



Universidade Federal
de Campina Grande



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

HÉLVIO RUBENS REIS DE ALBUQUERQUE

**MODELAGEM MATEMÁTICA DA CURVA DE CALIBRAÇÃO DO SENSOR DE
VAZÃO YF-S201 APLICADA AO CONTEXTO DA INDÚSTRIA 4.0**

Campina Grande, Paraíba
2018

HÉLVIO RUBENS REIS DE ALBUQUERQUE

MODELAGEM MATEMÁTICA DA CURVA DE CALIBRAÇÃO DO SENSOR DE VAZÃO YF-S201
APLICADA AO CONTEXTO DA INDÚSTRIA 4.0

*Trabalho de Conclusão de Curso submetido à
Unidade Acadêmica de Engenharia Elétrica da
Universidade Federal de Campina Grande,
como parte dos requisitos necessários para
obtenção do grau de Bacharel em Ciências no
domínio da Engenharia Elétrica.*

Área de Concentração: Processamento da Informação

Orientador:

Professor Jalberth Fernandes de Araujo, D. Sc.

Campina Grande, Paraíba
2018

HÉLVIO RUBENS REIS DE ALBUQUERQUE

MODELAGEM MATEMÁTICA DA CURVA DE CALIBRAÇÃO DO SENSOR DE VAZÃO YF-S201
APLICADA AO CONTEXTO DA INDÚSTRIA 4.0

*Trabalho de Conclusão de Curso submetido à
Unidade Acadêmica de Engenharia Elétrica da
Universidade Federal de Campina Grande,
como parte dos requisitos necessários para
obtenção do grau de Bacharel em Ciências no
domínio da Engenharia Elétrica.*

Área de Concentração: Processamento da Informação

Aprovado em: ____ / ____ / _____

Professor Ronimack Trajano de Souza, D. Sc.

Universidade Federal de Campina Grande
Avaliador, UFCG

Professor Jalberth Fernandes de Araujo, D. Sc.

Universidade Federal de Campina Grande
Orientador, UFCG

Campina Grande, Paraíba
2018

AGRADECIMENTOS

Primeiramente, gostaria de agradecer aos meus pais, Jane e Rubens. Em todos os desafios, batalhas e vitórias da minha vida, vocês estavam sempre ao meu lado. Obrigado por todo o amor, renúncias e dificuldades que vocês enfrentaram para que eu pudesse estar aqui hoje. Nada foi em vão!

Agradeço ao professor Jalberth Fernandes por acreditar que era possível realizar este trabalho, mesmo quando eu achava o contrário. Aprendi muito com suas dicas, conversas e estratégias traçadas a cada novo passo que eu dava em direção à conclusão desse projeto. Obrigado por toda a transformação, aprendizado e orientação. Isso é pancada!

Agradeço ao professor Raimundo Freire, por todo o apoio demonstrado, pelos ensinamentos e oportunidades que me fizeram crescer pessoal e profissionalmente durante a graduação. Obrigado, mestre.

Agradeço também às professoras Andrea e Danielle, por me ajudarem a entender um pouco mais sobre os sistemas de abastecimento, fluidos e como poderia integrar este tema em um novo contexto. Muito obrigado por me ajudarem a entender um pouco mais sobre esse mundo fascinante.

Não posso deixar de agradecer ao Vicente Santos, que me ajudou substancialmente durante esse trabalho, ajudando a isolar vazamentos, encontrar bugs e por não desanimar, mesmo quando tudo dava errado.

Aos meus amigos mais antigos, meus mais sinceros agradecimentos, por todas as conversas, risadas, noites mal dormidas e por toda a amizade demonstrada durante todos esses anos. Obrigado Ewerton Brasil, Bianca Cartaxo, Helton Tavares, Ivana Pires, Welch Martiniano, Ana Carolina Medeiros e Sarah Albuquerque. Sem vocês, todo o caminho teria sido mais doloroso. Aos amigos mais novos, agradeço por toda a paciência que vocês demonstraram ter durante as últimas semanas. Não teria chegado até aqui sem Ianca Rocha, Izadora Soares, Ariadne Guedes. Obrigado por todas as vezes que me acolheram com uma palavra, com um gesto ou simplesmente dividindo uma pizza em uma semana complicada. Agradeço também a Mônica Willany por ser tão prestativa, autêntica e por saber ouvir mesmo quando eu não conseguia dizer nada. Vocês fazem parte da minha história e da minha vida.

Aos demais amigos, colegas e familiares, cada pessoa foi essencial para a formação de quem eu sou hoje. A vocês, muitíssimo obrigado!

RESUMO

Neste trabalho foi desenvolvido um modelo matemático para um sensor de vazão no contexto da Indústria 4.0. A modelagem do sensor foi essencial para determinar a sua curva de calibração, estabelecendo a relação entre as variáveis de interesse, volume e número de pulsos. O modelo foi determinado a partir do uso do método dos mínimos quadrados. A aplicação desse método permite obter a estimativa dos parâmetros da reta de regressão linear simples, de forma que o erro associado ao estimador seja o menor possível. Assim, possibilita-se ao microcontrolador ESP8266 calcular os dados com maior exatidão. As informações foram processadas pelo NodeMCU e transmitidas via internet para uma aplicação *online* baseada na computação em nuvem, o *Thingspeak*[®]. A ideia empregada neste trabalho, o sensor e o modelo matemático podem ser aplicados ao setor de *utilities*, especificamente no abastecimento de água associado ao contexto da Quarta Revolução Industrial ou Indústria 4.0. Como o NodeMCU é um dispositivo de Internet das Coisas, foi possível integrá-los a outros sistemas de informação e possibilitar a gestores e consumidores, acompanharem o consumo de água remotamente por meio da internet.

Palavras-chave: mínimos quadrados, regressão linear, *utilities*, indústria 4.0, internet das coisas.

ABSTRACT

In this paper a mathematical model was developed for a flow sensor in the context of Industry 4.0. The modeling of the sensor is essential to determine its calibration curve, establishing the relation between the variables of interest, volume and number of pulses. The model was determined using the least squares method. The application of this method allows to obtain the estimation of the parameters of the line of simple linear regression, so that the error associated with the estimator is the smallest possible. Thus, it is possible for the ESP8266 microcontroller to calculate the data with greater accuracy. The information was processed by NodeMCU and transmitted over the internet to a cloud-based online application, Thingspeak®. The idea used in this paper, the sensor and the mathematical model can be applied to the utilities sector, specifically in the water supply associated with the context of the Fourth Industrial Revolution or Industry 4.0. As NodeMCU is a Internet of Thing device, it is possible to integrate them with other information systems and to enable managers and consumers to monitor water consumption remotely through the internet.

Keywords: *least square method, linear regression, utilities, industry 4.0, internet of things.*

LISTA DE ILUSTRAÇÕES

Figura 1: Representação vetorial da decomposição de uma força que atua sobre uma superfície.	3
Figura 2: Aplicação de um jato de tinta em um tubo com fluido apresentando (a) regime laminar e (b) regime turbulento.	4
Figura 3: Análise gráfica da região de entrada, para uma variação da tensão de cisalhamento em função do comprimento do tubo.	5
Figura 4: Classificação dos medidores de vazão para condutos forçados.	6
Figura 5: (a) Sensor YF-S201 e o (b) princípio de funcionamento.	8
Figura 6: Histograma de resíduos padronizados, indicando a presença de um outlier além de -4.	15
Figura 7: Microcontrolador ESP8266, da Espressif.	17
Figura 8: Módulo ESP8266 NodeMCU.	18
Figura 9: Procedimento de modelagem estatística para determinação de uma nova curva de calibração.	20
Figura 10: Esquema do sistema de calibração de sensor de vazão.	21
Figura 11: Aquisição de 10 amostras para diversos volumes observados do sensor de vazão YF-S201.	24
Figura 12: Valores para o teste de Grubbs para as amostras de pulsos do sensor de vazão YF-S201.	25
Figura 13: Gráfico da reta de regressão linear pelos Mínimos Quadrados.	26
Figura 14: Histograma dos resíduos padronizados, para a expressão modificada.	27
Figura 15: Janelas de dados no Thingspeak®, indicando a localização atual e o consumo.	28

LISTA DE TABELAS

Tabela 1: Valores críticos, T_c , para distintos níveis de significância e número de amostras.....	12
Tabela 2: Comparativo entre as placas de desenvolvimento da Espressif e Arduino.	18
Tabela 3: Especificações técnicas do medidor de vazão Saga® US-3,0.....	19
Tabela 4: Comprimentos mínimos para cada seção do sistema hidráulico.	21
Tabela 5: Indicadores de erro de predição para a equação original e equação modificada da reta de mínimos quadrados do sensor YF-S201.....	27

SUMÁRIO

1	Introdução	1
2	Fundamentação Teórica	2
2.1	Hidrodinâmica	2
2.1.1	Fluidos	2
2.1.2	Tensão de Cisalhamento	2
2.1.3	Escoamento em Tubos	3
2.1.4	Regimes de Escoamento	3
2.1.5	Região de Entrada	4
2.1.6	Vazão	5
2.2	Medidores de Vazão	6
2.3	Sensor de Vazão YF-S201	7
2.4	Incerteza de Medição	8
2.5	Análise Estatística de Dados	9
2.5.1	Medidas Estatísticas Descritivas	9
2.5.2	Medidas de Dados Bivariados	10
2.5.3	Valores Extremos	11
2.6	Regressão Linear Simples	12
2.7	Mínimos Quadrados	13
2.8	Indicadores de Erros de Predição	14
2.9	Indústria 4.0	15
2.10	Computação em Nuvem com a <i>Thingspeak</i> ®	16
2.11	Projetos de IoT com o ESP8266	17
3	Metodologia	17
3.1	ESP8266 NodeMCU	18
3.2	Medidor e Sensor de Vazão	19
3.3	Sistema de Calibração do Sensor	20
3.3.1	Parâmetros do Sistema Hidráulico	20
3.3.2	Aquisição de dados do Sensor de Vazão	21
3.3.3	Diagrama de Dispersão e Teste de Grubbs	22
3.3.4	Coefficiente de Correlação e Mínimos Quadrados	23
3.3.5	Validação do Sistema de Calibração	23
3.4	Sistema Integrado à Nuvem	23
4	Resultados e Discussão	24
4.1	Coleta e Pré-Tratamento de Dados	24
4.2	Análise Estatística para a Regressão Linear	25
4.3	Computação em Nuvem com <i>Thingspeak</i> ®	27
5	Conclusão	29
	Referências	30

1 INTRODUÇÃO

A Quarta Revolução Industrial, ou Indústria 4.0, é um processo de transformação cada vez mais presente no contexto industrial, caracterizada pela presença de ambientes colaborativos cada vez mais integrados a processos físicos, por meio de elementos computacionais integrados. Essa característica possibilitou o surgimento das chamadas tecnologias habilitadoras, como por exemplo, a Internet das Coisas (IoT), *big data*, *analytics* e computação em nuvem (SCHWAB, 2016).

No entanto, o setor industrial de *utilities*, cujo produto é essencial e crítico para a sociedade, como as empresas de energia, gás, água e transportes (MCNABB, 2005), ainda apresenta resistência à integração ao novo modelo industrial. (BOOTH, MOHR e PETERS, 2016). Esse comportamento protecionista, no entanto, pode conduzir a uma baixa eficiência no processo de gerência e planejamento das empresas, como foi constatado em um diagnóstico sobre o sistema de abastecimento de água (SAA) na cidade de Campina Grande – PB (MENESES, 2011). Em sua pesquisa, Meneses (2011) afirmou que um SAA pode apresentar diversos problemas, como identificação da capacidade atual dos reservatórios de distribuição, consumo de água em tempo real e pouco ou nenhum conhecimento sobre o perfil do usuário final.

Apesar desta pesquisa encontrar-se atualmente desatualizada, visto que a CAGEPA iniciou em 2012 um processo de automatização dos seus processos, é proposto neste trabalho um sistema de medição conectado à internet, com o objetivo de monitorar o consumo de água remotamente por meio da comunicação *Wi-Fi*. Contudo, o sensor de vazão YF-S201 apresenta forte dependência do tempo, visto que a sua curva de calibração é dada em frequência proporcional à vazão. Como o microcontrolador utilizado também depende do tempo, para realizar o processamento dos dados, atrasos sucessivos podem contribuir para uma medição incoerente da vazão. Assim, foi realizado um estudo acerca da calibração do sensor por meio da medição do volume, utilizando a regressão linear em conjunto com o método dos mínimos quadrados. Para obter a vazão, basta dividir o volume acumulado pelo tempo.

Associado a isso, uma plataforma online de desenvolvimento de aplicações em IoT, a *Thingspeak*[®], foi utilizada para exibir os dados coletados e revelar aos usuários o seu perfil de consumo, assim como introduzir mais uma ferramenta de análise aos gestores, para que decisões institucionais cada vez mais assertivas sejam tomadas.

2 FUNDAMENTAÇÃO TEÓRICA

Serão discutidos nesta seção, a teoria sobre vazão volumétrica e alguns conceitos importantes na hidráulica, que possibilitem gerar uma curva de calibração para o sensor a ser utilizado neste trabalho. Também serão apresentadas ferramentas matemáticas para determinação dos principais parâmetros inerentes ao processo de calibração e aquisição dos dados.

2.1 HIDRODINÂMICA

2.1.1 FLUIDOS

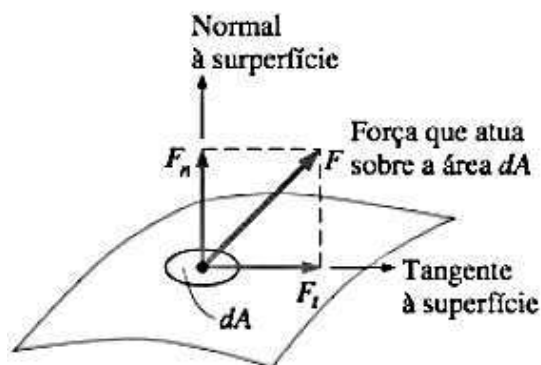
A matéria possui três estados fundamentais: sólido, líquido e gasoso. Os sólidos possuem um arranjo de moléculas que se repete em um padrão por todo o corpo, devido às forças atrativas que as mantêm fixas. Nos líquidos, as forças intermoleculares são mais fracas e grupos de moléculas se movem uns em relação aos outros, mantendo o volume constante. Essa característica permite que o líquido preencha o recipiente que o contém. Um gás, no entanto, expande-se até encontrar as paredes do recipiente, preenchendo todo o espaço disponível, enquanto que as suas moléculas apresentam um movimento completamente desordenado e aleatório (ÇENGEL e CIMBALA, 2015).

Devido ao arranjo das moléculas dos líquidos e gases não serem fixos, são também denominados de fluidos.

2.1.2 TENSÃO DE CISALHAMENTO

Tensão mecânica é definida como o valor da distribuição de forças por unidade de área e é obtida pelo quociente entre a força e a área sobre a qual ela atua (ÇENGEL e CIMBALA, 2015). Decompondo essa força em componentes normal e tangencial à superfície, como esboçado na Figura 1, obtém-se respectivamente a tensão normal e tensão de cisalhamento ou tensão tangencial.

Figura 1: Representação vetorial da decomposição de uma força que atua sobre uma superfície.



Fonte: (ÇENGEL e CIMBALA, 2015).

A principal característica dos fluidos é a capacidade da substância em se deformar continuamente sobre a atuação de uma tensão de cisalhamento. Para um fluido em repouso, a tensão normal é conhecida como pressão e a tensão de cisalhamento é nula, pois as paredes do recipiente impedem o surgimento de forças tangenciais. Porém, quando o recipiente é inclinado ou uma de suas paredes é removida, as forças tangenciais fazem com que o líquido se mova para manter a superfície da horizontal (ÇENGEL e CIMBALA, 2015).

2.1.3 ESCOAMENTO EM TUBOS

O escoamento de líquido ou gás por meio de tubos ou dutos é usado em diversas aplicações de aquecimento ou resfriamento, além das redes de distribuição de água e coleta de esgotos. O fluido pode ser forçado a percorrer esse caminho por meio de uma diferença de pressão, ou por meio de uma bomba hidráulica.

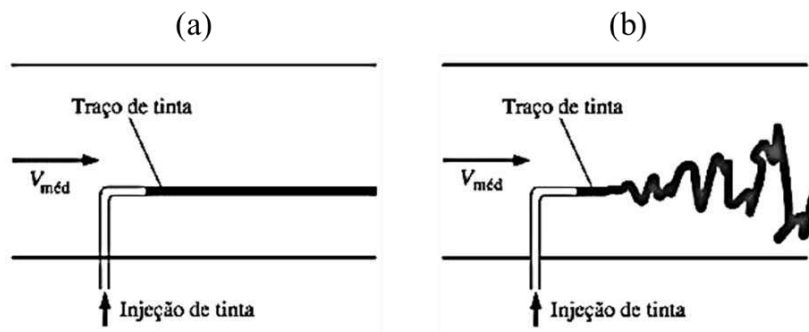
Çingel e Cimballa (2015) definem tubo, como seções de escoamento de seção transversal circular e as seções de escoamento de seção transversal não circular de dutos. Os tubos são usados majoritariamente para líquidos, pois a seção transversal circular pode suportar grandes diferenças de pressão entre interior e exterior, enquanto que os dutos são mais utilizados para conduzirem gases.

2.1.4 REGIMES DE ESCOAMENTO

Um fluido é considerado aerodinâmico para uma dada velocidade média ($V_{méd}$), mas pode se tornar completamente caótico, caso esta velocidade ultrapasse um valor crítico. Na Figura 2 (a) mostrou-se um exemplo prático, em que um jato de tinta foi

lançado no interior de um tubo contendo um fluido. Observa-se que o jato de tinta possui um comportamento suave e ordenado, caracterizado pelo regime laminar. Para o jato de tinta da Figura 2 (b), no entanto, o fluido torna-se completamente desordenado e com flutuações, caracterizado pelo regime turbulento.

Figura 2: Aplicação de um jato de tinta em um tubo com fluido apresentando (a) regime laminar e (b) regime turbulento.



Fonte: (ÇENGEL e CIMBALA, 2015).

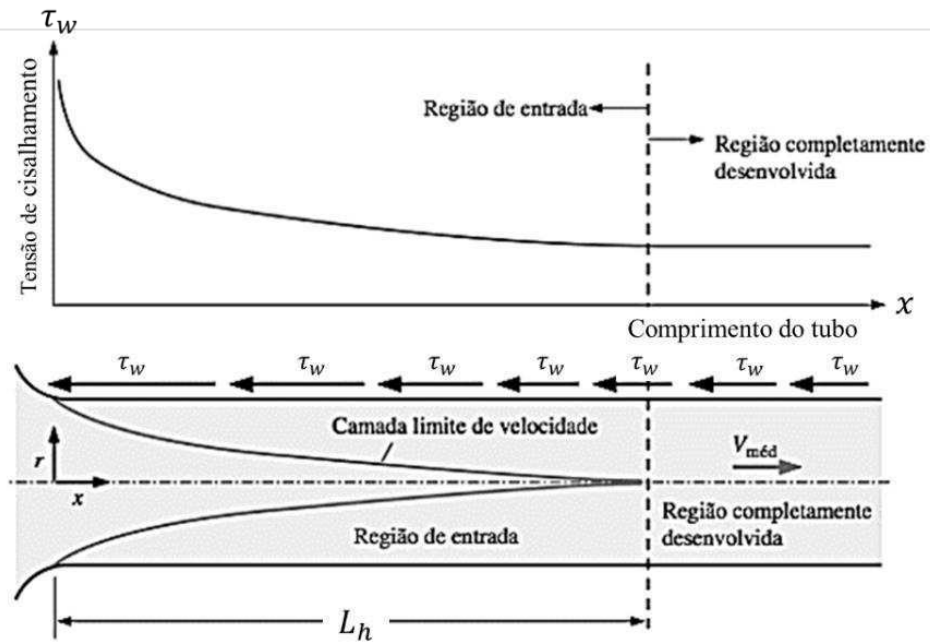
O movimento turbulento ocorre sobretudo devido a perturbações ou agitações presentes no fluido, em que as quantidades envolvidas apresentam variações estocásticas no tempo e espaço. Essa flutuação rápida implica no aumento da transferência de quantidade de movimento entre as partículas, ocasionando uma maior força de atrito com as paredes do tubo, aumentando assim, a potência de bombeamento necessária para deslocar o fluido (ÇENGEL e CIMBALA, 2015).

2.1.5 REGIÃO DE ENTRADA

Çengel e Cimbala (2015) afirmam que uma característica importante para um fluido líquido atravessando um tubo circular, é que sua velocidade seja uniforme em todos os pontos observados, para que o movimento seja considerado permanente e constante. Todavia, em virtude da condição de não escorregamento, as partículas em contato com a superfície interna do tubo apresentam uma velocidade menor do que na parte mais central do fluido. Isso é devido à presença do fator de atrito entre o tubo e o líquido, conhecido como tensão de cisalhamento (τ_w).

Para resolver esse problema, é necessário que o comprimento do tubo (x) seja tal que o fluido possa desenvolver um gradiente de velocidade adequado, região em que sua velocidade média ($V_{méd}$) pode ser considerada constante. Esse comportamento está descrito na Figura 3:

Figura 3: Análise gráfica da região de entrada, para uma variação da tensão de cisalhamento em função do comprimento do tubo.



Fonte: Adaptado de (ÇENGEL e CIMBALA, 2015).

Por meio do gráfico da Figura 3, percebe-se que existe um comprimento ótimo, em que a velocidade média do fluido está completamente desenvolvida. Esse comprimento é definido como a distância entre a entrada do tubo até a posição em que a tensão de cisalhamento atinge 2% do valor máximo e é denominado *comprimento de entrada hidrodinâmica* (L_h).

Para fins práticos de projeto, Çengel e Cimbala (2015) consideram que um comprimento de entrada ideal, para um regime turbulento, mais comum na maioria dos casos, pode ser aproximado por 10 vezes o comprimento do diâmetro do tubo (D):

$$L_{h,turbulento} \approx 10 \cdot D. \quad (1)$$

2.1.6 VAZÃO

Vazão ou descarga em uma seção (Q), pode ser definida como o volume de um fluido (V) distribuído no tempo (t), necessário para atravessar esta seção:

$$Q = \frac{V}{t}. \quad (2)$$

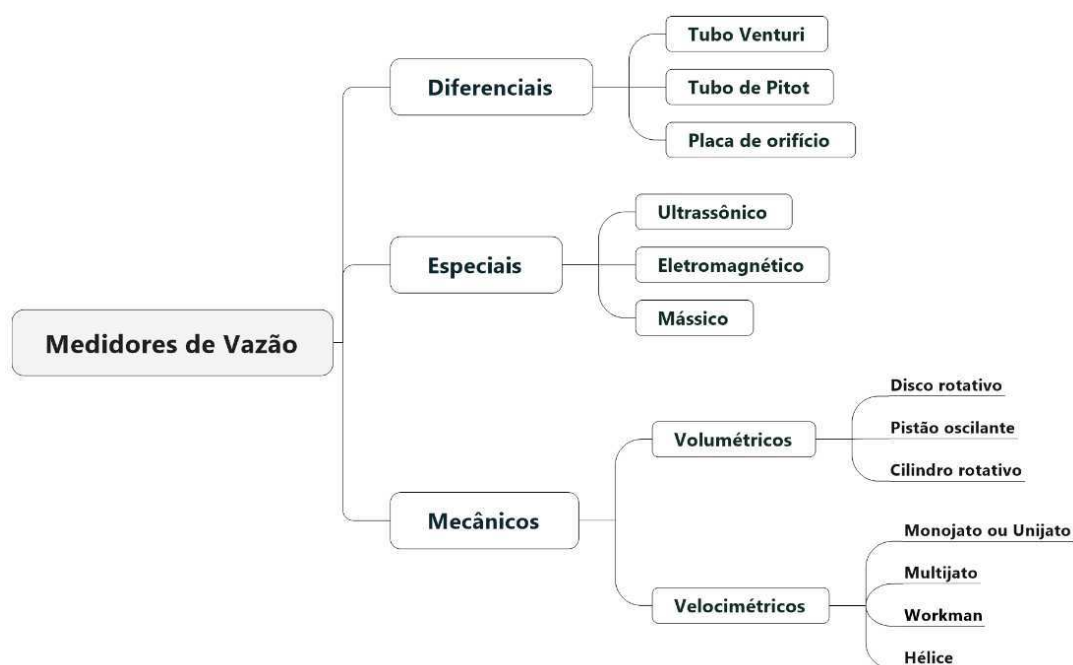
Essa definição é convencionada em outras referências como vazão volumétrica, visto que também é possível definir a vazão por meio da massa específica, velocidade de escoamento e área da seção, convencionado como vazão mássica (ÇENGEL e CIMBALA, 2015). Para este trabalho, utilizou-se o termo “vazão” de forma geral para se referir à vazão volumétrica.

Por inspeção da expressão (2), define-se a unidade de vazão pelo Sistema Internacional (SI) como o metro cúbico por segundo (m^3/s), contudo em alguns casos também são usadas outras unidades derivadas, como litros por hora (L/h).

2.2 MEDIDORES DE VAZÃO

As medições de vazão são realizadas indiretamente por meio de outras grandezas como pressão e rotação, além de outros métodos que podem ser totalmente mecânicos, totalmente eletrônicos ou uma combinação de ambos. Os medidores de vazão para condutos forçados, são classificados em três tipos: os diferenciais, os especiais e os mecânicos, como disposto na Figura 4.

Figura 4: Classificação dos medidores de vazão para condutos forçados.



Fonte: Elaborado a partir de (COELHO, 1983).

Por meio da restrição da passagem do fluido em um tubo ou duto, os medidores do tipo diferenciais utilizam a diferença de pressão entre dois pontos, para calcular a vazão que é diretamente proporcional à pressão medida. Um exemplo de medidor que

utiliza esse princípio é o tubo de Venturi, que utiliza a restrição de área de seção transversal para desenvolver uma diferença de pressão ($P_H - P_L$) em um fluido (DUNN, 2005). Não necessitam de alimentação elétrica e operam de forma invasiva, sendo necessário o seccionamento do tubo ou duto para a sua instalação.

Os medidores especiais recebem essa classificação devido ao seu princípio de funcionamento bem específico. Um exemplo é o medidor ultrassônico, que gera pulsos sonoros em alta frequência e por meio da velocidade de propagação do som no fluido, é capaz de determinar a sua velocidade, volume e vazão em um determinado instante de tempo. Sua principal vantagem em relação aos medidores diferenciais é que são não invasivos e apresentam boa exatidão para tubos metálicos. Contudo, necessitam de alimentação elétrica e apresentam custos elevados de aquisição e manutenção (COELHO, 1983).

Os medidores mecânicos operam por meio de um mecanismo associado ao movimento do fluido no tubo ou duto. São amplamente empregados na medição do consumo de água em residências, sendo popularmente conhecidos como hidrômetros. São subdivididos em duas categorias de acordo com o tipo de mecanismo de rotação.

Os medidores mecânicos volumétricos mais comuns, operam por meio de um pistão rotativo que gira em uma câmara cilíndrica, deslocando volumes definidos de água a cada rotação. Por meio de engrenagens, o número de revoluções é convertido em vazão ou volume. Não possuem partes móveis, apresentam elevada exatidão, são compactos, porém devem ser utilizados apenas para projetos de longo prazo, devido ao custo elevado de aquisição (DUNN, 2005).

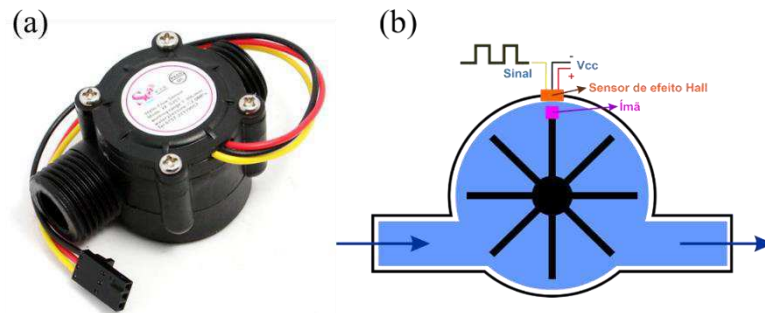
Para os medidores velocimétricos o volume de água é proporcional ao número de revoluções de uma turbina, que também gira em uma câmara cilíndrica. São largamente utilizados na maioria das residências, devido ao custo ser menor do que os medidores volumétricos, porém apresentam menor exatidão (COELHO, 1983).

2.3 SENSOR DE VAZÃO YF-S201

O sensor de vazão YF-S201 na Figura 5 (a), é do tipo mecânico velocimétrico unijato, porém conta com um sistema eletrônico em que é possível obter dados por meio de sinais elétricos. A turbina interna do YF-S201 é formada por oito pás e uma delas possui um ímã acoplado, de forma que ao passar por um sensor de efeito Hall, localizado

em um circuito eletrônico dentro da carcaça, é gerado um pulso quadrado, como esquematizado na Figura 5 (b).

Figura 5: (a) Sensor YF-S201 e o (b) princípio de funcionamento.



Fonte: (SEED STUDIO, 2018) e próprio autor.

Cada pulso gerado está associado a um valor de vazão, por meio de uma expressão disponibilizada pelo *datasheet* do sensor, indicando que existe uma relação linear entre a frequência de pulsos (F) e a vazão de água (Q) em L/min:

$$F = 7 \cdot Q \quad (3)$$

Contudo, como a frequência em (3) é dada pelo número de pulsos por segundo, utilizar esse sensor com microcontroladores é um problema, devido aos tempos inerentes ao dispositivo, como tempo decorrente do cálculos e execuções das instruções. Esses tempos representam pequenos atrasos no programa principal do microcontrolador, contribuindo para uma possível inadequação entre frequência observada e vazão.

2.4 INCERTEZA DE MEDIÇÃO

Para qualquer grandeza medida existe uma incerteza associada, que pode ser decorrente da sua própria natureza ou estar associada ao processo de medição. Para obter resultados com maior confiabilidade, é necessário definir quais as características físicas da grandeza, adotar um método específico e aplicar uma técnica adequada (INMETRO, 2012).

O Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), define um conjunto de diretrizes com o objetivo de garantir a reprodutibilidade e reduzir a incerteza nas medições. Dessa forma, é recomendado a utilização do *Guide of Uncertainty in Measurement* (GUM), que é considerado o método internacional de referência para a expressão da incerteza de medição.

Como a grandeza pode ser aleatória ou sistemática, é recomendado no GUM classificar a metodologia empregada, para que seja possível garantir a reprodutibilidade da técnica, independentemente da natureza do mensurando. Por essa abordagem, tem-se que medições do Tipo A são definidas por meio de uma série de observações independentes. Entretanto, se existirem informações previamente conhecidas sobre a variabilidade da grandeza, devemos classificar a avaliação como do Tipo B (INMETRO, 2012).

2.5 ANÁLISE ESTATÍSTICA DE DADOS

Nesta seção serão abordadas algumas técnicas de análise estatística, para identificar e determinar associações entre dados de um experimento.

2.5.1 MEDIDAS ESTATÍSTICAS DESCRITIVAS

Na análise estatística de uma única variável de uma amostra, destacam-se duas medidas descritivas importantes, que auxiliam a caracterizar os dados quanto a sua centralidade e dispersão: a média aritmética e o desvio-padrão amostral (BECKER, 2015; NAVIDI, 2012).

A média amostral (ou média aritmética) determina o centro de uma distribuição de dados, por meio da soma dos valores de uma variável x , dividida pelo número de n observações. Matematicamente, expressamos a média como:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i. \quad (4)$$

A expressão em (4) é importante para definir em torno de qual valor uma amostra estaria localizada, se todos os seus valores fossem iguais entre si, ou seja, se não houvesse variabilidade nos dados (NAVIDI, 2012). Contudo, devido à incerteza nas medições, considera-se como variância amostral, o desvio (ou diferença) quadrático entre o valor das observações e a sua média, dado por:

$$s_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5)$$

Entretanto, observa-se na expressão (5) que a unidade da variância é dada por um fator quadrático da unidade da amostra, portanto, torna-se mais intuitivo para fins de comparação, a utilização do conceito de desvio-padrão amostral:

$$s_x = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (6)$$

A expressão (6) possui as mesmas características de análise de variabilidade em uma amostra, com a vantagem de ter a mesma unidade que a variável de interesse.

2.5.2 MEDIDAS DE DADOS BIVARIADOS

A aquisição simultânea de duas variáveis de uma amostra, x e y , para n observações, resulta na formação de um conjunto de pares ordenados $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, denominados de dados bivariados (BECKER, 2015; NAVIDI, 2012).

Define-se diagrama de dispersão, a representação cartesiana desse conjunto de pontos, geometricamente localizados em um plano, cujas coordenadas são os valores de cada membro do par de dados. Ao dispor os dados bivariados em um diagrama de dispersão, examina-se a ocorrência de tendências, agrupamento de variáveis, espalhamento de um atributo em relação ao outro e verificar a ocorrência de pontos discrepantes (BECKER, 2015).

A verificação da influência de uma variável sobre a outra, possibilita determinar a relação linear existente entre elas e assim inferir sobre a amostra. A medida estatística que possibilita essa análise, é a covariância:

$$s_{xy} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}). \quad (7)$$

Quando a covariância entre duas variáveis é positiva, pode-se afirmar por meio da expressão (7), que quando x aumenta, y também aumenta; todavia, caso o resultado da covariância seja negativo, as variáveis crescem em sentidos opostos. Se o valor da covariância tender a zero, os dados não apresentam relação linear (BECKER, 2015).

Apesar da covariância ser uma medida de variabilidade conjunta, não é possível comparar graus de relação entre os atributos de uma amostra, devido à presença das unidades de medida das variáveis. Diante disso, basta dividir a covariância pelo produto do desvio-padrão das variáveis, denominado coeficiente de correlação ou coeficiente de Pearson:

$$r = \frac{s_{xy}}{s_x \cdot s_y}. \quad (8)$$

O coeficiente de correlação é adimensional e apresenta uma variação entre -1 e +1. Valores positivos de r indicam que os dados possuem uma tendência de crescimento, enquanto que valores negativos de r indicam uma tendência de decréscimo. Valores de r próximo a +1 ou -1, representam uma forte relação linear e valores próximos de 0, indicam uma ausência de linearidade (NAVIDI, 2012).

2.5.3 VALORES EXTREMOS

A *International Organization of Standardization* (ISO), uma organização responsável por gerir métodos e normas de padronização, recomenda o Teste de Grubbs para detectar pontos discrepantes presentes nas medições. Esses pontos são denominados de valores extremos, exceções ou *outliers* e são decorrentes da variabilidade própria das observações das amostras. Contudo, na maioria dos casos, são gerados devido a erros na aquisição dos dados, na conversão de medidas, entre outros (GRUBBS, 1969).

Para identificar se o valor de uma observação é um valor extremo, faz-se necessário adotar a metodologia proposta por Grubbs:

- (i) Ordenar um conjunto de n observações de uma amostra em ordem crescente, ou seja, $x_1 < x_2 < \dots < x_{n-1} < x_n$, para $n \geq 3$;
- (ii) Supor que os valores nos extremos, x_1 e x_n podem ser *outliers*;
- (iii) Calcular o critério de teste, T_n , para o caso em que x_n é considerado um *outlier*, como apresentado em (9):

$$T_n = \frac{x_n - \bar{x}}{s_x}, \quad (9)$$

em que, \bar{x} é a média amostral das n observações e s_x o desvio-padrão da amostra.

- (iv) Para x_1 , o critério de teste, T_1 , é dado pela expressão em (10):

$$T_1 = \frac{\bar{x} - x_1}{s_x}, \quad (10)$$

- (v) Selecionar o nível de significância que será aplicado ao conjunto de dados, entre 1%, 2,5% ou 5%. Grubbs recomenda a utilização de 5% como valor

padrão. Basicamente, o nível de significância está associado com o risco que alguém está disposto a aceitar, na rejeição de uma observação. Assim sendo, esse valor deve ser utilizado com base em uma escolha adequada para cada problema;

- (vi) Verificar se o critério de teste, T_n ou T_1 , é maior que o valor crítico, T_c , disposto na Tabela 1, considerando o número de amostras e nível de significância adotados. Se $T_n > T_c$ ou $T_1 > T_c$, então o valor inicialmente suposto, x_n ou x_1 , é considerado um valor extremo.

Tabela 1: Valores críticos, T_c , para distintos níveis de significância e número de amostras.

Nº de Amostras n	Nível de Significância			Nº de Amostras n	Nível de Significância		
	5%	2,5%	1%		5%	2,5%	1%
3	1,15	1,15	1,15	20	2,56	2,71	2,88
4	1,46	1,48	1,49	21	2,58	2,73	2,91
5	1,67	1,71	1,75	22	2,60	2,76	2,94
6	1,82	1,89	1,94	23	2,62	2,78	2,96
7	1,94	2,02	2,10	24	2,64	2,80	2,99
8	2,03	2,13	2,22	25	2,66	2,82	3,01
9	2,11	2,21	2,32	30	2,75	2,91	-
10	2,18	2,29	2,41	35	2,82	2,98	-
11	2,23	2,36	2,48	40	2,87	3,04	-
12	2,29	2,41	2,55	45	2,92	3,09	-
13	2,33	2,46	2,61	50	2,96	3,13	-
14	2,37	2,51	2,66	60	3,03	3,20	-
15	2,41	2,55	2,71	70	3,09	3,26	-
16	2,44	2,59	2,75	80	3,14	3,31	-
17	2,47	2,62	2,79	90	3,18	3,35	-
18	2,50	2,65	2,82	100	3,21	3,38	-
19	2,53	2,68	2,85	-	-	-	-

Fonte: (GRUBBS, 1969).

2.6 REGRESSÃO LINEAR SIMPLES

Os modelos de regressão são utilizados em diversas áreas do conhecimento, cujo principal objetivo é representar por meio de uma equação matemática, a relação entre pelo menos duas variáveis de uma amostra. Cada variável é obtida experimentalmente ou a partir de outros modelos de regressão, para um conjunto de n observações. Com isso, realiza-se predições de valores da variável de interesse (BECKER, 2015).

O cálculo do coeficiente de correlação, discutido na seção 2.5.2, permite identificar o comportamento do conjunto de dados, para classificá-los como linear ou não linear. Dessa forma, define-se quais as técnicas utilizadas para modelar a amostra.

A regressão linear simples é usada para estimar o valor esperado de uma variável dependente y , quanto à influência de uma única variável independente x . Contudo, em algumas situações, uma única variável independente não é suficiente para descrever a grandeza medida. Assim, a variável y é dada em função de diversas variáveis independentes, sendo necessário a utilização das técnicas de regressão linear múltipla (KUTNER, 2005; MONTGOMERY, 2012).

Após n observações de um experimento, obtém-se um conjunto de dados bivariados $(x_1, y), (x_2, y), \dots, (x, y_n)$, então, se x e y estiverem linearmente correlacionados, o modelo de regressão linear simples pode ser representado matematicamente, considerando a i -ésima amostra:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{para } i = 1, 2, \dots, n \quad (11)$$

em que y_i é a variável resposta, x_i é a variável conhecida, β_0 é o coeficiente linear, β_1 é o coeficiente angular e ε_i é a variável que representa o erro experimental.

2.7 MÍNIMOS QUADRADOS

Para analisar a regressão linear, é necessário reagrupar a equação (11) em função da variável erro experimental:

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i. \quad (12)$$

Para que os desvios ε_i entre os valores observados e estimados sejam mínimos, reduz-se todos os valores de erros $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, por meio do método dos Mínimos Quadrados (MONTGOMERY, 2012). Esse método consiste em minimizar a soma dos quadrados de todos os desvios:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (13)$$

Substituindo β_0 e β_1 por $\hat{\beta}_0$ e $\hat{\beta}_1$, para indicar valores particulares dos parâmetros que minimizam $S(\beta_0, \beta_1)$, e igualando as derivadas parciais a zero, obtemos um conjunto de equações, cujas soluções são:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (14)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (15)$$

Os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ são chamados de Estimadores de Mínimos Quadrados e o modelo de regressão linear simples ajustado é dado por:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \quad (16)$$

Assim, para cada conjunto de dados bivariados medido (x_i, y_i) , tem-se o respectivo par ordenado sobre a reta (x_i, \hat{y}_i) , em que a distância entre esses pontos é definida como resíduo (e_i):

$$e_i = y_i - \hat{y}_i \quad (17)$$

Dessa forma, quanto mais próximos os resíduos estiverem de 0, mais fidedigno é o modelo matemático ajustado pelos mínimos quadrados. Pode-se interpretar um resíduo negativo como um valor subestimado de sua previsão, enquanto que um resíduo positivo indica um valor superestimado.

2.8 INDICADORES DE ERROS DE PREDIÇÃO

Um indicador de erro de predição analisa o comportamento dos resíduos de uma reta de mínimos quadrados, a fim de obter uma estimativa acerca de suas características (MORETTIN, 2008).

Também é possível identificar valores extremos por meio da representação gráfica de um histograma de resíduos. Assim, para uma melhor interpretação dos dados, faz-se necessário calcular os valores dos resíduos padronizados, também conhecido como escore-Z (RUMSEY, 2014):

$$Z = \frac{e_i - \bar{e}_i}{s_e}. \quad (18)$$

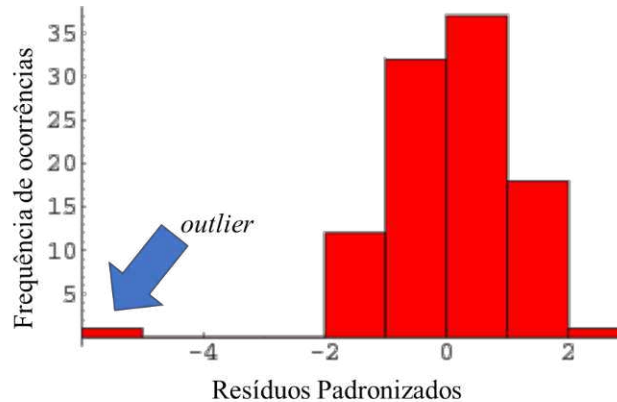
A análise dos escores-Z nos indica se os resíduos padronizados estão na forma de uma distribuição normal padrão, suposição essencial para que os resultados da regressão linear sejam confiáveis (NAVIDI, 2012).

Quaisquer valores acima de +3 ou -3 em um histograma de resíduos, podem ser considerados *outliers*, devido à característica de distribuição normal apresentada pela reta

de mínimos quadrados, que exige que os dados estejam concentrados dentro deste intervalo (RUMSEY, 2014).

É apresentado na Figura 6, o exemplo de um *outlier* detectado por meio da análise gráfica de um histograma de resíduos padronizados com escore-Z.

Figura 6: Histograma de resíduos padronizados, indicando a presença de um *outlier* além de -4.



Fonte: Adaptado de (WOLFRAM, 2018).

A Média Percentual Absoluta do Erro (MAPE) é uma medida estatística de análise, que expressa a exatidão em porcentagem, com relação ao ajuste do modelo de regressão linear.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100 \quad (19)$$

O MAPE é amplamente utilizado e isso se deve em grande parte à sua facilidade de interpretação, pelo fato de ser expresso em termos percentuais. No entanto, ao analisar a expressão (19), percebe-se que não deve ser utilizado valores nulos de y_i , por causar uma indeterminação devido à divisão por zero no somatório.

2.9 INDÚSTRIA 4.0

A chamada Quarta Revolução Industrial, ou Indústria 4.0, é caracterizada por sistemas ciber-físicos ou pela sigla em inglês, CPS. Os sistemas ciber-físicos são responsáveis por criarem um ambiente colaborativo, que tem por objetivo controlar processos físicos por meio de elementos computacionais integrados (SCHWAB, 2016).

Essas tecnologias habilitadoras e convergentes, são caracterizadas sobretudo por mudarem rapidamente o panorama tecnológico vigente e transformarem a sociedade de forma radical, promovendo o desenvolvimento econômico, social e cultural de um povo,

segundo a definição do Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC, 2018).

Essa transformação na indústria de manufatura é observada também na indústria de *utilities*. No entanto, a *McKinsey & Company*, empresa de consultoria nas áreas de distribuição de energia elétrica e gás natural, afirmou em um relatório de 2016 que muitos gestores norte-americanos do setor de *utilities* veem a Indústria 4.0 como uma ameaça ao modelo atual de gerência e negócios (BOOTH, MOHR e PETERS, 2016).

Essa visão também é observada nas indústrias de *utilities* aqui no Brasil, como foi apresentado no relatório da Logicalis, empresa global de soluções e serviços de Tecnologia da Informação e Comunicação (TIC). De acordo com o estudo, apenas 37% dos 176 empresários afirmam que soluções em IoT são importantes para seus negócios. Apenas 18% já adotam alguma solução conectada à internet, enquanto 19% afirmaram estar em processo de habilitação (LOGICALIS, 2017).

2.10 COMPUTAÇÃO EM NUVEM COM A THINGSPEAK®

A computação em nuvem ou em inglês, *cloud computing*, é um novo tipo de aplicação caracterizado por acesso rápido, *on demand*, rapidamente adaptável, flexível e capaz de compartilhar dados de forma simplificada (GENG, 2015).

A computação em nuvem proporciona uma extensão da internet já existente, visto que por meio de uma infraestrutura própria, consegue promover o acesso de suas informações de forma remota. Essa característica está sendo amplamente usada, sobretudo, por empresas que não necessitam mais manter um banco de dados próprio, basta apenas utilizar um serviço de computação em nuvem, para obter mais fluidez e autonomia nas decisões, além de poupar custos fixos com manutenção e operação (GENG, 2015).

Dessa forma, uma aplicação integrada à computação em nuvem, deve ser capaz de realizar operações básicas de coleta de dados, análise e responder a ações pré-definidas ou devido à presença de um dado específico (OLIVEIRA, 2017).

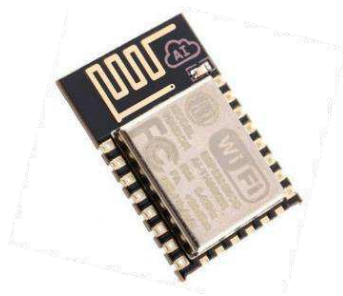
A *Thingspeak*® é uma plataforma de análise de dados, que permite que seja possível agregar dados provenientes de sensores e informações em rede. Também é permitido enviar dados do *Thingspeak*® para um dispositivo IoT e vice-versa, além de integração com o MATLAB®, para desenvolvimento de algoritmos de análise e predição, como algoritmos de aprendizado de máquina, *big data* e *analytics* (OLIVEIRA, 2017).

2.11 PROJETOS DE IOT COM O ESP8266

O termo Internet das Coisas é definido pela *International Telecommunication Union* (ITU), por meio da recomendação ITU-T Y.2060, como uma infraestrutura global para a sociedade da informação, em que serviços avançados são ofertados por meio da conexão de dispositivos, por meio da tecnologia de informação e comunicação já existente (ITU, 2012).

Dado esse cenário, atualmente existem diversos dispositivos baseados no princípio da Internet das Coisas. Um dispositivo que vem sendo utilizado com frequência em projetos de IoT, é o microcontrolador ESP8266, mostrado na Figura 7. Ele é fabricado pela Espressif, possui um microprocessador Tensilica L106, de 32 bits, operando com frequência padrão de 80 MHz, podendo chegar a 160 MHz. Possui suporte embutido a *Wi-Fi* (802.11) e memória *flash* de 4 MB.

Figura 7: Microcontrolador ESP8266, da Espressif.



Fonte: (INDIA MART, 2018).

A tensão nominal de operação é 3,3 V e apresenta uma corrente de consumo baixo, na ordem de 170 mA quando operando pacotes do protocolo *Wi-Fi*. Apresenta alguns modos configuráveis de economia de energia, fazendo com que a corrente de operação atinja 20 μA no modo *sleep*, em que algumas funções do microprocessador são desabilitadas.

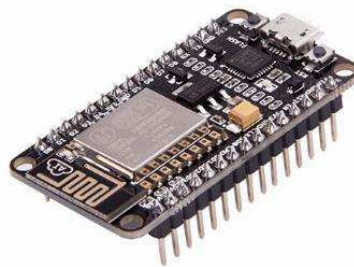
3 METODOLOGIA

Nesta seção é abordado os materiais e métodos necessários para a execução deste trabalho, com a descrição dos procedimentos necessários para a aquisição, calibração e validação do sistema proposto. Também é abordado os componentes e desenvolvimento de aplicações com soluções no contexto da Internet das Coisas.

3.1 ESP8266 NODEMCU

Existem vários módulos com o ESP8266 embarcado, em que variam número de interfaces de entrada e saída (GPIO), memória flash disponível ou microprocessador. O módulo utilizado neste trabalho, o NodeMCU, apresentado na Figura 8, é baseado no módulo ESP-12 e possui alguns recursos adicionais, como conversor USB-serial, regulador de tensão de 5 V para 3,3 V e tratamento de capacitâncias parasitas, de forma a criar um ambiente de desenvolvimento muito parecido com as placas Arduino®.

Figura 8: Módulo ESP8266 NodeMCU.



Fonte: (SEED STUDIO, 2018).

Contudo, ao comparar o NodeMCU com o Arduino Due, que apresenta características semelhantes, percebe-se na Tabela 2 que o NodeMCU é uma placa bastante versátil, com configurações adequadas para a maioria dos projetos em IoT.

Tabela 2: Comparativo entre as placas de desenvolvimento da Espressif e Arduino.

Características	NodeMCU	Arduino Due
<i>Tensão de operação</i>	3,3 V (digital)/ 1 V (analógico)	3,3 V
<i>Frequência de operação</i>	80 MHz/ 160 MHz	84 MHz
<i>Processador</i>	32 bits (Tensilica)	32 bits (ARM)
<i>Microcontrolador</i>	ESP8266	AT91SAM3X8E
<i>Memória Flash</i>	4 MB	512 kB
<i>Portas Digitais de E/S</i>	11	54
<i>Módulo Wireless</i>	Wi-Fi	Nenhum
<i>Módulo ADC</i>	1 canal, 10 bits	12 canais, 12 bits
<i>Módulo DAC</i>	Nenhum	2 canais, 12 bits
<i>Preço¹</i>	US\$ 8,20	US\$ 49,95

Fonte: (ESPRESSIF, 2018) e (ARDUINO, 2018).

¹ Os preços foram consultados na mesma loja, a *Seed Studio*, em 28 de fevereiro de 2018.

Para o desenvolvimento deste trabalho, foi necessário apenas sensor, sendo considerado mais importante a frequência de operação, memória *flash* disponível para armazenamento de dados e comunicação nativa *Wi-Fi*. Essas características possibilitam que o projeto possa ser inserido em nós de rede de comunicação sem fio, tornando-o uma aplicação IoT.

Outra vantagem que o NodeMCU tem sobre o Arduino Due, é que ele pode ser configurado por meio de diversas linguagens como C, C++, Python, além de sua linguagem nativa, a Lua, desenvolvida por brasileiros. Além disso, o ambiente de desenvolvimento integrado (IDE) da Arduino é compatível com o NodeMCU, de modo que a grande maioria das bibliotecas podem ser usadas em ambas as plataformas.

3.2 MEDIDOR E SENSOR DE VAZÃO

Para realizar o procedimento de calibração, foi utilizado um medidor de vazão mecânico velocimétrico unijato da marca Saga[®], cujas especificações estão apresentadas na Tabela 3. O medidor não foi calibrado e considerou-se apenas a sua calibração de fábrica.

Tabela 3: Especificações técnicas do medidor de vazão Saga[®] US-3,0.

<i>Diâmetro Nominal</i>	<i>Pressão Máxima</i>	<i>Leitura Mínima de Medição</i>
20 mm (3/4")	1 MPa	0,002 L
<i>Vazão Máxima</i>	<i>Vazão Nominal</i>	<i>Vazão Mínima</i>
3,0 m ³ /h	1,5 m ³ /h	30 L/h

Fonte: (SAGA, 2017).

A NBR 12.218/1994, sobre projetos de rede de distribuição de água para abastecimento público, afirma que as pressões estática e dinâmica de um sistema de abastecimento de água são respectivamente 500 kPa e 100 kPa (ABNT, 1994). O medidor de vazão foi escolhido por operar em uma faixa de operação 10 vezes maior, garantindo a integridade do sistema e das medições.

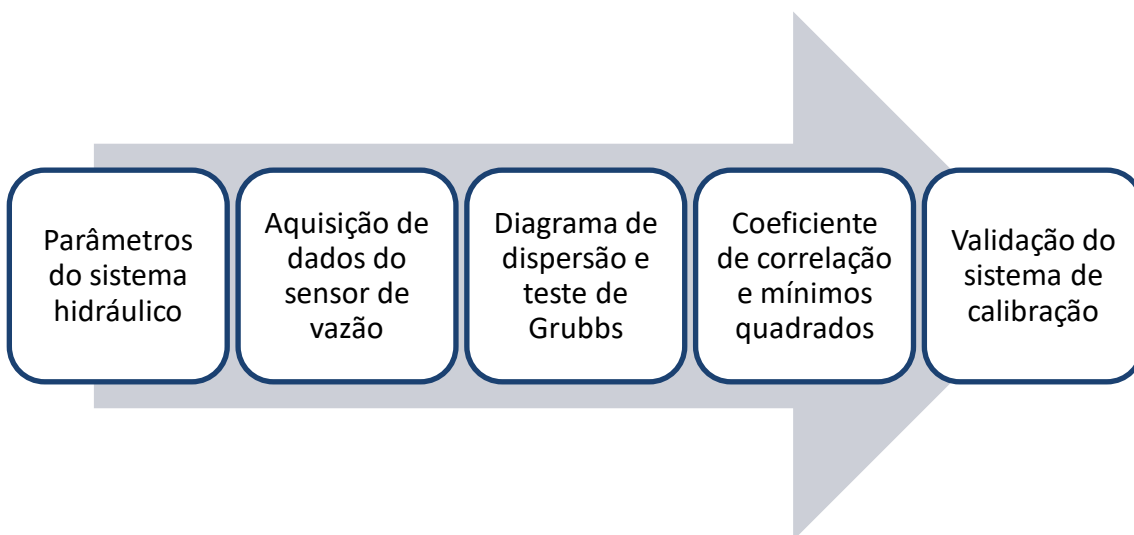
O medidor de vazão mecânico foi escolhido por estar presente na grande maioria das edificações que utilizam tubulações de água fria, além de possuir uma leitura mínima de medição com grande exatidão. Assim, o medidor utilizado foi utilizado como um padrão de comparação para o sensor de vazão calibrado.

3.3 SISTEMA DE CALIBRAÇÃO DO SENSOR

A fim de evitar uma intensa dependência do tempo, foi proposta uma nova abordagem para calcular a vazão por meio da determinação de uma nova reta de calibração, relacionando número de pulsos e volume de água. Com a informação de volume, basta fazer uma observação em tempos pré-definidos (a cada hora, por exemplo) e então foi possível obter a vazão diretamente em termos de suas unidades do SI, o m³/h.

Para implementar esse novo modelo de calibração, foi importante seguir alguns procedimentos, a fim de determinar um modelo estatístico que possua a menor incerteza associada. Esse processo é apresentado na Figura 9.

Figura 9: Procedimento de modelagem estatística para determinação de uma nova curva de calibração.



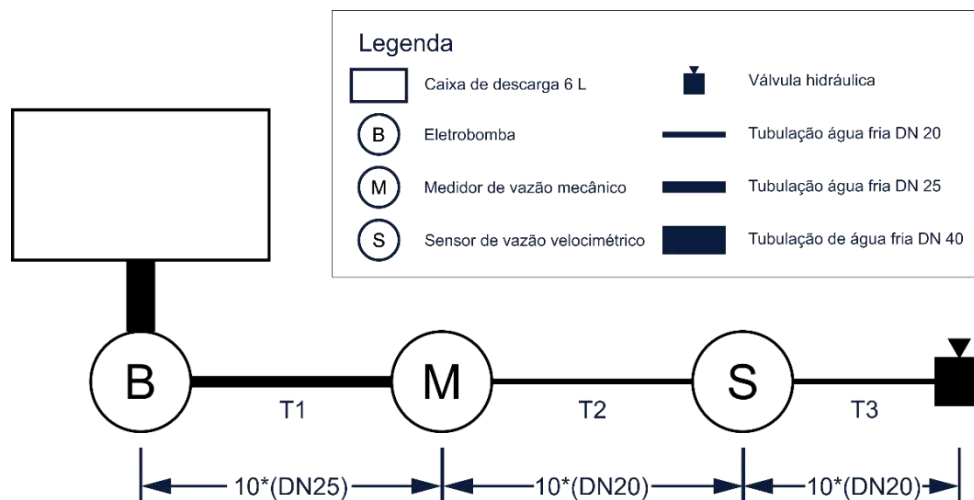
Fonte: Próprio autor.

3.3.1 PARÂMETROS DO SISTEMA HIDRÁULICO

O sistema de calibração do sensor YF-S201, foi concebido como uma estrutura hidráulica com tubulações de água fria em PVC, com diâmetro de 38mm (DN40), 25 mm (DN25) e 20 mm (DN20). Acoplado ao sistema, foi utilizada uma caixa de descarga com 6 L de capacidade e uma eletrobomba de 34 W, modelo BAV1126, usado em algumas máquinas de lavar roupa, para manter o fluxo de água constante em toda a tubulação, evitando o surgimento de bolhas de ar.

Na saída da caixa de descarga, instalou-se a eletrobomba e o medidor de vazão mecânico, para simular o comportamento de um sistema de abastecimento de água. A jusante do medidor, foi inserido o sensor de vazão e uma torneira como válvula de controle. O sistema está ilustrado na Figura 10.

Figura 10: Esquema do sistema de calibração de sensor de vazão.



Fonte: Próprio autor.

Os diâmetros das tubulações nas seções T1, T2 e T3 foram calculados com o auxílio dos conceitos de região de entrada para fluidos dinâmicos. Por isso, o comprimento mínimo de cada seção foi calculado de acordo com o diâmetro nominal dos dispositivos. Na Tabela 4 estão dispostos os valores de comprimento para cada seção.

Tabela 4: Comprimentos mínimos para cada seção do sistema hidráulico.

	Seção T1	Seção T2	Seção T3
DN	25 mm	20 mm	20 mm
L_h	250 mm	200 mm	200 mm

Fonte: Próprio autor.

3.3.2 AQUISIÇÃO DE DADOS DO SENSOR DE VAZÃO

Para obter os pontos de interesse, realizou-se a aquisição de pontos de volume (em m^3 ou L) e de números de pulsos (adimensional). A caixa de descarga foi abastecida com água até o limite máximo e a válvula de controle foi aberta para que o ar presente nas tubulações fosse liberado. Após 1 min a válvula foi fechada e o líquido liberado, devolvido de volta à caixa de descarga. Para garantir que o nível do reservatório não ficasse muito baixo e pudesse criar vórtices que acumulassem ar na tubulação, manteve-se durante o experimento, pelo menos 1/3 da caixa preenchida com água (2 L).

Utilizou-se o mínimo de conexões entre as tubulações, para evitar vazamentos e criação de áreas turbulentas, que pudessem interferir nas medições. Para isso, utilizou-se fita veda-rosca e cola específica para canos de PVC, para realizar as conexões devidas.

O NodeMCU foi utilizado para obter os dados do sensor de vazão YF-S201 e a transmissão de dados foi realizada via cabo USB, por comunicação serial, para um computador pessoal. Os dados foram visualizados pelo “*Monitor Serial*”, disponibilizado pelo ambiente de desenvolvimento do Arduino, que é compatível com o NodeMCU.

O microcontrolador apresenta um pino de interrupção externa (D2) que associado a um algoritmo específico desenvolvido, formado apenas por um contador unitário, possibilitou contar o número de pulsos gerados pelo sensor. Para obter esses valores, liberou-se volumes definidos de água, controlados pelo medidor de vazão mecânico, começando por um valor de 0,100 L e incrementando o volume com passos de 0,100 L.

Para garantir uma melhor distribuição para os dados, observou-se 10 amostras para cada 0,100 L de volume liberado pelo sistema e por fim, foram calculados a média e o desvio padrão de cada conjunto de amostras. As observações ocorreram até um volume de 2 L, de forma a garantir que o reservatório não ficasse abaixo de 50% do seu volume máximo.

3.3.3 DIAGRAMA DE DISPERSÃO E TESTE DE GRUBBS

O diagrama de dispersão foi calculado com base no referencial teórico abordado na seção 2.5.2 deste trabalho. O diagrama é essencial para identificar a possível natureza da relação entre volume de água e número de pulsos do sensor YF-S201.

Para identificar possíveis valores de exceção ou *outliers*, aplicou-se o teste de Grubbs, que prevê a identificação de valores extremos em um conjunto de dados. Portanto, para cada conjunto de 10 amostras foi aplicado o teste de Grubbs, com o auxílio de um software de computação algébrica, o MATLAB®. Caso seja identificado um *outlier*, ele deve ser removido das amostras e o cálculo da média será realizado com as demais amostras.

No entanto, se em um conjunto de amostras for identificado mais de 5 valores extremos, o conjunto inteiro deve ser excluído e um novo experimento deve ser feito, para o volume de água referente ao conjunto. Se o surgimento de mais de 5 *outliers* permanecer, pode-se concluir que o experimento e volume observado possuem alguma

relação estocástica e, portanto, os dados devem ser levados em consideração para o cálculo da média.

Por fim, os pontos de análise são formados pelas médias relativas a cada conjunto de amostras de volume e seus respectivos números de pulso do sensor de vazão. Esse conjunto de dados é utilizado para o cálculo do coeficiente de correlação e reta de mínimos quadrados.

3.3.4 COEFICIENTE DE CORRELAÇÃO E MÍNIMOS QUADRADOS

Como foi dito, apenas a observação visual não garante a linearidade entre as amostras. Assim, fez-se necessário o cálculo do coeficiente de correlação, para poder determinar a reta de mínimos quadrados que representa o comportamento do sensor. Para o cálculo do coeficiente, utilizou-se a ferramenta “*cftool*”, nativa do *software* MATLAB®, para determinação do coeficiente de correlação (*r-square*) e a equação da reta de mínimos quadrados (*polynomial, degree 1*). Considerou-se que o volume de água é uma variável independente conhecida (*x*) e o número de pulsos a variável dependente ou resposta (*y*).

Com a equação da reta de mínimos quadrados foi possível obter os valores de previsão (\hat{y}_i) e os resíduos (e_i) de cada observação.

3.3.5 VALIDAÇÃO DO SISTEMA DE CALIBRAÇÃO

Para validar o sistema calibrado, foi necessário calcular os resíduos padronizados, o *score-Z* e plotar a informação por meio de um histograma. Por conseguinte, deve-se detectar se os resíduos estão entre -3 e +3, por meio da utilização das funções do MATLAB® “*zscore*” e “*histogram*”.

3.4 SISTEMA INTEGRADO À NUVEM

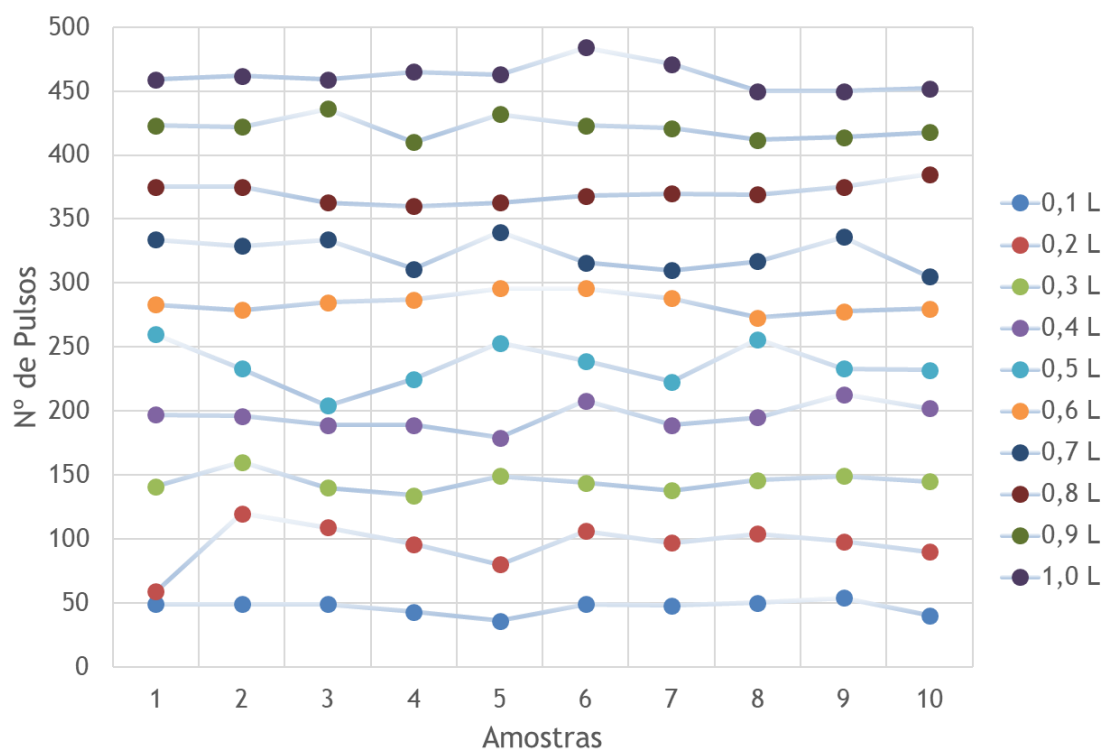
Devido à possibilidade de integração com o MATLAB® e por dispor de uma plataforma completamente desenvolvida, o *Thingspeak*® foi escolhido como plataforma de aquisição de dados para o projeto deste trabalho. Além disso, a plataforma também possui integração com o microcontrolador ESP8266, por meio de uma biblioteca de funções.

4 RESULTADOS E DISCUSSÃO

4.1 COLETA E PRÉ-TRATAMENTO DE DADOS

Seguindo o procedimento padrão discutido na seção 3.3 deste trabalho, foi coletada uma coleção de pontos para diversos volumes, com passo de 0,100 L, começando pelo valor de 0,100 L até 4,0 L. Uma parte do conjunto de dados está ilustrado no gráfico da Figura 11.

Figura 11: Aquisição de 10 amostras para diversos volumes observados do sensor de vazão YF-S201.



Fonte: Próprio autor.

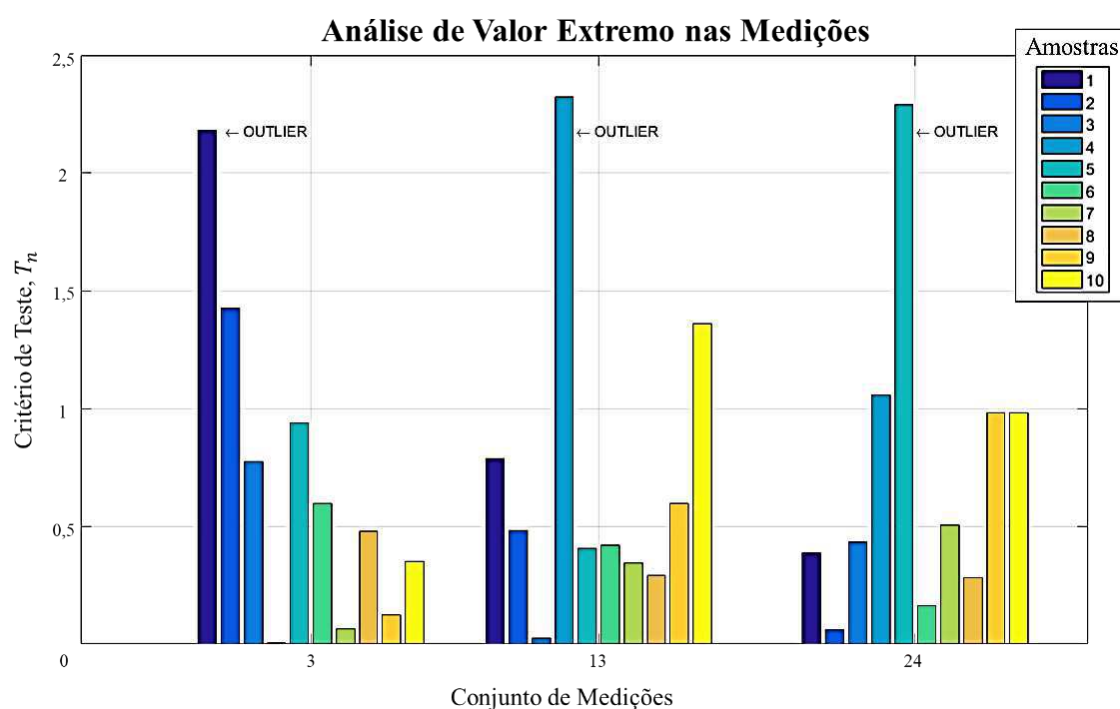
Devido à variação estocástica presente nas amostras para um mesmo valor de volume, como pode ser observado para o conjunto de amostras de 0,2 L, por exemplo, foi utilizada a média das 10 observações em detrimento da escolha de apenas um desses valores.

Portanto, para a construção do diagrama de dispersão, foram utilizados os dados bivariados volume e média de pulsos. Com o auxílio computacional do *software* MATLAB®, foi possível determinar graficamente o comportamento do diagrama de dispersão e o cálculo do coeficiente de correlação, por meio da ferramenta “*cftool*”.

O coeficiente de correlação obtido foi de $r = 0,9998$, que indica uma forte linearidade positiva entre a média de pulsos e o volume de água. Dessa forma, foi possível calcular os componentes da reta de mínimos quadrados. Contudo, de acordo com a recomendação da ISO, deve-se aplicar um método para detecção de valores extremos. Diante disso, foi aplicado o teste de Grubbs para detecção de pontos extremos nas amostras dos pulsos gerados pelo sensor de vazão.

Utilizando o procedimento descrito na seção 2.5.3, para 5% de significância dos dados e 10 amostras, o que implica em $T_c = 2,18$, foi composto um novo conjunto de dados. Os dados que apresentaram valores extremos estão dispostos na Figura 12.

Figura 12: Valores para o teste de Grubbs para as amostras de pulsos do sensor de vazão YF-S201.



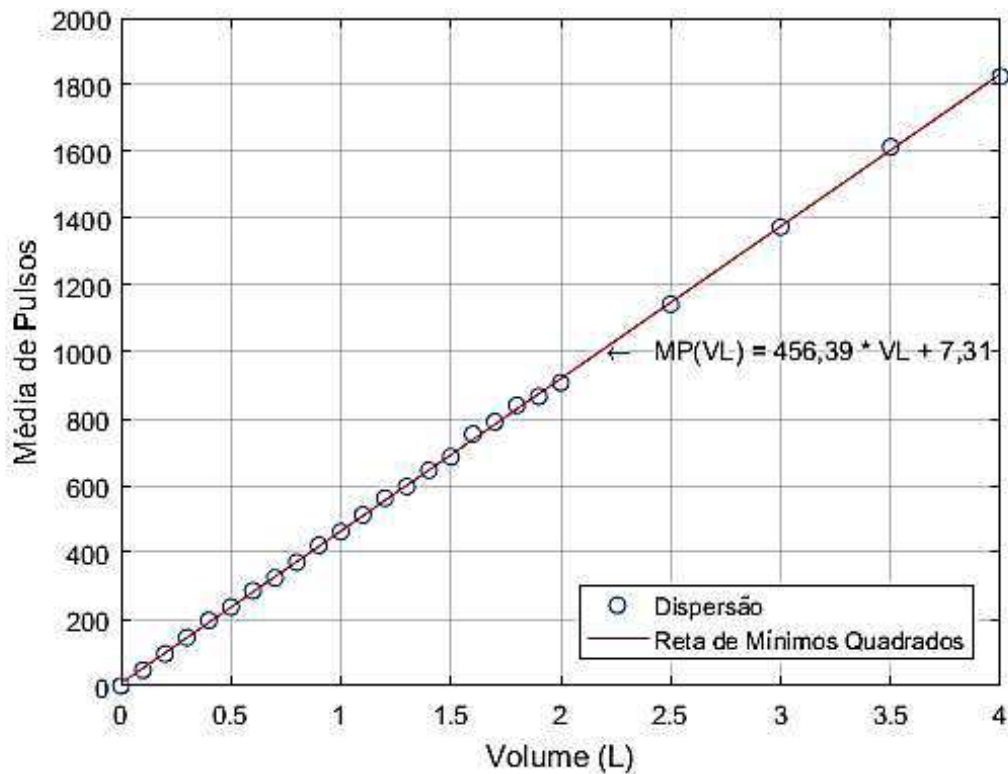
Fonte: Próprio autor.

Apenas três amostras apresentaram valor extremo, em um conjunto total de 250 pontos observados. Para os conjuntos de medições 3, 13 e 24, os *outliers* foram removidos e a média e desvio padrão de cada conjunto foi recalculado.

4.2 ANÁLISE ESTATÍSTICA PARA A REGRESSÃO LINEAR

Utilizando a média e o desvio padrão sem *outliers*, foi realizado o cálculo dos parâmetros da reta de mínimos quadrados, por meio da ferramenta “*cftool*” do MATLAB® e o gráfico está representado na Figura 13.

Figura 13: Gráfico da reta de regressão linear pelos Mínimos Quadrados.



Fonte: Próprio autor.

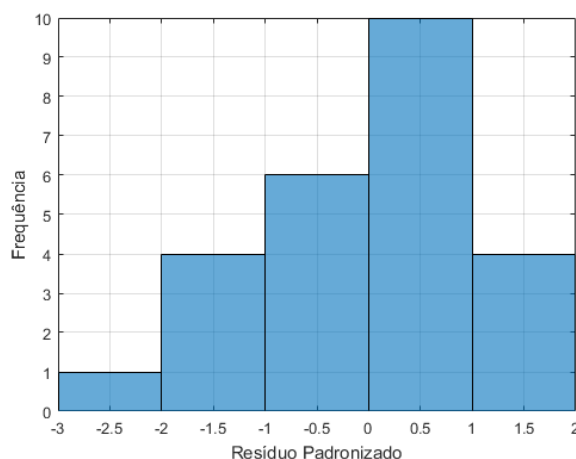
A aplicação dos métodos para obtenção da reta de mínimos quadrados resultou na equação em (20), que relaciona média de pulsos (MP) e volume de água (VL).

$$MP = 456,39 \cdot VL + 7,31 \quad (20)$$

Porém, é válido analisar o contexto em que a equação está inserida. Considerando um $VL = 0$, o sensor deveria indicar uma média de 7,31 pulsos, resultando em um dado incongruente. Para isso, o ideal seria remover o fator “7,31” da equação. Dessa forma, a equação (17) reproduziria com fidelidade o comportamento físico observado. Entretanto, o impacto dessa alteração deve ser analisado sob o ponto de vista estatístico, para que o modelo desenvolvido inicialmente não produza dados equivocados.

Para analisar o comportamento da nova expressão, utilizou-se o conceito de indicadores de erro de predição, da seção 2.8. Inicialmente, foram calculados os escores-Z, para criar o histograma dos resíduos padronizados, como apresentado na Figura 14.

Figura 14: Histograma dos resíduos padronizados, para a expressão modificada.



Fonte: Próprio autor.

Por meio da Figura 14, observa-se que todos os resíduos padronizados, encontram-se na faixa entre ± 3 , portanto, pode-se inferir que a equação modificada não apresenta *outliers*. Para analisar o erro devido ao ajuste do modelo pelos mínimos quadrados, foi calculada a média percentual absoluta do erro (MAPE).

Tabela 5: Indicadores de erro de predição para a equação original e equação modificada da reta de mínimos quadrados do sensor YF-S201.

Indicador	$MP_o = 456,39 \cdot VL + 7,31$	$MP_m = 456,39 \cdot VL$
MAPE	1,45%	2,05%

Fonte: Próprio autor.

A MAPE em ambos os casos, seja na equação original (MP_o) ou na equação modificada (MP_m), não atingiu mais do que 5%, relativos ao intervalo de confiança definido no teste de Grubbs. Assim, pode-se afirmar que é facultativo o uso de ambas as equações, porém devido ao fator explicado, a equação modificada da reta é a melhor escolha por representar a natureza do processo de forma mais completa e fidedigna.

4.3 COMPUTAÇÃO EM NUVEM COM *THINGSPEAK*[®]

O ambiente escolhido para desenvolver a aplicação de computação em nuvem, foi o *Thingspeak*[®], que apresenta uma versão gratuita e não escalonável, com direito a três milhões de mensagens por ano, ou cerca de 8200 mensagens por dia, suficiente para um teste com apenas um sensor.

Uma importante limitação da versão gratuita do *Thingspeak*[®] foi o tempo de envio entre uma mensagem e outra, que foi de 15 segundos. Para aplicações que

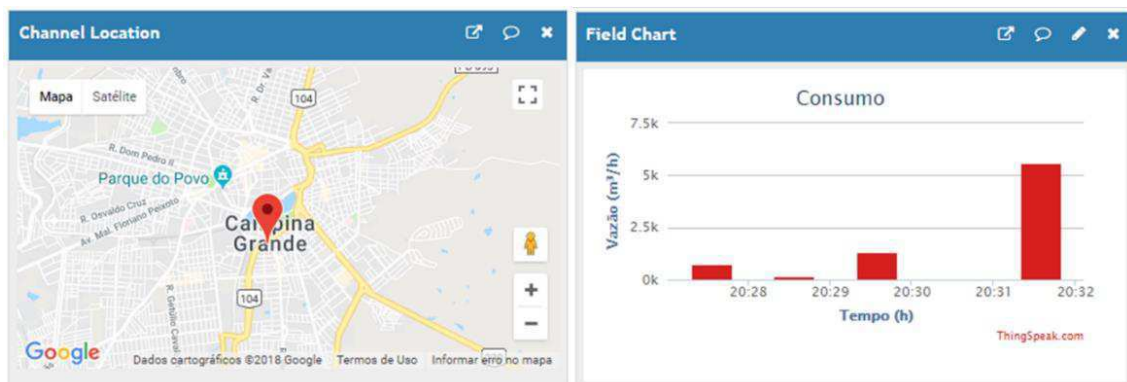
necessitem de atualizações frequentes e em tempo real, a versão gratuita pode representar uma dificuldade. Para resolver esse contratempo, foi desenvolvido um algoritmo específico para a atualização dos dados no banco de dados do *Thingspeak*[®]. Uma interrupção externa por borda de descida realiza a indicação de que ocorreu um pulso, enquanto que uma interrupção de temporização habilita o envio dos dados coletados para o *Thingspeak*[®], em intervalos de tempo bem definidos.

A princípio essa foi uma possível solução para evitar o uso de funções como o “*delay()*”, que inserem atrasos na execução das instruções do *hardware*, promovendo envios de dados não sincronizados. Entretanto, essa solução não apresentou uma resposta satisfatória quando a interrupção temporizada enviou o dado, enquanto a interrupção também estava realizando a contagem dos pulsos. Percebeu-se que durante o envio de dados para o *Thingspeak*[®], as duas interrupções realizavam simultaneamente um atraso de aproximadamente dois segundos na execução do programa. E além disso, os dados capturados pelo sensor durante esse período eram perdidos, pois o sistema embarcado não conseguia armazenar a informação.

Dessa forma, criou-se um algoritmo capaz de impedir que o dado fosse enviado, caso as duas interrupções estejam ativas no mesmo instante. Assim, quando a interrupção temporizada completar seu ciclo, só será permitido enviar os dados coletados, caso a interrupção externa responsável pela contagem dos pulsos esteja inoperante.

Foi observado durante os testes, que a nova abordagem de *software* funcionou corretamente. Assim foram desenvolvidas duas janelas na plataforma *Thingspeak*[®], uma com a localização atual do medido, obtido por meio de uma aplicação de geolocalização por IP, da Google[®] e outra janela com as últimas medições de consumo, como pode ser visto na Figura 15.

Figura 15: Janelas de dados no *Thingspeak*[®], indicando a localização atual e o consumo.



Fonte: Próprio autor.

5 CONCLUSÃO

O estudo acerca das incertezas associadas ao novo modelo de medição de consumo proposto, para o sensor de vazão YF-S201, apresentou elevado grau de fidelidade. A aplicação do método dos mínimos quadrados e o estudo acerca dos valores extremos, foi fundamental para dar base científica e estatística à metrologia do processo empregado. Contudo, não foram realizados testes para determinar se a presença de ar nos tubos poderia modificar algum parâmetro da curva de calibração do sensor.

Para a obtenção de resultados com menor incerteza associada, faz-se necessário a utilização de um medidor de vazão calibrado, a fim de realizar estudos mais aprofundados acerca da variabilidade dos dados obtidos. Recomenda-se a calibração do medidor em um laboratório de metrologia creditado pelo INMETRO, para garantir uma maior isonomia no processo e aferir os dados com maior exatidão.

Entretanto, a construção do modelo matemático para o sensor de vazão YF-S201, representou uma nova metodologia de medição. Portanto, é possível definir uma curva de calibração de volume em função do número de pulsos, o que possibilitou obter dados com menor influência do tempo, o que é fundamental para sistemas embarcados.

Estudar e propor novas metodologias de aquisição para os sensores, representa, portanto, um novo conjunto de regras a fim de trazer dados cada vez mais consistentes e que representem com fidelidade a natureza do mensurando. Além disso, a plataforma *online* de Internet das Coisas, também foi fundamental para promover mobilidade e agilidade em uma possível tomada de decisões. No âmbito do usuário final, foi dado a ele a possibilidade de acompanhar visualmente o consumo de água pessoal e ter um maior planejamento sobre seus gastos. Para gestores, essa pode ser uma nova ferramenta para obtenção de informações cada vez mais abrangentes sobre o sistema e seus usuários.

Todavia, devido a limitações físicas e operacionais, não foi possível ir além no estudo desse sistema e testar outros sensores e capacidades de vazão, assim como medir a pressão interna das tubulações, para inferir sobre a influência dos mesmos na incerteza da medição. É sugerido para futuros trabalhos, utilizar *machine learning* e *big data*, a fim de abordar essas outras características. Essa abordagem pode permitir análises mais profundas acerca do sistema de abastecimento de água, e por meio de algoritmos de predição, determinar possíveis soluções para os problemas apresentados.

REFERÊNCIAS

- ABNT. **NBR 12218: Projeto de rede de distribuição de água para abastecimento público**. Rio de Janeiro. 1994.
- ARDUINO. Arduino Due. **Arduino.cc**, 2018. Disponível em: <<https://store.arduino.cc/usa/arduino-due>>. Acesso em: 28 fevereiro 2018.
- BECKER, J. L. **Estatística básica: transformando dados em informação**. Porto Alegre: Bookman, 2015.
- BOOTH, A.; MOHR, N.; PETERS, P. The digital utility: new opportunities and challenges. **McKinsey & Company**, 2016. Disponível em: <<https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/the-digital-utility-new-opportunities-and-challenges>>. Acesso em: 20 fevereiro 2018.
- ÇENGEL, Y. A.; CIMBALA, J. M. **Mecânica dos fluidos: fundamentos e aplicações**. 3ª. ed. São Paulo: AMGH Editora, 2015.
- COELHO, A. C. **Medição de água e controle de perdas**. Rio de Janeiro: ABES, 1983.
- DUNN, W. C. **Fundamentals of industrial instrumentation and process control**. New York: McGraw-Hill, 2005.
- ESPRESSIF. ESP8266, 2018. Disponível em: <<https://www.espressif.com/products/hardware/esp8266ex/overview/>>. Acesso em: 28 fevereiro 2018.
- GENG, H. **Data center handbook**. New Jersey: John Wiley & Sons, 2015.
- GRUBBS, F. E. Procedures for detecting outlying observations in samples. **Technometrics**, v. 11, n. 1, p. 21, fevereiro 1969.
- INDIA MART. ESP8266 -12E WiFi Module. **IndiaMart**, 2018. Disponível em: <<https://www.indiamart.com/proddetail/esp8266-12e-wifi-module-16469774997.html>>. Acesso em: 28 fevereiro 2018.
- INMETRO. Avaliação de dados de medição: guia para a expressão de incerteza de medição (GUM 2008), p. 138, 2012. Disponível em: <http://www.inmetro.gov.br/inovacao/publicacoes/gum_final.pdf>. Acesso em: 21 fevereiro 2018.
- ITU. ITU-T Y.4000/Y.2060 (06/2012). **ITU-T Recommendations**, 2012. Disponível em: <<https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060>>. Acesso em: 10 março 2018.
- KUTNER, M. H. **Applied linear statistical models: regression, analysis of variance, and experimental designs**. 2ª. ed. New York: McGraw-Hill, 2005.
- LOGICALIS. IoT Snapshot 2017. **Logicalis**, 2017. Disponível em: <https://www.la.logicalis.com/globalassets/latin-america/advisors/pt/_iot_snapshot_2017_vfinal_web.pdf>. Acesso em: 28 fevereiro 2018.
- MCNABB, D. E. **Public utilities: management challenges fo the 21st century**. Cornwall: Elgar Publishing, 2005.

MCTIC. Tecnologias habilitadoras e convergentes. **Ministério da Ciência, Tecnologia, Inovações e Comunicações**, 2018. Disponível em: <http://www.mctic.gov.br/mctic/opencms/tecnologia/tecnologias_convergentes/paginas/tecnologias_conv ergentes_e_habilitadoras/TECNOLOGIAS_CONVERGENTES_E_HABILITADORAS.html>. Acesso em: 28 fevereiro 2018.

MENESES, R. A. **Diagnóstico operacional de sistemas de abastecimento de água: o caso Campina Grande**. UFCG. Campina Grande, p. 161. 2011.

MONTGOMERY, D. C. **Introduction to linear regression analysis**. 5ª. ed. New Jersey: John Wiley & Sons, 2012.

MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. 1ª. ed. São Paulo: Edgard Blücher, 2008.

NAVIDI, W. **Probabilidade e Estatística para Ciências Exatas**. Porto Alegre: AMGH Editora, 2012.

OLIVEIRA, S. D. **Internet das Coisas com ESP8266, Arduino e Raspberry PI**. São Paulo: Novatec, 2017.

RUMSEY, D. **Estatística Para Leigos**. Rio de Janeiro: Alta Books, v. II, 2014.

SAGA. Catálogo completo unijato residencial. **Saga Medição**, 2017. Disponível em: <<http://www.sagamedicao.com.br/download/1008/>>. Acesso em: 20 fevereiro 2018.

SCHWAB, K. **The fourth industrial revolution**. London: Penguin Books, 2016. 192 p.

SEED STUDIO. G1&2" Water Flow Sensor. **Seed Studio**, 2018. Disponível em: <https://www.seeedstudio.com/g12-water-flow-sensor-p-635.html?cPath=84_87>. Acesso em: 28 fevereiro 2018.

SEED STUDIO. NodeMCU v2 - Lua based ESP8266 development kit. **Seed Studio**, 2018. Disponível em: <<https://www.seeedstudio.com/NodeMCU-v2-Lua-based-ESP8266-development-kit-p-2415.html>>. Acesso em: 28 fevereiro 2018.

WOLFRAM. Wolfram Math World: Outlier. **Wolfram**, 2018. Disponível em: <<http://mathworld.wolfram.com/Outlier.html>>. Acesso em: 28 fevereiro 2018.