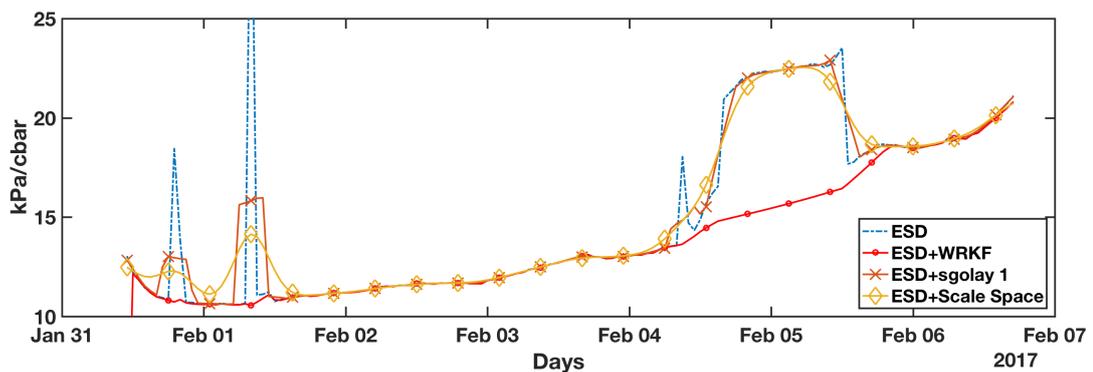


Figura 21 – Dados filtrados dos sensores a 45cm de profundidade (b)

E nos dados dos sensores a 75cm de profundidade, a performance de quase todos os filtros foi similar (Figuras 22 e 23), com exceção do filtro de Kalman, que apresentou um viés (*bias*) negativo.

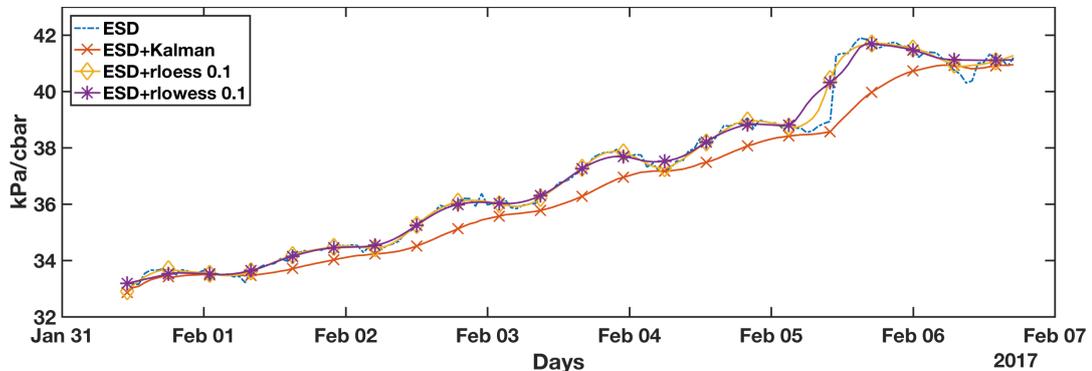


Figura 22 – Dados filtrados dos sensores a 75cm de profundidade (a)

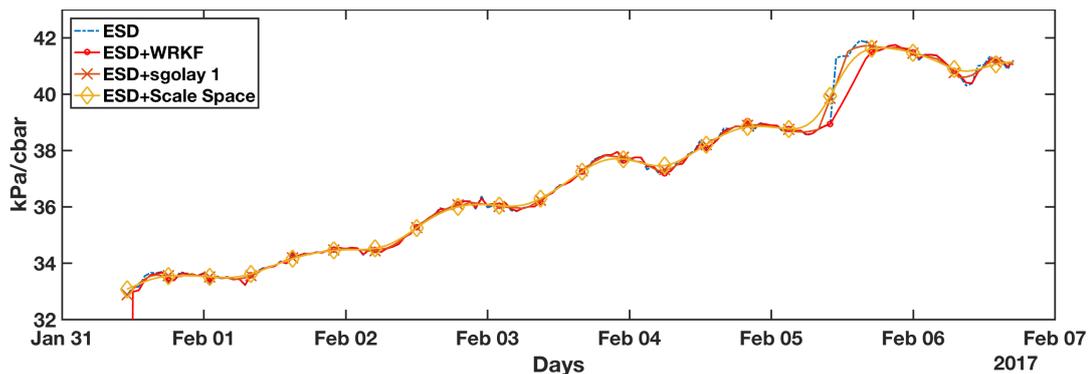


Figura 23 – Dados filtrados dos sensores a 75cm de profundidade (b)

Sendo assim, o filtro WRKF demonstrou ser o mais adequado, tendo maior resiliência a *outliers* residuais e picos de dados anômalos (possivelmente causados por sensores falhos).

5.3.3 Conclusão do Experimento I

Com base nos dados coletados e no experimento realizado, decidimos adotar o método ESD em conjunto com o filtro WRKF por terem sido capazes de detectar os *outliers* óbvios e por não serem tão sensíveis ao ponto de remover dados relevantes.

Estes algoritmos foram utilizados para tratar os dados da coleta 02, mais especificamente na camada de fusão de médio nível para identificar e remover *outliers* na fusão de sensores de umidade de solo.

5.4 Discussão do Experimento II: aplicações de evapotranspiração e umidade do solo

Este experimento trata da execução da aplicação de umidade de solo e da aplicação de evapotranspiração, com base nos dados da coleta 02, e da avaliação dos dados obtidos utilizando um software de análise e mineração de dados.

5.4.1 Aplicação de Umidade de Solo

Conforme apresentado na subseção 4.1.1, o resultado da camada de nível médio da aplicação de umidade de solo é uma decisão referente ao início da irrigação do cultivo. Esta decisão é realizada com base nas informações de chuva e de umidade do solo.

O período monitorado coincidiu com um período de chuvas constantes, então a umidade do solo nunca chegou perto do valor de referência adequado para o cultivo (60 kPa), valor este informado pelo especialista da Embrapa. Mas, para ilustrar o funcionamento da decisão, o valor mínimo de umidade de solo foi definido como 10 kPa, resultando nos dados da Figura 24. Sendo assim, é possível observar que a aplicação determinou corretamente quando deve ser iniciada a irrigação, de acordo com as regras determinadas.

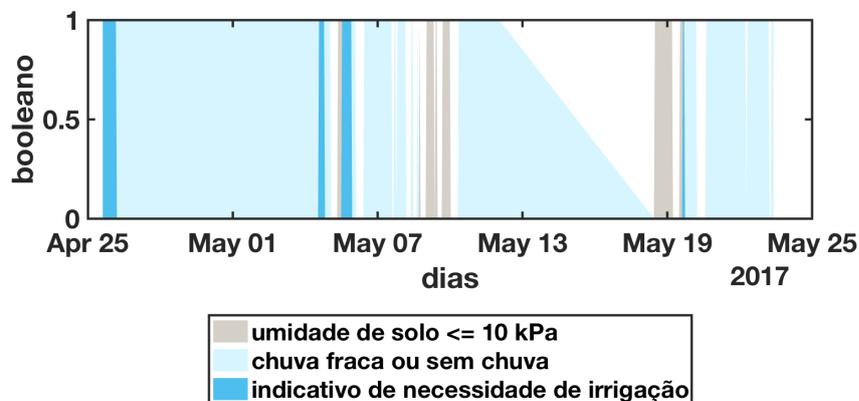


Figura 24 – Resultados da decisão de início de irrigação

5.4.2 Aplicação de Evapotranspiração

Enquanto a decisão da aplicação de umidade se refere quando deve ser iniciada a irrigação, a decisão da aplicação de evapotranspiração se refere ao tempo de irrigação ideal para o cultivo, utilizando uma gama de sensores de uma estação meteorológica para efetuar o cálculo de ETo pelo método Penman-Monteith.

De posse dos dados ETo horários e dos parâmetros do cultivo, obtivemos o tempo ideal de irrigação para o cultivo de coqueiros no campo experimental de Paraipaba (Figura 25). Essa decisão é específica para este cultivo, pois leva em consideração a eficiência da irrigação (valor obtido empiricamente pelo especialista) e um coeficiente específico para a cultura.

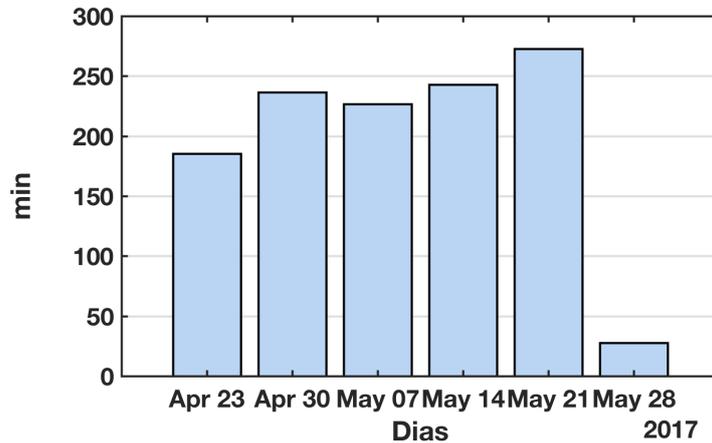


Figura 25 – Resultados da decisão de tempo de irrigação

5.4.3 Validação dos Dados

O objetivo desta etapa é validar os dados gerados pelos sensores e pelas aplicações de umidade de solo e de evapotranspiração, verificando sua correlação com os dados da estação meteorológica de Itapipoca. Para realizar essa validação, utilizamos um método preditivo para gerar um ETo com todos os dados dos sensores e da estação, e verificamos se seu resultado é tão confiável quanto o método Penman-Monteith.

Utilizamos o software WEKA (aplicativo *opensource* multiplataforma de análise e mineração de dados), em conjunto com o algoritmo de classificação M5P (algoritmo que cria árvores de decisões combinadas com modelos de regressão linear), realizando uma validação cruzada em 10 partes (*folds*).

O algoritmo M5P gerou uma árvore de decisão com 6 regras, tendo como base o parâmetro de radiação solar. O Código-fonte 1 lista o resultado de processamento do algoritmo. Na Tabela 7 são listadas as regras criadas e a árvore de decisão resultante está na Figura 26.

Código-fonte 1 – WEKA: resultado do algoritmo M5P com dados da estação meteorológica e dados fusionados com método ESD + WRKF

```
1 === Run information ===
2
3 Scheme:          weka.classifiers.trees.M5P -M 4.0
4 Relation:        dissertacao_fusao -weka.filters.unsupervised.attribute.
                    Remove-R1
5 Instances:       340
6 Attributes:      24
7                 umidade_solo_15cm
8                 umidade_solo_45cm
9                 umidade_solo_75cm
10                chuva
11                temperatura_solo
12                temperatura_ar_instante
13                temperatura_ar_maxima
14                temperatura_ar_minima
15                umidade_instante
16                umidade_maxima
17                umidade_minima
18                pto_orvalho_instante
19                pto_orvalho_maximo
20                pto_orvalho_minima
21                pressao
22                pressao_maxima
23                pressao_minima
24                vento_velocidade
25                vento_direcao
26                vento_rajada
27                radiacao
28                precipitacao
29                temperatura_media
30                EToPM
31 Test mode:       10-fold cross-validation
32
33 === Classifier model (full training set) ===
34
35 M5 pruned model tree:
36 (using smoothed linear models)
37
38 radiacao <= 0.627 :
```

```

39 | radiacao <= 0.03 : EQ1 (178/3.468\%)
40 | radiacao > 0.03 : EQ2 (45/3.992\%)
41 radiacao > 0.627 :
42 | radiacao <= 2.289 :
43 | | radiacao <= 1.819 : EQ3 (38/11.045\%)
44 | | radiacao > 1.819 :
45 | | | temp_min <= 29.05 : EQ4 (21/3.551\%)
46 | | | temp_min > 29.05 : EQ5 (9/10.528\%)
47 | radiacao > 2.289 : EQ6 (49/3.67\%)
48
49 Number of Rules : 6
50
51 Time taken to build model: 0.41 seconds
52
53 === Cross-validation ===
54 === Summary ===
55
56 Correlation coefficient          0.9958
57 Mean absolute error             0.018
58 Root mean squared error        0.0291
59 Relative absolute error        6.7513 \%
60 Root relative squared error    9.1476 \%
61 Total Number of Instances      340

```

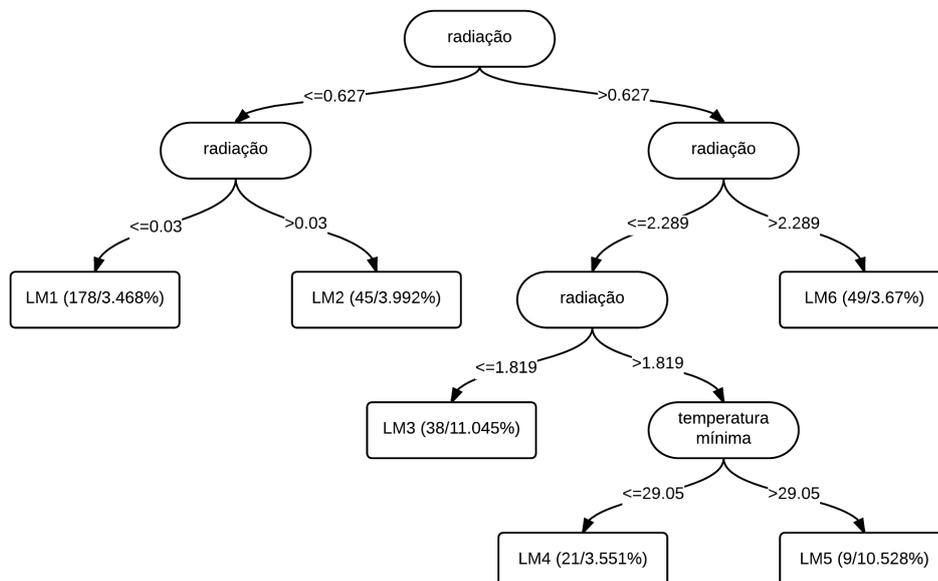


Figura 26 – WEKA: árvore de decisão gerada pelo algoritmo M5P

Tabela 7 – WEKA: regras criadas pelo método M5P

<p>EQ num: 1 EToPM = -0.0001 * umidade_solo_15cm - 0.0002 * umidade_solo_45cm + 0.0002 * umidade_solo_75cm - 0.0014 * temperatura_solo + 0.0031 * temperatura_ar_instante - 0.0011 * temperatura_ar_minima - 0.0007 * umidade_instante - 0.0001 * umidade_maxima + 0 * umidade_minima - 0.0105 * pto_orvalho_instante + 0.0066 * pressao + 0.0133 * vento_velocidade + 0.0002 * vento_rajada + 0.0351 * radiacao - 0.3821</p>	<p>EQ num: 2 EToPM = -0.0024 * umidade_solo_15cm - 0.0026 * umidade_solo_45cm - 0.0169 * temperatura_solo - 0.0008 * temperatura_ar_instante - 0.0011 * temperatura_ar_minima - 0.002 * umidade_instante - 0.0001 * umidade_maxima + 0 * umidade_minima - 0.0061 * pto_orvalho_instante - 0.0043 * pto_orvalho_minima + 0.0066 * pressao + 0.0022 * vento_velocidade + 0.0002 * vento_rajada + 0.2303 * radiacao + 0.3554</p>
<p>EQ num: 3 EToPM = 0.0003 * umidade_solo_45cm - 0.0036 * temperatura_solo + 0.0115 * temperatura_ar_instante - 0.013 * temperatura_ar_minima + 0.0005 * umidade_instante - 0.0002 * umidade_maxima - 0.001 * umidade_minima - 0.0033 * pto_orvalho_instante + 0.2437 * pressao - 0.0047 * vento_velocidade + 0.003 * vento_rajada + 0.2401 * radiacao - 24.0246</p>	<p>EQ num: 4 EToPM = 0.0004 * umidade_solo_45cm - 0.0007 * umidade_solo_75cm - 0.0014 * temperatura_solo + 0.0123 * temperatura_ar_instante - 0.0187 * temperatura_ar_minima + 0.0005 * umidade_instante - 0.0002 * umidade_maxima - 0.0021 * umidade_minima - 0.0033 * pto_orvalho_instante + 0.0988 * pressao - 0.0055 * vento_velocidade + 0.0034 * vento_rajada + 0.2545 * radiacao - 9.3707</p>
<p>EQ num: 5 EToPM = 0.0004 * umidade_solo_45cm - 0.0011 * umidade_solo_75cm - 0.004 * temperatura_solo + 0.0123 * temperatura_ar_instante - 0.0248 * temperatura_ar_minima + 0.0005 * umidade_instante - 0.0002 * umidade_maxima - 0.0011 * umidade_minima - 0.0033 * pto_orvalho_instante + 0.0988 * pressao - 0.0055 * vento_velocidade + 0.0034 * vento_rajada + 0.1804 * radiacao - 9.0589</p>	<p>EQ num: 6 EToPM = -0.0003 * umidade_solo_75cm + 0.0038 * temperatura_solo + 0.0082 * temperatura_ar_instante - 0.0075 * temperatura_ar_minima + 0.0005 * umidade_instante - 0.004 * umidade_maxima - 0.0005 * umidade_minima - 0.0033 * pto_orvalho_instante + 0.0535 * pressao + 0.0003 * vento_rajada + 0.2828 * radiacao - 5.058</p>

O resultado final é um modelo com alta correlação (0.9958) com os valores de referência de Penman-Monteith (Figura 27), em que os dados dos sensores de umidade e de temperatura do solo são utilizados em todas as regras, mostrando a sua relevância para o cálculo de ETo.

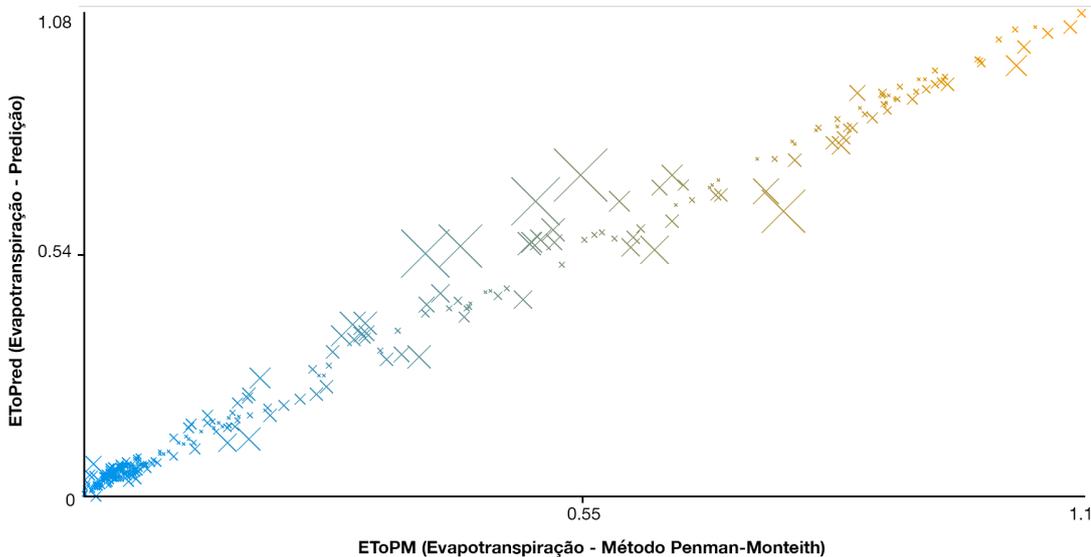


Figura 27 – WEKA: visualização do ETo previsto (y) e o ETo real (x)

5.4.4 Conclusão do Experimento II

Neste experimento, foram apresentados os resultados das decisões de nível alto das aplicações de evapotranspiração e de umidade do solo, além de uma análise da correlação dos valores dos sensores (em particular de temperatura e de umidade do solo) no cálculo de evapotranspiração, utilizando o método de Penman-Monteith como referência. Foi possível determinar um alto grau de correlação entre os valores obtidos pelo método gerado utilizando o algoritmo M5P e os valores alcançados pelo método Penman-Monteith.

5.5 Discussão do Experimento III: novo modelo de evapotranspiração

O terceiro e último experimento explora a possibilidade de utilizar os dados coletados para melhorar a precisão de um modelo de evapotranspiração existente, a ser utilizado no lugar do método Penman-Monteith. O método escolhido foi o Hargreaves-Samani, considerado adequado para climas semiáridos, que requer apenas dados de temperatura do ar, mas tem a tendência de superestimar ETo (SILVA *et al.*, 2015). Ao não depender de tantos sensores quanto o método Penman-Monteith, a utilização do novo modelo reduziria o custo de implementação e

manutenção do sistema.

Os dados fusionados de umidade de solo (obtidos pelos sensores em Paraipaba) e de temperatura do ar (obtidos pela estação de Itapipoca) foram carregados no MATLAB, no qual utilizamos a aplicação *Regression Learner* (aprendizado de regressão) para realizar a análise. Esta aplicação faz parte do módulo de análise estatística e de aprendizagem de máquina para realizar experimentos de aprendizado de regressão linear.

O *Regression Learner* tem funcionalidade similar à função *Classify* (classificação) do WEKA, disponibilizando diversos algoritmos diferentes de regressão linear, desde o método mais simples a árvores complexas e modelos gaussianos. Uma vantagem sobre o WEKA é a análise automática em todos os algoritmos disponíveis. Os resultados dos testes com os modelos de regressão estão listados na Tabela 8. O método que obteve o melhor desempenho foi o *Support Machine Vector* - Máquina de Suporte de Vetores (SVM) com kernel quadrático, resultando em um modelo que obteve o melhor índice de *Root Mean Square Error* - Raiz do erro quadrático médio (RMSE), de 0.79 (quanto mais próximo de zero mais acurado é o valor estimado).

Nas Figuras 28 e 29, são apresentados os dados reais do método Penman-Monteith e os valores estimados utilizando o novo modelo.

A Figura 30 apresenta uma comparação entre os valores de ETo pelo método de Hargreaves-Samani, de Penman-Monteith e pelo novo modelo. Confirmando a informação provida por Silva *et al.* (2015), o método Hargreaves-Samani superestimou o valor de ETo em comparação ao método Penman-Monteith (correlação de 0.8169), enquanto o novo modelo apresenta maior precisão (correlação de 0.9299).

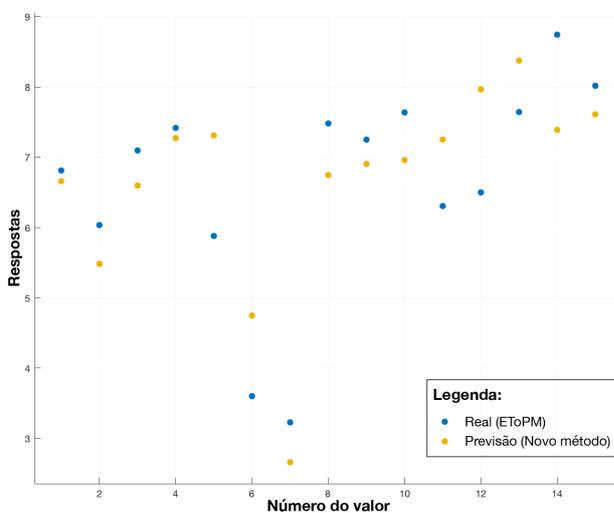


Figura 28 – MATLAB: gráfico de valores reais e estimados

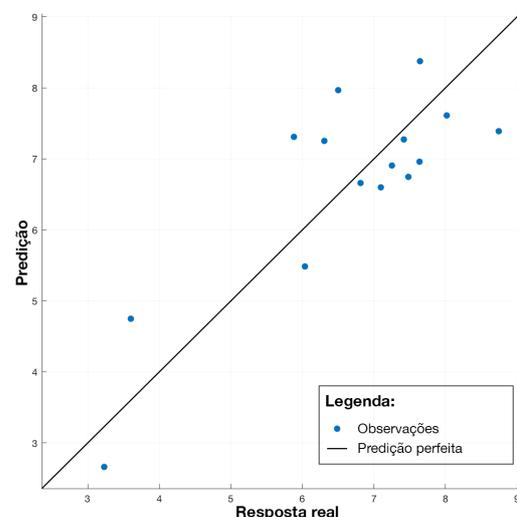


Figura 29 – MATLAB: gráfico de precisão da predição

Tabela 8 – Resultados dos modelos de regressão linear do MATLAB

Métodos	RMSE	R ²	MSE	MAE	Velocidade de predição (obs/seg)	Tempo de treino (seg)
Regressão Linear						
Regressão Linear	1,12	0,48	1,25	0,9	~91	5,3134
C/ Interações	6,58	-17,10	43,28	4,28	~130	0,92712
Robusta	1,11	0,49	1,22	0,89	~260	0,93718
Stepwise	1,06	0,53	1,13	0,93	~210	18,881
Árvore de Decisão						
Simples	1,55	-0,00	2,39	1,14	~640	0,22344
Média	1,55	-0,00	2,39	1,14	~840	0,21824
Complexa	1,78	-0,32	3,15	1,42	~290	3,8375
SVM						
Linear	1,12	0,48	1,25	0,88	~420	2,6551
Quadrático	0,79	0,74	0,63	0,69	~590	0,26581
Cúbico	0,85	0,70	0,72	0,72	~1000	0,17695
Gaussiano Fino	1,50	0,07	2,24	1,04	~820	0,19472
Gaussiano Médio	1,29	0,30	1,67	0,49	~1100	0,14973
Gaussiano Grossoiro	1,37	0,21	1,88	0,93	~620	0,17864
Agrupamento (Ensemble)						
<i>Boosted Trees</i>	1,57	-0,03	2,46	1,21	~260	1,5438
<i>Bagged Trees</i>	1,55	-0,00	2,39	1,13	~240	1,2767
Processo Gaussiano de Regressão (GPR)						
GPR Exponencial Quadrado	1,33	0,26	1,76	1,00	~440	1,3399
GPR Matern 5/2	1,17	0,42	1,38	0,91	~720	0,29289
GPR Exponencial	1,15	0,49	1,32	0,91	~800	0,28968
GPR Quadrático Racional	1,30	0,30	1,68	0,99	~860	0,2812

Fonte – O autor

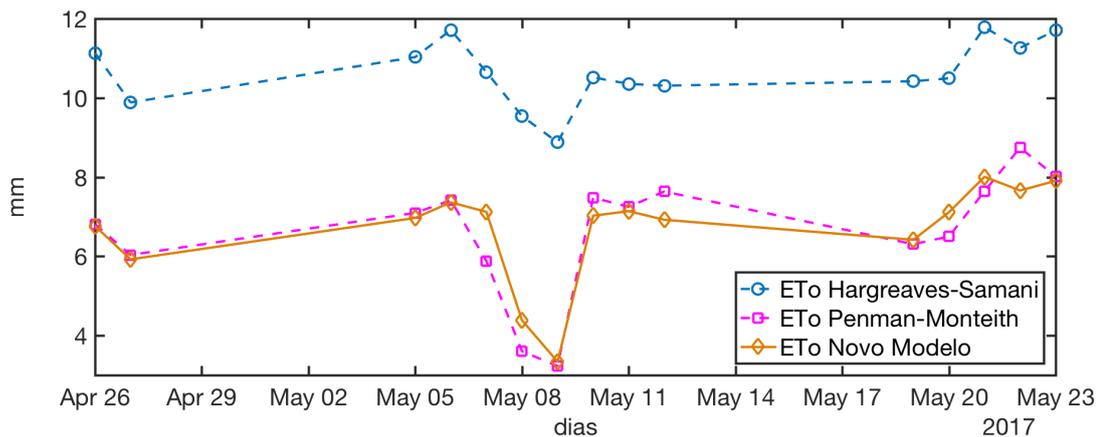


Figura 30 – Comparação entre os métodos de cálculo de ETo

5.5.1 Conclusão do Experimento III

Ao utilizar os dados fusionados de umidade de solo e os de temperatura do ar, foi possível gerar um novo modelo preditivo com acurácia maior do que o método Hargreaves-

Samani. Isto indica que, novamente, os dados coletados são válidos e que é viável a criação de uma aplicação de evapotranspiração com boa acurácia, mesmo sem a existência de uma estação meteorológica completa.

5.6 Conclusão do Capítulo

Neste capítulo, foram apresentados o material (ambiente, hardware, software, comunicação entre nós), a execução dos três experimentos, as coletas de dados e os resultados de cada experimento. O primeiro experimento focou na análise de técnicas de detecção de *outliers* e resultou na escolha do método mais adequado para os dados coletados, voltados para o domínio de agricultura. Os dois experimentos seguintes focaram na instanciação da arquitetura Hydra e na análise dos dados das aplicações de evapotranspiração e de umidade de solo, resultando em dois métodos para o cálculo de evapotranspiração, em que ambos são independentes do método Penman-Monteith e o segundo apresenta uma grande redução na quantidade de sensores requeridos para o funcionamento.

6 CONCLUSÕES E TRABALHOS FUTUROS

Foi apresentado neste trabalho uma arquitetura de fusão de dados multinível Hydra, com aplicação no domínio da Agricultura Inteligente, baseada em uma infraestrutura de baixo custo. Em parceria com a Embrapa, foi realizada uma implementação de teste em um campo experimental para monitoramento de dois cultivos: caju anão-precoce e coqueiros.

A utilização da arquitetura Hydra nas aplicações desenvolvidas resultou na confirmação da hipótese levantada de que uma arquitetura de fusão de dados multinível propicia uma melhoria da acurácia dos dados coletados, permitindo a detecção de eventos (chuva, por exemplo), e a tomada de decisões (como o momento de início de irrigação). O experimento I apresentou a melhoria da acurácia ao realizar a fusão de sensores de umidade de solo, enquanto o experimento II demonstrou as decisões obtidas com base nos eventos detectados. Já o experimento III resultou numa proposta de um novo modelo de cálculo de evapotranspiração.

A arquitetura Hydra pode ser aplicada em qualquer outro domínio, como o de ambientes inteligentes ou cidades inteligentes, e sua estrutura de fusão dos dados é desacoplada da infraestrutura, permitindo sua aplicação de maneira concentrada nos próprios nós sensores ou totalmente na nuvem, com os nós enviando apenas os dados brutos.

Em conjunto com a Hydra, foi desenvolvido um sistema IoT, utilizando o protocolo MQTT e com uma infraestrutura de RSSF baseada em Arduino¹ e no projeto Vinduino (LEE, 2017), que irá alimentar a arquitetura proposta com dados de sensores heterogêneos. Um dos benefícios de se adotar a plataforma Arduino é que, por se tratar de *hardware* aberto, o protótipo poderá eventualmente ser customizado e uma nova placa desenvolvida de acordo com as necessidades levantadas nesta pesquisa.

Sendo assim, eis as contribuições desta pesquisa:

- arquitetura de fusão de dados multinível para IoT com foco em:
 - flexibilidade na execução (desacoplado da infraestrutura);
 - tomada de decisões, com combinações de decisões;
- sistema IoT para agricultura inteligente;
- validação e avaliação da arquitetura.

¹ Arduino: <<http://www.arduino.cc>>

6.1 Artigos Publicados

TORRES, Andrei B. B.; ROCHA, Atslands Rego da; SOUZA, José N. de *Análise de Desempenho de Brokers MQTT em Sistema de Baixo Custo*. Anais do XV Workshop em Desempenho de Sistemas Computacionais e de Comunicação - WPerformance. SBC, Porto Alegre, RS, p. 2804-2815, 2016. (TORRES *et al.*, 2016)

TORRES, Andrei B. B.; FILHO, José A.; ROCHA, A. R. da; GONDIM, Rubens S.; SOUZA, José N. de *Outlier detection methods and sensor data fusion for precision agriculture*. IX Simpósio Brasileiro de Computação Ubíqua e Pervasiva - SBCUP. SBC, Porto Alegre, RS, 2017. (TORRES *et al.*, 2017)

6.2 Capítulo de Livro Publicado

ROCHA, Atslands Rego da; SOUZA, Flavio Rubens de Carvalho; TORRES, Andrei Bosco B.; SOUZA, José Neuman de. *Internet of Things: Architectures, Technologies and Applications*. In: Silvia Watts. (Org.). *The Internet of Things (IoT): Applications, Technology, and Privacy Issues*. 1ed. Nova Publishers, Hauppauge, NY, p. 1-21, 2016. (ROCHA *et al.*, 2012)

6.3 Dificuldades Encontradas

Durante o desenvolvimento desta dissertação diversas dificuldades e barreiras foram encontradas, podendo-se citar:

- superaquecimento e dano nas pilhas que alimentavam os nós;
- bloqueio do sinal do módulo de rádio por causa da vegetação (problema contornado ajustando a altura das antenas);
- consumo de energia do Raspberry Zero (apesar de ser o modelo de menor consumo);
- falta de cobertura celular na região, que impediu o uso de módulo GSM (*Global System for Mobile Communications* - Sistema Global para Comunicações Móveis);
- coincidência do período de chuva com o das coletas de dados.

6.4 Trabalhos Futuros

Os trabalhos futuros baseados neste projeto podem ser divididos entre melhorias de hardware e de software. Quanto ao hardware, é possível a realização da captação de energia solar para alimentação dos módulos, cortando a dependência de pilhas e a necessidade de manutenção constante; criação de uma placa customizada para remover a dependência do Arduino, utilizando diretamente o microcontrolador ATmega328 ou equivalente e controlando melhor o consumo energético dos componentes; e a avaliação de outros tipos de sensores de baixo custo, possibilitando inclusive a criação de miniestações meteorológicas.

Já quanto ao software, seria interessante implementar uma rede *mesh* para facilitar a inserção/remoção de módulos, criar uma *Application programming interface* - Interface de Programação de Aplicativos (API) que permita a comunicação da arquitetura Hydra com outras plataformas ou interfaces e também melhorar a implementação da camada de gerência, com a criação de uma interface flexível para que o usuário final tenha acesso fácil aos dados e à criação de regras e parâmetros.

E, finalmente, também é interessante efetuar coleta de dados por períodos mais longos. Os dados coletados já apresentaram alguma correlação com a fórmula de evapotranspiração de Penman-Monteith, porém é necessário aprofundar este estudo e também analisar outros sensores de umidade de solo.

REFERÊNCIAS

- ABDELGAWAD, A.; BAYOUMI, M. **Resource-Aware Data Fusion Algorithms for Wireless Sensor Networks**. 1. ed. Boston, MA: Springer US, 2012. v. 118. (Lecture Notes in Electrical Engineering, v. 118). ISBN 978-1-4614-1350-9. Disponível em: <<http://link.springer.com/10.1007/978-1-4614-1350-9>>.
- ALLEN, R. G.; PEREIRA, L.; RAES, D.; SMITH, M. **Crop evapotranspiration: Guidelines for computing crop water requirements** - FAO Irrigation and drainage paper 56. 1998. Disponível em: <<http://www.fao.org/docrep/XO490E/XO490E00.htm>>. Acesso em: 01 jun. 2017.
- Amazon Web Services. **AWS IoT**. 2016. Disponível em: <<https://aws.amazon.com/pt/iot/>>. Acesso em: 25 jun. 2017.
- ANDRADE, J. Pesquisadores preveem agravamento da seca no nordeste entre fevereiro e abril. **Agência Brasil**, 26 jan 2017. Disponível em: <<http://agenciabrasil.ebc.com.br/geral/noticia/2017-01/pesquisadores-preveem-agravamento-da-seca-no-nordeste-entre-fevereiro-e-abril>>. Acesso em: 03 jun. 2017.
- ASHTON, K. **That 'Internet of Things' Thing**. 2009. Disponível em: <<http://www.rfidjournal.com/articles/view?4986>>.
- ATZORI, L.; IERA, A.; MORABITO, G. The Internet of Things: A survey. **Computer Networks**, Elsevier B.V., v. 54, n. 15, p. 2787–2805, 2010. ISSN 13891286. Disponível em: <<http://dx.doi.org/10.1016/j.comnet.2010.05.010>>.
- BERNARDI, A. C. d. C.; NAIME, J. d. M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: resultados de um novo olhar**. Brasília, DF: Embrapa, 2014. 1–600 p. ISBN 9788570353528. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/114434/1/Agricultura-de-precisao.pdf>>.
- BISH, S.; ROHRER, M.; SCHEFFEL, P.; BENNETT, K. Multi-sensor fusion development. **SPIE Vol. 9831**, v. 9831, p. 98310T, 2016. ISSN 1996756X. Disponível em: <<http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2224138>>.
- BOSTRÖM, H.; ANDLER, S. F.; BROHEDE, M.; JOHANSSON, R.; KARLSSON, A.; LAERE, J. V.; NIKLASSON, L.; NILSSON, M.; PERSSON, A.; ZIEMKE, T. On the Definition of Information Fusion as a Field of Research. **IKI Technical Reports**, n. October, p. 1–8, 2007. Disponível em: <<http://www.isif.org/sites/isif.org/files/FULLTEXT01.pdf>>.
- BRAGA, L. Doze cidades cearenses só têm água garantida até julho. **OPOVO Online**, 30 jun 2017. Disponível em: <<http://www.opovo.com.br/jornal/cotidiano/2017/06/doze-cidades-cearenses-so-tem-agua-garantida-ate-julho.html>>. Acesso em: 24 jun. 2017.
- BRASIL. Portaria nº 852 - Art. 1º Criar a Comissão Brasileira de Agricultura de Precisão – CBAP. **Diário Oficial [da] República Federativa do Brasil**, Ministério da Agricultura, Pecuária e Abastecimento, Brasília, DF, 21 set. 2012. Seção 1, n. 184. Disponível em: <http://www.cnpt.embrapa.br/biblio/do/p_do42.htm>. Acesso em: 25 jun. 2017.
- CALLEGARO, R.; MONTEZ, C.; PINTO, A. R.; MORAES, R. Uma Arquitetura para Fusão de Dados e Detecção de Outliers em Sensores de Baixo Custo de Redes de Sensores sem Fio. **Anais do II Workshop de Comunicação em Sistemas Embarcados Críticos - WoCCES**, p. 3–16, 2014.

DASARATHY, B. V. Sensor fusion potential exploitation-innovative architectures and illustrative applications. **Proceedings of the IEEE**, v. 85, n. 1, p. 24–38, 1997. ISSN 00189219.

De Paola, A.; FERRARO, P.; GAGLIO, S.; Lo Re, G.; DAS, S. An Adaptive Bayesian System for Context-Aware Data Fusion in Smart Environments. **IEEE Transactions on Mobile Computing**, p. 1–1, 2016. ISSN 1536-1233. Disponível em: <<http://ieeexplore.ieee.org/document/7539530/>>.

ELMENREICH, W. **Sensor fusion in time-triggered systems**. Tese (Doutorado) — Institut für Technische Informatik, Vienna University of Technology, Áustria, 2002.

G1 Economia. **Agronegócio cresce com ajuda da tecnologia**. 2017. Disponível em: <<http://g1.globo.com/economia/agronegocios/agro-a-industria-riqueza-do-brasil/noticia/agronegocio-cresce-com-ajuda-da-tecnologia.ghtml>>. Acesso em: 02 ago. 2017.

GEBBERS, R.; ADAMCHUK, V. I. Precision Agriculture and Food Security. **Science**, v. 327, n. 5967, p. 828–831, 2010. ISSN 0036-8075. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20150492>>.

IBM. **IBM Bluemix Docs: MQTT messaging**. 2017. Disponível em: <<https://console-regional.ng.bluemix.net/docs/services/IoT/reference/mqtt/index.html>>. Acesso em: 25 jun. 2017.

IYENGAR, S. S.; CHAKRABARATY, K.; QI, H. Introduction to special issue on “distributed sensor networks for real-time systems with adaptive configuration”. **Journal of the Franklin Institute**, v. 338, p. 651–653, 2001.

LEE, R. V. D. **Vinduino**: Open license project for agricultural irrigation management. 2017. Disponível em: <<https://github.com/ReiniervdL/Vinduino>>. Acesso em: 01 jun. 2017.

LEONARD, E. Precision agriculture. In: WRIGLEY, C.; CORKE, H.; SEETHARAMAN, K.; FAUBION, J. (Ed.). **Encyclopedia of Food Grains**. Second edition. Oxford: Academic Press, 2016. p. 162 – 167. ISBN 978-0-12-394786-4. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780123944375002035>>.

MELO, U. T.; TREINISH, L. A. **Precision agriculture**: Using predictive weather analytics to feed future generations. 2012. Disponível em: <http://www.research.ibm.com/articles/precision_agriculture.shtml>. Acesso em: 06 jun. 2017.

Microsoft. **Communicate with your IoT hub using the MQTT protocol**. 2017. Disponível em: <<https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-mqtt-support>>. Acesso em: 25 jun. 2017.

MOLIN, G. D. **Na contramão da crise, agronegócio deve puxar PIB brasileiro**. Gazeta do Povo, 2017. Disponível em: <<http://www.gazetadopovo.com.br/agronegocio/agricultura/na-contramao-da-crise-agronegocio-deve-puxar-pib-brasileiro-120108yf1mom1oaf8vua57y9f>>. Acesso em: 02 ago. 2017.

NAKAMURA, E. F.; LOUREIRO, A. a. F.; FRERY, A. C. Information fusion for wireless sensor networks. **ACM Computing Surveys**, v. 39, n. 3, 2007. ISSN 03600300. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1267070.1267073>>.

OASIS STANDARD. **MQTT Version 3.1.1 Plus Errata 01**. [S.l.], 2015. Disponível em: <<http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html>>. Acesso em: 24 jun. 2017.

- OBERSTEIN, T. G.; GOEDDE, A. **The Web Application Messaging Protocol**. [S.l.], 2017. <http://wamp-proto.org/static/rfc/draft-oberstet-hybi-crossbar-wamp.html>. Acesso em: 25 jun. 2017.
- PINA, R. Seca no nordeste corre risco de bater recorde e completar seis anos em 2017. **Brasil de Fato**, 16 dez 2016. Disponível em: <<https://www.brasildefato.com.br/2016/12/16/seca-no-nordeste-corre-risco-de-bater-recorde-e-completar-seis-anos-em-2017/>>. Acesso em: 05 jun. 2017.
- RAWAT, P.; SINGH, K. D.; CHAOUCHI, H.; BONNIN, J. M. Wireless sensor networks: A survey on recent developments and potential synergies. **Journal of Supercomputing**, v. 68, n. 1, p. 1–48, 2014. ISSN 15730484.
- ROCHA, A. R. da; SOUZA, F. R. de C.; TORRES, A. B. B.; SOUZA, J. N. de. Internet of Things: Architectures, Technologies and Applications. In: WATTS, S. (Ed.). **The Internet of Things (IoT): Applications, technology, and privacy issues**. 1. ed. Hauppauge, NY: Nova Publishers, 2012. p. 1–21.
- SAMANI, Z. Estimating solar radiation and evapotranspiration using minimum climatological data. **Journal of Irrigation and Drainage Engineering**, American Society of Civil Engineers, v. 126, n. 4, 2000.
- SANCHEZ, L.; MUÑOZ, L.; GALACHE, J. A.; SOTRES, P.; SANTANA, J. R.; GUTIERREZ, V.; RAMDHANY, R.; GLUHAK, A.; KRICO, S.; THEODORIDIS, E.; PFISTERER, D. SmartSantander: IoT experimentation over a smart city testbed. **Computer Networks**, Elsevier, v. 61, p. 217–238, mar 2013. ISSN 13891286. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-84897112118{&}partnerID=tZOtx>>.
- SEMATECH, N. e-Handbook of Statistical Methods. **Engineering Statistics Handbook**, p. 633, 2003. Disponível em: <<http://www.itl.nist.gov/div898/handbook/>>.
- SHELBY, Z.; HARTKE, K.; BORMANN, C. **The Constrained Application Protocol (CoAP)**. [S.l.], 2014. Disponível em: <<https://tools.ietf.org/html/rfc7252>>. Acesso em: 25 jun. 2017.
- SHOCK, C. C.; BARNUM, J. M.; SEDDIGH, M. Calibration of watermark soil moisture sensors for irrigation management. In: **Proceedings of the 1998 Annual Meeting of the Irrigation Association**. [S.l.: s.n.], 1998. p. 139–146.
- SILVA, M. G. da; OLIVEIRA, I. d. S.; CARMO, F. F.; LÊDO, E. R. F.; da Silva Filho, J. A. Estimativa da Evapotranspiração de Referência pela Equação de Hargreaves-Samni no Estado do Ceará, Brasil. **Brazilian Journal of Biosystems Engineering**, v. 9, n. 2, p. 132–141, 2015.
- SINGH, D.; TRIPATHI, G.; JARA, A. J. A survey of Internet-of-Things: Future vision, architecture, challenges and services. **2014 IEEE World Forum on Internet of Things, WF-IoT 2014**, p. 287–292, 2014.
- SKERRETT, I. **Case Study MQTT: Why Open Source and Open Standards Drive Adoption**. 2015. Disponível em: <<https://ianskerrett.wordpress.com/2015/03/04/case-study-mqtt-why-open-source-and-open-standards-drives-adoption/>>. Acesso em: 25 jun. 2017.
- THORNTHWAITE, C. W. An Approach toward a Rational Classification of Climate. **Geographical Review**, v. 38, n. 1, p. 55–94, 1948.

- 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. **IX Simpósio Brasileiro de Computação Ubíqua e Pervasiva - SBCUP**, SBC, Porto Alegre, RS, 2017.
- TORRES, A. B. B.; ROCHA, A. R. da; SOUZA, J. N. de. Análise de desempenho de brokers MQTT em sistema de baixo custo. **Anais do XV Workshop em Desempenho de Sistemas Computacionais e de Comunicação - WPerformance**, SBC, Porto Alegre, RS, p. 2804–2815, 2016. Disponível em: <<http://ebooks.pucrs.br/edipucrs/anais/csbc/#/evento/15wperformance>>.
- U.S. DEPARTMENT OF DEFENSE. Data Fusion Lexicon. In: **Tecnical Panel for C3: F.E. White**, Code 4202. San Diego: Data Fusion Sub-Panel, Naval Ocean Systems Center, 1991. Disponível em: <<http://www.dtic.mil/dtic/tr/fulltext/u2/a529661.pdf>>.
- WALD, L. Some terms of reference in data fusion. **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 3 I, p. 1190–1193, 1999. ISSN 01962892.
- WALTZ, E. L.; LLINAS, J. **Multisensor Data Fusion**. Norwood, MA, USA: Artech House, Inc., 1990. ISBN 0890062773.
- WANG, J.; GAO, Y.; RAN, C.; HUO, Y. State estimation with two-level fusion structure. In: **2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF)**. [S.l.: s.n.], 2015. p. 105–109.
- WICHIT, N. Multisensor data fusion model for activity detection. **12th International Conference on ICT and Knowledge Engineering**, 2014.
- YINBIAO, S.; LEE, K.; LANCTOT, P.; JUANBIN, F.; HAO, H.; CHOW, B.; DESBENOIT, J.-P.; STEPHAN, G.; HUI, L.; GUODONG, X.; CHEN, S.; FAULK, D.; KAISER, T.; SATOH, H.; JINSONG, O.; SHOU, W.; YAN, Z.; JUNPING, S.; HAIBIN, Y.; PENG, Z.; DONG, L.; QUI, W. Internet of Things: Wireless Sensor Networks. **Internation Electronic Commision**, n. December, p. 1–78, 2014. Disponível em: <<http://www.iec.ch/whitepaper/pdf/iecWP-internetofthings-LR-en.pdf>>.
- ZHANG, L. **Building Facebook Messenger**. 2011. Disponível em: <<https://www.facebook.com/notes/facebook-engineering/building-facebook-messenger/10150259350998920>>. Acesso em: 25 jun. 2017.