

Proposta para Identificação e Substituição de Outliers em Dados de Velocidade de Pipeline Inspection Gauges

Gustavo Fernandes de Lima¹, Andrés Ortiz Salazar²

¹Instituto Federal do Rio Grande do Norte (IFRN)
Rua Antônia de Lima Paiva, 155, CEP: 59143-455, Parnamirim, RN

²Departamento de Engenharia de Computação e Automação
Laboratório de Avaliação e Medição de Petróleo
Universidade Federal do Rio Grande do Norte (UFRN), CEP 59078-970, Natal, RN

gustavo.lima@ifrn.edu.br, andres@dca.ufrn.br

Abstract. *The speed measurement of a Pipeline Inspection Gauge suffer with noises and eventually some readings extrapolate average of the others. The objective this work is to present a computational code implementation to identify and replace outliers, which harms the graphic analysis of speed. A Prototype PIG and Testing Pipeline are used to getting speed data with outliers. The developed code is applied in these data, identifying points above calculated upper limit, successfully. The result show which is possible to identify and replace outliers speed data of the Prototype PIG.*

Resumo. *A medição de velocidade de um Pipeline Inspection Gauge (PIG) sofre com ruídos e eventualmente algumas leituras extrapolam a média das demais. O objetivo deste trabalho é apresentar a implementação de um código computacional para identificar e substituir pontos fora da curva (outlier), que prejudicam a análise de gráficos de velocidade. Um PIG Protótipo e uma Tubulação de Testes são utilizados para obtenção de dados de velocidade com outliers. O código desenvolvido é aplicado nesses dados, identificando e substituindo com sucesso os pontos acima do limite superior calculado. O resultado mostra que é possível, de forma bem sucedida, identificar e substituir outliers de velocidade do PIG Protótipo.*

Introdução

Com o envelhecimento dos dutos industriais cresceu a necessidade de realizar inspeções e avaliações sobre as condições atuais dessas tubulações, com o objetivo de detectar anomalias como vazamentos, corrosão, amassamentos, rachaduras, etc. e, assim, reduzir custos com manutenção. A operação de inspeção é aplicável em dutos da indústria química, de óleo e gás, farmacêutica e de alimentos, sem interromper o fluxo do produto no duto, usando uma técnica conhecida por *pigging*.

O *pigging* é um conceito de manutenção de dutos em que uma ferramenta esférica ou cilíndrica chamada *Pipeline Inspection Gauge* (PIG) percorre toda a extensão da tubulação para limpá-lo e/ou coletar dados para avaliação a respeito do duto. O PIG tem como característica principal a capacidade de ser empurrado pela pressão, atrás dele, produzida pelo fluido em movimento dentro do duto.

Há basicamente dois tipos de PIGs, de acordo com o propósito que o mesmo vai executar no duto, podendo ser: de limpeza (*Cleaning PIG*) ou instrumentado (*Smart PIG*). Os PIGs de limpeza (Fig. 1a) realizam função de limpeza, separação de produtos, remoção de água e desamassamento dos dutos [Diaz 2008]. Os PIGs instrumentados (Fig. 1b) portam sensores e salvam dados como velocidade e distância percorrida.



(a) Limpeza



(b) Instrumentado

Figura 1. Exemplos de PIGs.

Fontes: <http://www.polocs.com.br/> e <http://tecma.gdoit.net/>

Vários pesquisadores desenvolveram trabalhos relacionados aos PIGs. Por exemplo, [Nguyen et al. 2001] modelaram e simularam o controle de fluxo pelo PIG. O método foi eficaz para estimar a posição e velocidade do PIG com uma dada condição operacional. [Sadovnychiy et al. 2006] investigaram o problema de precisão de medição de posição por um odômetro. Os resultados mostram que a medição de posição é influenciada pela velocidade que o PIG atinge. [Sabino 2009] destaca que a resolução de um PIG de perfilagem no sentido longitudinal do duto é definida pela taxa de amostragem e pela velocidade da ferramenta.

Este trabalho segue pesquisas anteriores desenvolvidas na Universidade Federal do Rio Grande do Norte (UFRN). [Freitas et al. 2016] detectaram a passagem de PIGs com transdutores de pressão e estimaram a velocidade média da ferramenta. [Silva et al. 2016] desenvolveram um odômetro para medição de velocidade de PIGs instrumentados. [Lima et al. 2017] validaram um laboratório de testes de PIGs comparando velocidades médias obtidas do PIG e da Tubulação de Testes. [Silva et al. 2018] aplicaram tacômetro magnético e microeletrônica para determinação da velocidade de PIGs.

A medição de velocidade dos PIGs, por vezes, sofre com ruídos provocados pelo movimento inconstante da ferramenta dentro do duto, acarretando o surgimento de medições incorretas, que extrapolam a média dos demais pontos coletados. Tais distorções alteram a interpretação dos gráficos obtidos, dificultando a análise e o projeto de controladores de velocidade para essas ferramentas de inspeção.

O objetivo deste artigo é apresentar a implementação de um código para detecção de pontos fora da curva (*outlier*) dentro de dados de velocidade de um PIG Protótipo. O vetor de velocidade foi obtido de um ensaio experimental do PIG Protótipo percorrendo uma Tubulação de Testes. Depois, os dados foram tratados e os *outliers* detectados, que

foram substituídos por pontos de valor menor. Essa substituição eliminou as variações bruscas do gráfico de velocidade obtido, permitindo uma análise confiável desses dados.

Para a continuação deste trabalho pretende-se, futuramente, realizar um estudo comparativo com outros algoritmos que utilizam outras métricas, levantar o esforço computacional do algoritmo proposto etc.

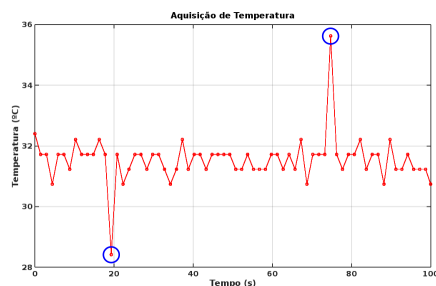
Referencial Teórico

Pontos fora da curva ou *outliers*, de acordo com [Machado 2016] não possuem definição rigorosa, sendo geralmente denominados como valores que se distanciam muito da maioria das observações, sendo um problema na análise de dados por mascarar a amostra.

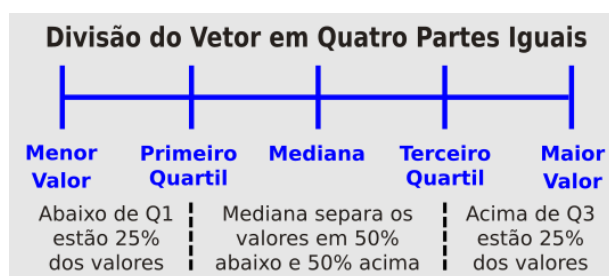
Segundo [Medri 2011] as observações atípicas alteram enormemente as médias e variabilidade dos grupos a que pertencem e podem até mesmo distorcer as conclusões obtidas através de uma análise estatística padrão. A representação de dois pontos atípicos em um gráfico de temperaturas pode ser visualizado na Fig. 2a. Portanto, é de fundamental importância identificar esses *outliers* e dar um tratamento adequado a elas.

Dentre as possíveis causas do aparecimento de *outliers*, pode citar as seguintes [Medri 2011]: leitura, anotação ou transição incorreta dos dados; erro na execução do experimento ou na tomada da medida; e mudanças não controláveis nas condições experimentais ou dos pacientes.

Um dos métodos para identificação de *outliers*, segundo [Andriotti 2005], é com a utilização de estatística robusta (mediana, *interquartile range*). Neste caso, chamando-se **Q3** ao terceiro quartil, **Q1** ao primeiro quartil, e **IQR** à diferença (**Q3 - Q1**), seriam classificados como *outliers* aos valores inferiores a (**Q1 - 1,5 · IQR**) e superiores a (**Q3 + 1,5 · IQR**). Este método é conhecido por “Regra do 1,5 · IQR”, ela afirma que um dado é um *outlier* quando ele é maior que **1,5 · IQR** acima do terceiro quartil (**Q3**) ou abaixo do primeiro quartil (**Q1**). Uma ilustração do primeiro quartil, da mediana e do terceiro quartil dentro de um vetor de dados é apresentada na Fig. 2b.



(a) Representação de *outliers*.



(b) Divisão de um vetor de dados.

Figura 2. Pontos fora da curva.

Os passos da utilização da “Regra do 1,5 · IQR” para identificar *outliers* são: (a) ordenar de forma crescente os dados, (b) encontrar a mediana dos dados, o primeiro quartil e o terceiro quartil, (c) calcular o intervalo interquartil (**IQR**), (d) calcular **Q1 - 1,5 · IQR** abaixo do primeiro quartil e verificar a existência de *outliers* e (e) calcular **Q3 + 1,5 · IQR** acima do terceiro quartil e verificar a existência de *outliers*.

O Código 1 mostra a rotina computacional desenvolvida em *Python*TM 2.7.12 rodando no sistema *Ubuntu*TM 16.04 LTS, com base nos passos acima.

```

1  clc; % limpa tela
2  veloc=csvread('dados.txt'); % dados
3  notas=sort(veloc'); % ordena crescente
4  comprim=length(notas); % tamanho vetor
5  meio=round(comprim/2); % arredonda cima
6  mediana=(notas(meio)+notas(meio+1))/2
7  umquarto=fix(comprim/4);
8  primquartil=(notas(umquarto)+notas(umquarto+1))/2 % Q1
9  tresquarto=3*round(comprim/4);
10  tercquartil=(notas(tresquarto)+notas(tresquarto+1))/2 % Q3
11  faixainter=tercquartil-primquartil %IQR
12  liminf=primquartil-1.5*faixainter
13  quantoutlierinf=sum(notas<liminf)
14  for i=1:quantoutlierinf
15     notas(i) % mostra outliers baixos
16  end;
17  limisup=tercquartil+1.5*faixainter
18  quantoutliersup=sum(notas>limisup)
19  for i=comprim:-1:comprim-quantoutliersup+1
20     notas(i) % mostra outliers altos
21  end;

```

Código 1. Identificação e substituição de outliers.

Além de identificar e apresentar os *outliers*, o código também os substitui (tratamento adequado) por pontos “novos” cujos valores são a média entre o antecessor e o sucesso do ponto considerado fora da curva.

Metodologia

PIG Protótipo e Eletrônica Embarcada

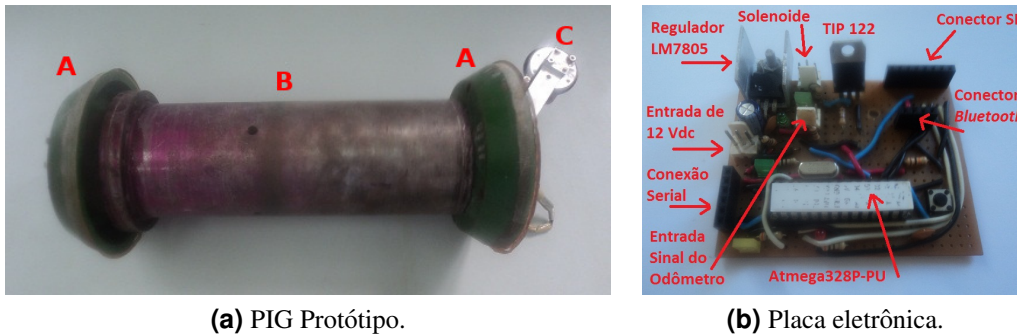
O PIG Protótipo é o equipamento projetado para realização dos ensaios experimentais dentro da Tubulação de Testes. Esse PIG é constituído de dois discos de borracha (A), uma cápsula metálica (B) e um tacômetro (C), conforme a Fig. 3a. Os discos centralizam o PIG Protótipo dentro do duto; a cápsula metálica possui espaço interno para a placa eletrônica (Fig. 3b) e baterias; e o tacômetro produz sinais para o cálculo de velocidade.

A placa eletrônica é responsável por ler o sinal do tacômetro e calcular a velocidade do PIG Protótipo. Os dados são salvos em um cartão de memória SD. A placa conta com um microcontrolador Atmega328P-PU, um regulador de tensão LM7805 e soquete para encaixe de um módulo de cartão.

A placa eletrônica consome energia de um pacote de baterias composto por 3 células de 3,7 Vdc de Lítio Íon ligadas em série para fornecer um total de 11,1 Vdc. São células reaproveitadas de baterias de computadores portáteis.

Tubulação de Testes

A Tubulação de Testes com 55 m de extensão é o equipamento construído para realização de ensaios experimentais com o PIG Protótipo. Ela é formada por dutos de $\phi 6''$ de



(a) PIG Protótipo. (b) Placa eletrônica.

Figura 3. PIG Protótipo e sua eletrônica embarcada.

diâmetro em aço carbono, unidas por flanges, no percurso principal e dutos de $\phi 8''$ de diâmetro no Lançador e no Recebedor, conforme a Fig. 4.

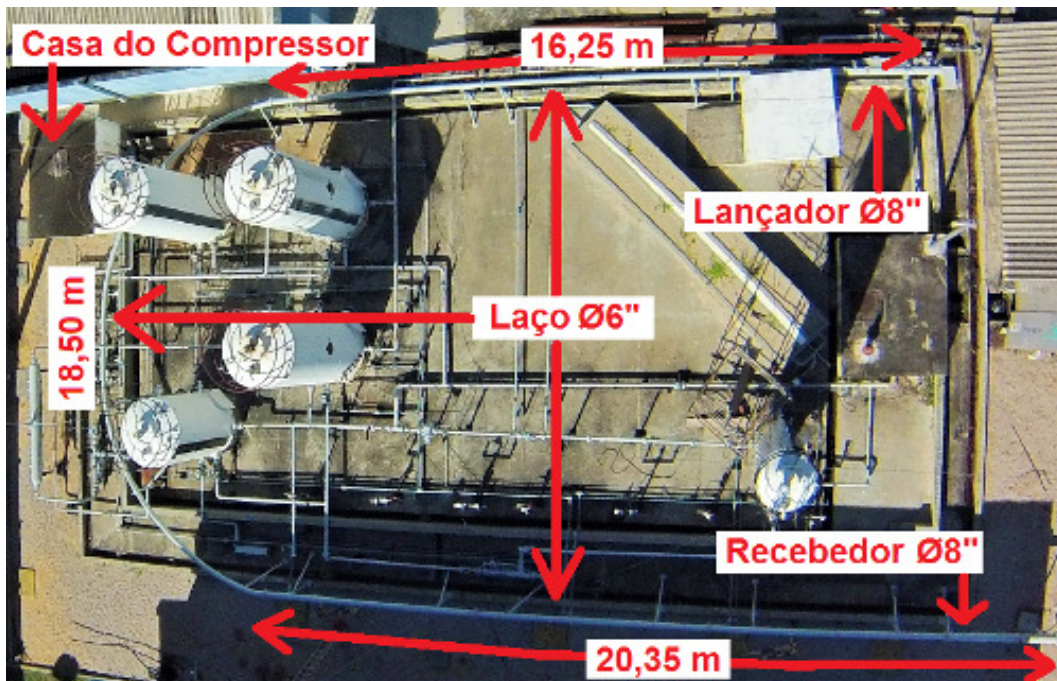


Figura 4. Tubulação de teste vista do alto.

A obtenção de dados de velocidade para testar o código proposto passa por operar uma corrida do PIG Protótipo dentro da Tubulação de Testes. Para isso, o PIG Protótipo é colocado no Lançador. O compressor de ar é ligado e o ar comprimido injetado atrás do PIG Protótipo, empurrando-o para frente. O atrito dos discos de borracha com a parede do duto produz um ruído audível, que permite saber quando o PIG Protótipo chega no Recebedor. Depois, o PIG Protótipo é retirado do Recebedor, desmontado e o cartão de memória recuperado para leitura dos dados.

Um exemplo de gráfico de espaço e velocidade do PIG Protótipo é apresentado na Fig 5. O PIG Protótipo percorre uma distância de $25,09\text{ m}$ até $62,19\text{ m}$, quando chega no Recebedor. A duração do movimento é inferior a 5 s . A velocidade do PIG Protótipo apresenta variações de até 80 m/s , um possível *outlier*.

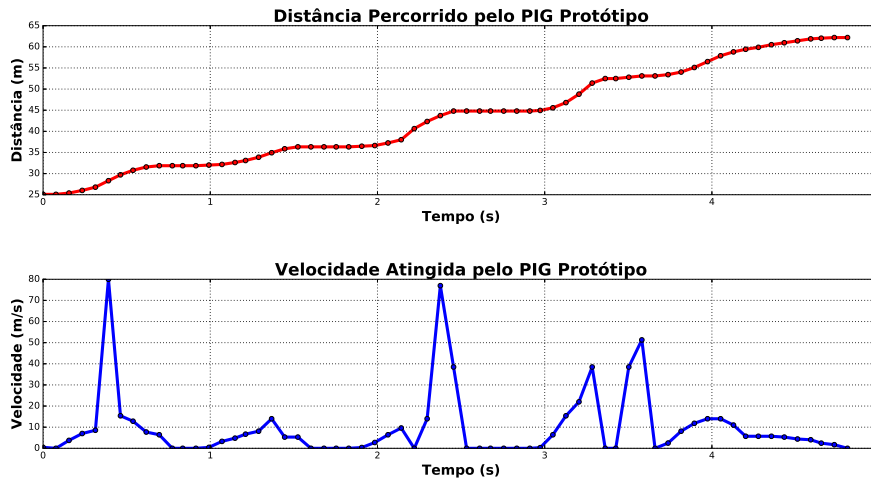


Figura 5. Gráfico da distância e velocidade.

Foram realizados 5 corridas do PIG Protótipo para obtenção de dados com possíveis *outliers*. A existência de pontos distantes dos demais provoca um mascaramento de outras regiões de possível interesse para análise. O conjunto de pontos escolhido apresentou um mascaramento de análise de velocidade em $t = 4s$. Onde ocorre uma diminuição de velocidade de quase $15 m/s$ para $5 m/s$, mas que não é evidente.

No gráfico acima é destaque os picos de 80 e $76,97 m/s$, enquanto que pontos abaixo de $20 m/s$ são ignorados. Por causa do mascaramento em $t = 4s$ e dos dois picos de velocidade, este gráfico é utilizado para testar a existência de *outliers*. Os outros conjuntos de dados apresentaram *outliers*, mas não mostraram mascaramento de dados.

Resultados e Discussões

Com base nos dados de velocidade acima e aplicando o código desenvolvido é apresentada a tela com a quantidade de pontos fora da curva e seus valores, conforme Cód. 2.

```

1 mediana = 5.0600
2 primquartil = 0
3 tercquartil = 10.310
4 faixa = 10.310
5 liminf = -15.465
6 outlierinf = 0
7 limisup = 25.775
8 outliersup = 6
9 =====
10      k          T(k)      veloc(k)
11      6.00000    0.39100    80.00000
12      32.0000    2.3760    76.9700
13      33.0000    2.4550    38.4800
14      44.0000    3.2840    38.4800
15      47.0000    3.5030    38.4800
16      48.0000    3.5810    51.3100
17 =====
18 >>>

```

Código 2. Resultado do código de *outliers*.

Nela, são apresentados valores relevantes dentro da teoria de *Outliers*, alguns deles são: mediana igual $5,06 \text{ m/s}$, primeiro quartil igual a zero, terceiro quartil igual a $10,31 \text{ m/s}$, limite superior $25,775 \text{ m/s}$ e quantidade de *outliers* superiores igual a 6.

Na parte inferior da tela são mostrados os 6 *outliers*, dentro dos dados de velocidade, onde ‘k’ é a posição dentro do vetor de dados, ‘T(k)’ é o valor de tempo e ‘veloc(k)’ é o valor de velocidade acima de $30,838 \text{ m/s}$ (limite superior).

Os 6 *outliers* identificados são substituídos dentro do vetor por outro valor, sendo a média entre o antecessor e o sucessor do referido *outlier*. Com isso, o gráfico de velocidade não apresenta variações bruscas, conforme a Fig. 6, onde os dados originais estão na cor vermelha e os dados “novos” estão na cor azul.

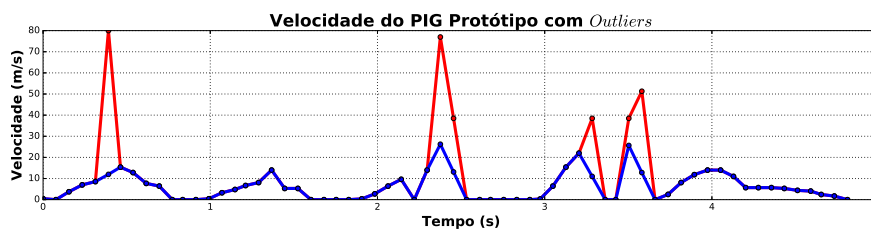


Figura 6. Sobreposição dos dados.

Com a substituição dos *outliers* (cor vermelha) o “novo” gráfico é apresentado na Fig. 7. Nele, agora, as maiores velocidades registradas são de $26,235$ e $25,655 \text{ m/s}$, sendo consideradas aceitáveis. E o mascaramento em $t = 4 \text{ s}$ não existe mais, ficando claro a diminuição de velocidade sofrida pelo PIG Protótipo.

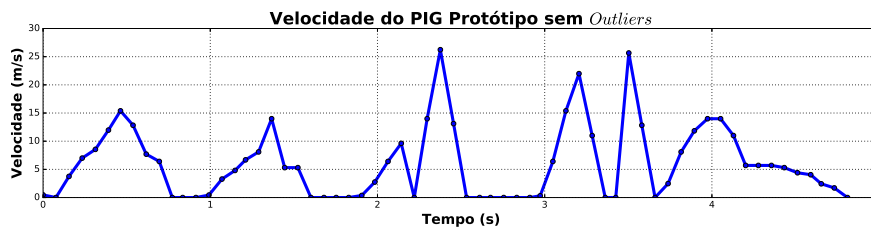


Figura 7. “Novo” Gráfico de Velocidade.

Por fim, sem os *outliers*, os dados de velocidade ficam melhores de serem analisados, uma vez que a maioria dos pontos está dentro de uma faixa curta (de 0 a 30 m/s) ao contrário do gráfico inicial, que distorcia as análises.

Conclusão

O objetivo deste trabalho é apresentar o estudo e a implementação de um código para identificação e substituição de *outliers* nos dados de velocidade do PIG Protótipo, utilizando a “Regra do $1,5 \cdot \text{IQR}$ ” implementado em *Python*TM no sistema *Ubuntu*TM.

Um PIG Protótipo e uma Tubulação de Testes são utilizados para obtenção de dados de velocidade. Os dados iniciais apresentam potenciais *outliers*, de modo que o código é utilizado e identifica 6 *outliers*, com sucesso. Na sequência os 6 *outliers* são substituídos por “novos” pontos de menor valor. Ao final, o gráfico “novo”, sem os *outliers* iniciais, mostra valores de velocidade menores de 30 m/s .

Com base no resultado apresentado é possível concluir que o código desenvolvido é um sucesso. A principal contribuição deste trabalho é propor e apresentar um código capaz de identificar e substituir *outliers* de dados utilizando ferramentas *open sources*, ficando o código disponível para professores, pesquisadores e estudantes que precisam excluir *outliers* de seus dados.

Agradecimento

O primeiro autor agradece ao Instituto Federal do Rio Grande do Norte (IFRN) *Campus Parnamirim* pelo afastamento em tempo integral para conclusão do Doutorado.

Referências

- Andriotti, J. L. S. (2005). *Técnicas Estatísticas Aplicáveis a Tratamento de Informações Oriundas de Procedimentos Laboratoriais*. Serviço Geológico do Brasil., Porto Alegre, RS.
- Diaz, J. F. A. (2008). Estudo e projeto conceitual de um robô para inspeção de linhas de serviço. Master's thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.
- Freitas, V. C. G., Lima, G. F., Salazar, A. O., and Maitelli, A. L. (2016). “pig” detection with pressure transducers. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)*, 5(9):7497–7503.
- Lima, G. F., Freitas, V. C. G., Araújo, R. P., Maitelli, A. L., and Salazar, A. O. (2017). PIG's speed estimated with pressure transducers and hall effect sensor : An industrial application of sensors to estimate speed pigs. *Sensors*, 17(9).
- Machado, M. A. D. (2016). *Deteção e Influência de Outliers na Qualidade de Modelos de Relação Hipsométrica sob o Ponto de Vista Preditivo*. Universidade Tecnológica Federal do Paraná., Dois Vizinhos, PR.
- Medri, W. (2011). *Análise Exploratório de Dados*. Universidade Estadual de Londrina., Londrina, PR. Notas de Aula.
- Nguyen, T. T., Yoo, H. R., Rho, Y. W., and Kim, S. B. (2001). Speed control of PIG using bypass flow in natural gas pipeline. *IEEE International Symposium on Industrial Electronics*, pages 863 –868.
- Sabino, J. M. (2009). Avaliação por pig de perfilagem de danos superficiais nos materiais das paredes de dutos de petróleo. Master's thesis, Universidade Federal do Rio Grande do Norte, Natal, RN.
- Sadovnychiy, S., López, J., Ponomaryov, V., and Sadovnychy., A. (2006). Evaluation of distance measurement accuracy by odometer for pipelines pigs. *Journal of the Japan Petroleum Institute*, pages 38–42.
- Silva, A. E. M., Lima, G. F., and Salazar, A. O. (2016). Desenvolvimento de tacômetro magnético para medição de velocidade em pigs instrumentos. *II Congresso Nacional de Engenharia de Petróleo, Gás Natural e Biocombustíveis (Conepetro)*.
- Silva, C. Y. F., da Silva, W. L. A., de Lima, G. F., Maitelli, A. L., and Salazar, A. O. (2018). Aplicação de tacômetro magnético e microeletrônica para determinação da velocidade de PIGs. *X Congresso Nacional de Engenharia Mecânica (CONEM)*.