

ESTUDO E IMPLEMENTAÇÃO DE CÓDIGO PARA IDENTIFICAÇÃO E SUBSTITUIÇÃO DE *OUTLIERS* EM DADOS DE VELOCIDADE DE *PIPELINE INSPECTION GAUGES* (PIGs)

LIMA, G. F.¹ e SALAZAR, A. O.²

¹ Professor Me. do Curso Técnico em Mecatrônica – IFRN – Campus Parnamirim – Parnamirim
gustavo.lima@ifrn.edu.br

² Professor Dr. do Departamento de Engenharia de Computação e Automação – UFRN – Natal
andres@dca.ufrn.br

ÁREA TEMÁTICA: Engenharias.

RESUMO

A medição de velocidade de um *Pipeline Inspection Gauge* (PIG) sofre com ruídos e por vezes algumas leituras extrapolam a média das demais. O objetivo deste trabalho é apresentar a implementação de um código (rotina computacional) para identificar e substituir pontos fora da curva (*outlier*), que prejudicam a análise de gráficos de velocidade de PIGs. Um PIG Protótipo e uma Tubulação de Testes são utilizados para obtenção de dados de velocidade com potenciais *outliers* presentes. O código

desenvolvido é aplicado nesses dados, identificando com sucesso 6 pontos acima do limite superior calculado. Estes *outliers* são substituídos por pontos de menor valor e, assim, o gráfico é mantido dentro de uma faixa curta de valores (0 a 30 m/s). O resultado mostra que é possível, de forma bem sucedida, identificar e substituir *outliers* de dados de velocidade do PIG Protótipo, de modo a melhorar o entendimento sobre seu movimento dentro da Tubulação de Testes.

PALAVRAS-CHAVE: Estatística. Medição atípica. Atmega328P-PU. Lítio íon. Pressão.

ABSTRACT

The speed measurement of a Pipeline Inspection Gauge suffer with noise and sometimes some readings extrapolate average of the others. The objective this work is to present the code (computational routine) implementation to identify and replace outliers, which harms the graphic analysis of PIG's speed. A Prototype PIG and Testing Pipeline are using to getting speed data with outliers potential. The

developed code is applied in these data, identifying 6 points above calculated upper limit, successfully. This outliers are replace by points of lesser values and, thus, the graphic is held within a short range of values (0 a 30 m/s). The result show which is possible to identify and replace outliers speed data of Prototype PIG, so to improve understanding about PIG's movement within Testing Pipeline.

KEYWORDS: Statistics. Atypical measurement. Atmega328P-PU. Lithium ion. Pressure.

1 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 PIGs*) ou instrumentado (*Smart PIGs*). 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. Possuem ainda um preciso sistema de localização que registra a posição dos defeitos encontrados ao longo da operação (PEREIRA et al, 2013).



(a)



(b)

Figura 1 – PIGs quanto a sua aplicação. (a) De limpeza; (b) Instrumentado.

Fontes – <http://www.polocs.com.br/wp-content/uploads/2016/07/PIG-de-limpeza.png>
<http://tecma.gdoit.net/wp-content/uploads/2015/02/MFL4.jpg>

Uma pesquisa na literatura revela alguns trabalhos sobre PIGs instrumentados e a importância da medição de velocidade e espaço percorridos por essas ferramentas. 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. Freitas et al. (2016) detectaram a passagem de PIGs com transdutores de pressão e estimaram a velocidade média da ferramenta. Silva, Lima e Salazar (2017) 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.

Com base na revisão bibliográfica acima fica claro a necessidade da medição correta e confiável da velocidade atingida pelos PIGs durante a inspeção. Tal medição, por vezes, sofre com ruídos provocados pelo movimento inconstante do PIG dentro do duto, acarretando o surgimento de medições incorretas, que extrapolam a média dos demais pontos coletados.

O objetivo deste artigo é apresentar o estudo e a implementação de um código (rotina computacional) 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, durante o tratamento dos dados surgiram *outliers*, que foram substituídos por pontos de valor menor. Essa substituição eliminou variações bruscas no gráfico de velocidade obtido.

As próximas seções deste artigo estão organizadas da seguinte maneira: na seção 2 é explicada a teoria sobre pontos fora da curva ou *outliers*; na seção 3 são apresentados os materiais para coleta de dados de velocidade de PIGs; na seção 4 são apresentados os resultados da identificação e substituição de *outliers* nos dados de velocidade; por fim, na seção 5 estão as considerações finais deste artigo.

2 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. Portanto, é de fundamental importância detectar e dar um tratamento adequado a elas.

A Figura 2 mostra um gráfico de dispersão com um ponto (atípico) muito longe dos demais.

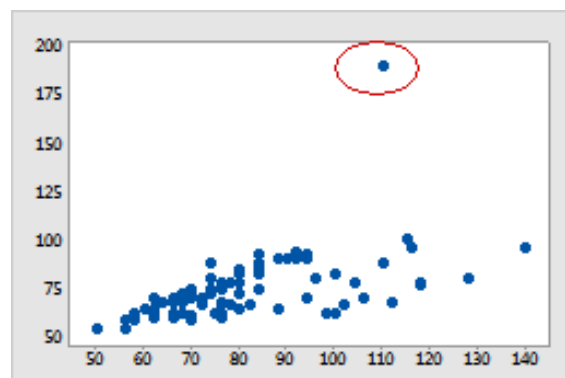


Figura 2 – Exemplo de *outlier* utilizando gráfico de dispersão.

Fonte – https://support.minitab.com/pt-br/minitab/18/outlier_scatterplot.png.

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, *inter quartile 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**).

O método proposto acima também 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 ou abaixo do primeiro quartil. Em outras palavras, *outliers* baixos estão abaixo de **Q1 – 1,5 · IQR** e *outliers* altos estão acima de **Q3 + 1,5 · IQR**.

A Figura 3 ilustra o que representa o 1º quartil **Q1**, a mediana **Q2** e o 3º quartil **Q3** para os dados dentro de um vetor.

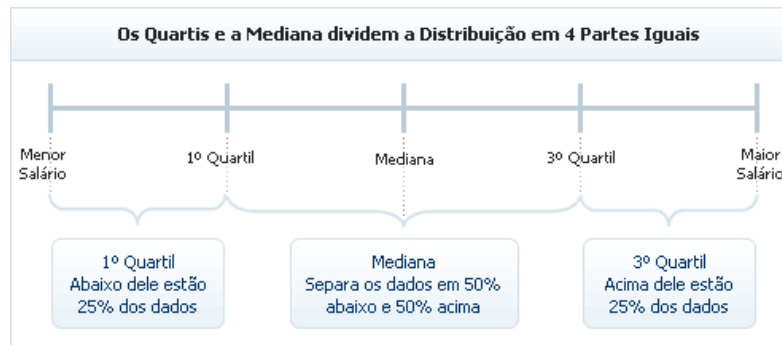


Figura 3 – Representação dos quartis e mediana dentro de um vetor de dados.

Fonte – <https://www.catho.com.br/salario/img/quartis.gif>.

Os passos da utilização da “Regra do $1,5 \cdot \text{IQR}$ ” para identificar *outliers* são:

1. Ordenar de forma crescente os dados de interesse;
2. Encontrar a mediana dos dados, o primeiro quartil e o terceiro quartil;
3. Calcular o intervalo interquartil (ou *inter quartile range*);
4. Calcular $Q1 - 1,5 \cdot \text{IQR}$ abaixo do 1º quartil e verificar a existência de *outliers*;
5. Calcular $Q3 + 1,5 \cdot \text{IQR}$ acima do 3º quartil e verificar a existência de *outliers*.

O Código 1 desenvolvido no programa GNU/Octave (version 4.2.2) rodando no sistema operacional Ubuntu 16.04 LTS, com base nos passos acima, é apresentado abaixo:

```

clc; % limpa tela
veloc=csvread('vetordados.txt'); % recebe dados velocidade
notas=sort(veloc); % ordena dados forma crescente
comprim=length(notas); % quant posições do vetor
resto=rem(comprim,2) % apresenta valor resto (par ou não)
meio=round(comprim/2); % arredonda para cima
media=sum(notas)/comprim % apresenta valor da media
mediana=(notas(meio)+notas(meio+1))/2 % apresenta valor mediana
umquarto=fix(comprim/4); % arredonda para baixo
primquartil=(notas(umquarto)+notas(umquarto+1))/2 % Q1
tresquarto=3*round(comprim/4);
tercquartil=(notas(tresquarto)+notas(tresquarto+1))/2 % Q3
faixainter=tercquartil-primquartil % faixa interquartil IQR
liminf=primquartil-1.5*faixainter % limite inferior
quantoutlierinf=sum(notas<liminf) % total de outliers inferiores
for i=1:quantoutlierinf
    notas(i) % apresenta o outliers baixos
end;
limisup=tercquartil+1.5*faixainter % limite superior
quantoutliersup=sum(notas>limisup) % total de outliers superiores
for i=comprim:-1:comprim-quantoutliersup+1
    notas(i) % apresenta os outliers altos
end;
  
```

Código 1 – Implementação de código para identificação de *outliers* nos dados de velocidade.

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

3 METODOLOGIA

Nesta seção são apresentados os principais materiais utilizados para obtenção de dados experimentais de velocidade de PIG a serem utilizados com o código de identificação e substituição de pontos fora da curva (*outliers*).

3.1 PIG Protótipo

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 de 6 polegadas cada (A), uma cápsula metálica (B) e um odômetro (C). Os discos centralizam o PIG dentro do duto, a cápsula metálica tem dimensões de 12,6 cm de diâmetro e 26,38 cm de comprimento e o odômetro emite pulsos elétricos para o cálculo de velocidade e espaço percorrido.

A Figura 4 apresenta o PIG Protótipo construído para os ensaios experimentais.

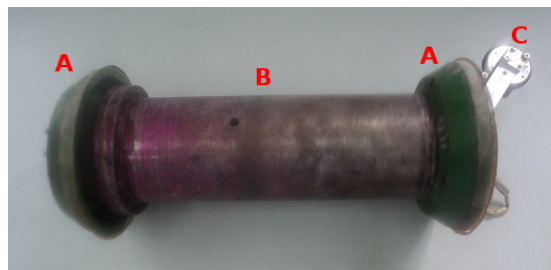


Figura 4 – PIG Protótipo.

A cápsula metálica tem espaço para acomodar a placa eletrônica, um pacote de baterias e uma eletroválvula pneumática 3/2 vias normalmente fechado.

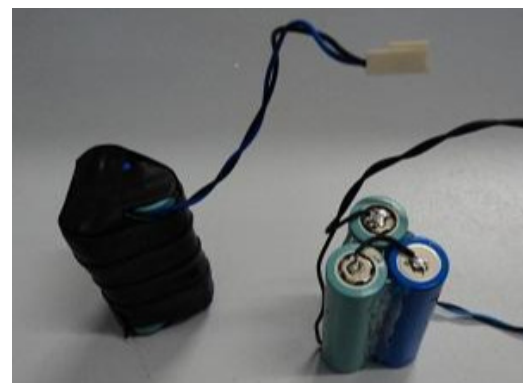
3.2 Placa eletrônica e baterias

A placa eletrônica é responsável por ler o sinal do odômetro e calcular a velocidade e o espaço percorrido pelo PIG Protótipo. Esses dados são salvos em um cartão de memória SD. A placa conta com um microcontrolador Atmega328P-PU, um regulador de tensão LM7805, soquete para entrada de sinal do odômetro, soquete de comunicação serial para programação do microcontrolador, soquete para encaixe de um módulo de cartão de memória SD.

As Figuras 5a e 5b mostram, respectivamente, a placa eletrônica montada para este projeto com identificação dos principais componentes e um detalhe do pacote de baterias.



(a)



(b)

Figura 5 – Componentes da cápsula metálica. (a) Placa eletrônica; (b) Baterias.

O pacote de baterias é 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 *notebooks*.

3.3 Tubulação de Testes

A Tubulação de Testes é o equipamento construído para realização de ensaios experimentais com o PIG Protótipo. Ela é formada por dutos de 6 polegadas no percurso principal e dutos de 8 polegadas no seu início (Lançador de PIG) e no seu final (Recebedor de PIG).

A Figura 6 apresenta uma vista aérea da Tubulação de Testes montada no Laboratório de Avaliação e Medição em Petróleo (LAMP).

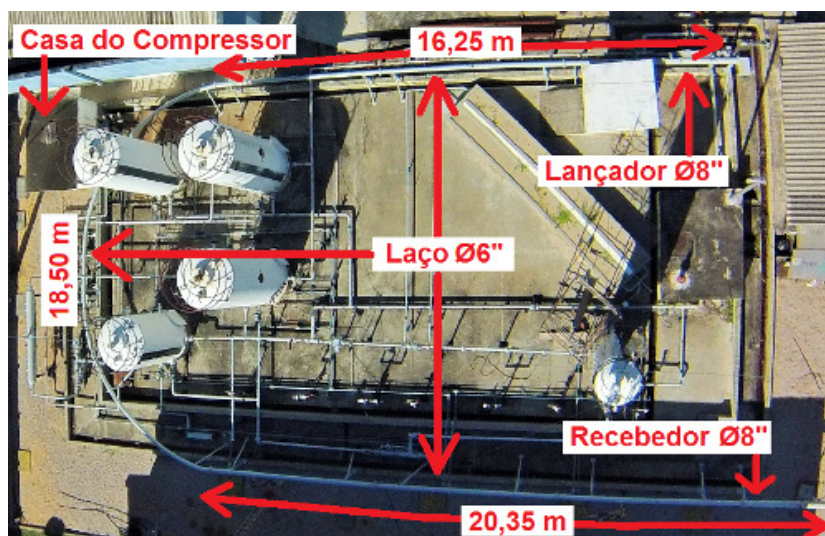


Figura 6 – Tubulação de teste vista do alto.

Ainda sobre a Tubulação de Testes são destaques o Lançador de PIG de 8 polegadas, por onde o PIG Protótipo é inserido do duto, e o Recebedor de PIG de 8 polegadas, por onde o PIG Protótipo é retirado do duto.

As Figuras 7a e 7b apresentam o Lançador e o Recebedor em vista lateral.

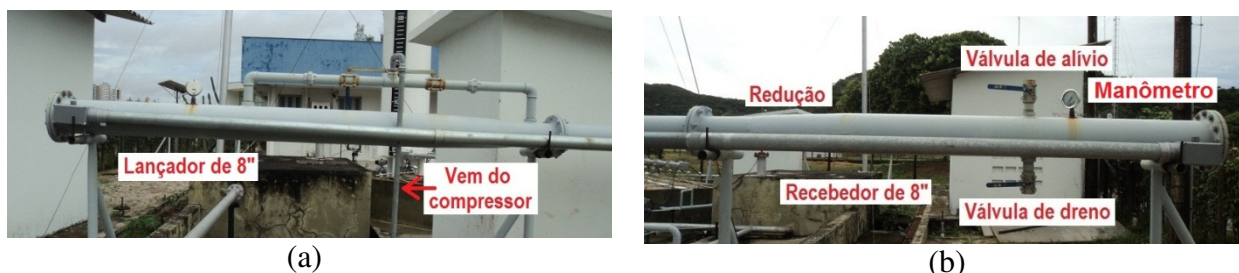


Figura 7 – Vista lateral. (a) Lançador de PIG; (b) Recebedor de PIG.

O funcionamento básico da Tubulação de Testes é ilustrado no diagrama de blocos da Fig. 8.



Figura 8 – Sequência de funcionamento da Tubulação de Testes.

Para operar a Tubulação de Testes com o PIG Protótipo é necessário cumprir quatro passos. Para colocar o PIG Protótipo no Lançador (1) é preciso retirar a tampa de 8 polegadas, colocar o PIG Protótipo dentro e repor a tampa para fechar o Lançador. Pressurizar a Tubulação de Testes (2) significa ligar o compressor de ar e injetar ar comprimido atrás do PIG Protótipo para que este seja empurrado para frente. Isto consiste em movimentar o PIG Protótipo pela Tubulação de Testes (3) percorrendo 60 m, aproximadamente. E ao final para retirar o PIG Protótipo do Recebedor (4) é necessário abrir a tampa de 8 polegadas, puxar o PIG Protótipo para fora e repor a tampa para fechar o Recebedor.

Ao concluir a operação acima, o PIG Protótipo é transportado até uma bancada, onde é desmontado, limpo e o cartão de memória SD é recuperado para obtenção dos dados de velocidade e espaço percorrido pelo PIG Protótipo dentro da Tubulação de Testes.

A Figura 9 mostra um exemplo de gráfico de espaço e velocidade do PIG Protótipo. Nela o espaço percorrido (curva vermelha) pelo PIG Protótipo é de 25,09 m até 62,19 m, quando chega no Recebedor, e com duração de 1.382,393 s até 1.387,048 s. E a velocidade (curva azul) apresenta variações de até 80 m/s, fortemente um candidato a ponto fora da curva (*outlier*).

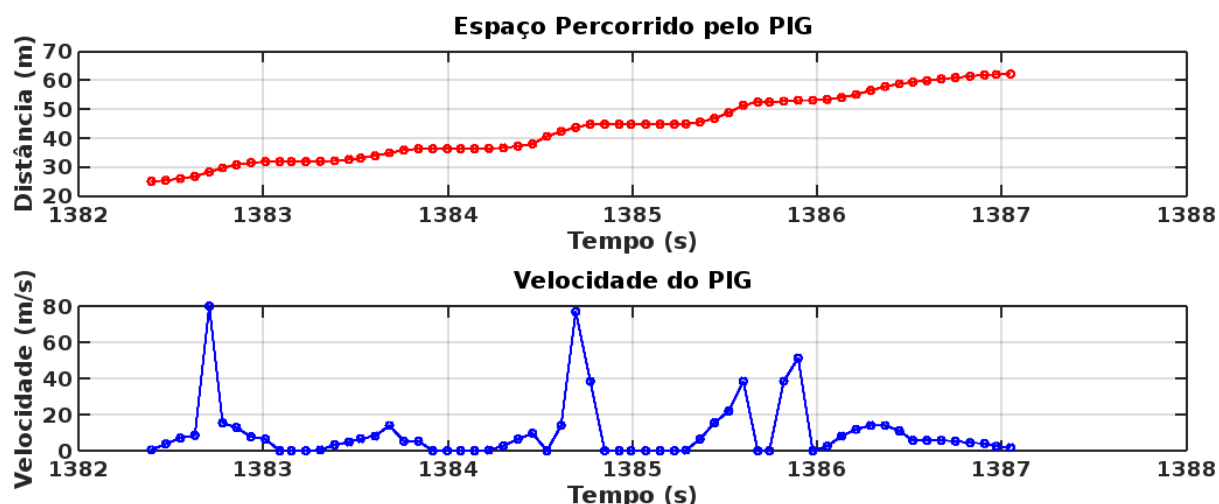


Figura 9 – Exemplo de gráfico com os dados de distância e velocidade do PIG Protótipo.

A existência de pontos distantes dos demais provoca um mascaramento de outras regiões de possível interesse para análise. Inicialmente, o gráfico acima chama a atenção do leitor para os valores próximos de 80 e 76,97 m/s, enquanto que pontos abaixo de 20 m/s são deixados de lado, uma vez que, nesta escala, esses pontos estão próximo de zeros sem mostrar muita importância. Por causa de dois pontos em destaque esse gráfico é utilizado para testar a existência ou não de *outliers*.

4 RESULTADOS E DISCUSSÕES

Com base nos dados de velocidade acima e aplicando o código de identificação e substituição de *outliers* é apresentada a tela com a quantidade de pontos fora da curva e seus valores, conforme Fig. 10. Nela, são apresentados valores relevantes dentro da teoria de *Outliers*, alguns deles são: mediana igual 5,31 m/s, primeiro quartil igual a 0, terceiro quartil igual a 12,335 m/s, limite superior 30,838 m/s e quantidade de *outliers* superiores igual a 6 pontos no total.

Na parte inferior da tela são mostrados os 6 pontos considerados fora da curva, 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 m/s (limite superior definido utilizando a teoria de *Outliers*).

```

Janela de Comandos
-----
resto = 0
media = 9.7968
mediana = 5.3100
primquartil = 0
tercquartil = 12.335
faixainter = 12.335
liminf = -18.503
quantoutlierinf = 0
limisup = 30.838
quantoutliersup = 6
-----
k          T(k)          veloc(k)
-----
5.0000    1382.7070    80.0000
31.000    1384.692         76.970
32.000    1384.771         38.480
43.000    1385.600         38.480
46.000    1385.819         38.480
47.000    1385.897         51.310
-----
>> |
-----
Janela de Comandos  Editor  Documentação
  
```

Figura 10 – Resultado do código implementado.

Os 6 pontos detectados são substituídos dentro do vetor de dados por outro valor, sendo a média entre o antecessor e o sucessor do referido ponto. Com isso, o gráfico de velocidade não apresenta variações bruscas de velocidade, conforme a Fig. 11, onde os dados originais estão na cor vermelha e os dados “novos” estão na cor azul.

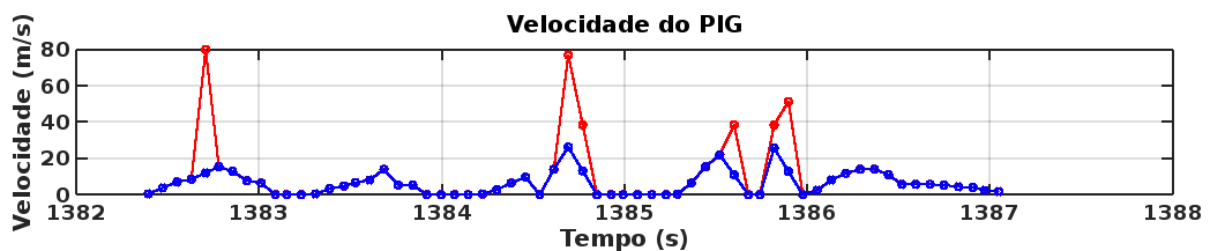


Figura 11 – Sobreposição dos dados depois da utilização do código implementado.

Com a substituição dos *outliers* (cor vermelha) o “novo” gráfico é apresentado na Fig. 12. Nele, agora, as maiores velocidades registradas são de 26,235 e 25,655 m/s, sendo consideradas aceitáveis.

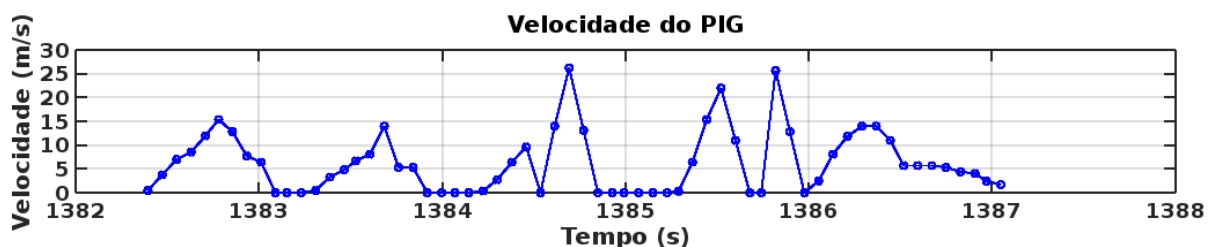


Figura 12 – “Novo” gráfico de velocidade sem os *outliers*.

Outra vantagem, 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, no qual a faixa de valores era longa (de 0 a 80 m/s), que distorcia as análises.

5 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho é apresentar o estudo e a implementação de um código para detecção e substituição de *outliers* nos dados de velocidade do PIG Protótipo, utilizando a “Regra do $1,5 \cdot IQR$ ” implementado no programa GNU/Octave (version 4.2.2) com Ubuntu 16.04 LTS.

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 pontos fora da curva com sucesso. Na sequência os 6 pontos são substituídos por “novos” pontos de menor valor. Ao final, o gráfico “novo” sem os *outliers* mostra valores de velocidade menores de 30 m/s, considerado aceitável.

Com base no resultado apresentado é possível concluir que o código desenvolvido é um sucesso. A principal contribuição deste trabalho é propor e apresentar uma rotina computacional capaz identificar e substituir *outliers* de dados utilizando ferramentas *open sources*.

REFERÊNCIAS

ANDRIOTTI, J. L. S. **Técnicas estatísticas aplicáveis a tratamento de informações oriundas de procedimentos laboratoriais**. 2005. Porto Alegre: CPRM – Serviço Geológico do Brasil, Superintendência Regional de Porto Alegre.

DIAZ, J. F. A. **Estudo e projeto conceitual de um robô para inspeção de linhas de serviço**. 2008. Dissertação (Mestrado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.

FREITAS, V. C. G., LIMA, G. F., SALAZAR, A. O., MAITELLI, A. L. **“PIG” detection with pressure transducers**. 2016. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Vol. 5, No. 9, PP. 7497-7503.

LIMA, G. F., FREITAS, V. C. G., ARAÚJO, R. P., MAITELLI, A. L., SALAZAR, A. O. **PIG’s speed estimated with pressure transducers and hall effect sensor: an industrial application of sensors to validate a testing laboratory**. 2017. Sensors, v. 17, n. 9, p. 2119.

MACHADO, M. A. D. **Detecção e influência de outliers na qualidade de modelos de relação hipsométrica sob o ponto de vista preditivo**. 2016. Trabalho de Graduação. Universidade Tecnológica Federal do Paraná – UTFPR. Dois Vizinhos, PR.

MEDRI, W. **Análise exploratória de dados**. 2011. Notas de aula - Curso de Especialização “Lato Sensu” em Estatística. Universidade Estadual de Londrina. Londrina, PR.

NGUYEN, T. T., KIM, S. B., YOO, H. R., RHO, Y. W. **Modeling and simulation for pig flow control in natural gas pipeline**. 2001. KSME International Journal, Vol. 15 No. 8, pp. 1165-1173.

PEREIRA, C. G., OLIVEIRA, R. R. G, LIMA, G. F., SALAZAR, A. O., MAITELLI, A. L., FONTES, F. A. O. **Proposta de tecnologia para controle de velocidade de pigs instrumentados**. 2013. INTERCON 2013, Trujillo, Perú.

SABINO, J. M. **Avaliação por pig de perfilagem de danos superficiais nos materiais das paredes de dutos de petróleo**. 2009. Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte - UFRN. Natal, RN.

SADOVNYCHYY, S., LÓPEZ, J., PONOMARYOV, V., SADOVNYCHYY, A. **Evaluation of distance measurement accuracy by odometer for pipelines pigs**. 2006. Journal of the Japan Petroleum Institute, v. 49, n. 1, p. 38-42.

SILVA, A. E. M.; LIMA, G. F.; SALAZAR, A. O. **Desenvolvimento de tacômetro magnético para medição de velocidade em pigs instrumentados**. 2017. HOLOS, v. 1, p. 336.

SILVA, C. Y. F., SILVA, W. L. A., LIMA, G. F., MAITELLI, A. L., SALAZAR, A. O. **Aplicação de tacômetro magnético e microeletrônica para determinação da velocidade de PIGs**. 2018. X Congresso Nacional de Engenharia Mecânica – CONEM 2018.