

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

**Gustavo Fernandes de Lima<sup>1</sup> e Andrés Ortiz Salazar<sup>2</sup>**

<sup>1</sup> Professor Me. do Curso Técnico em Mecatrônica – IFRN – *Campus Parnamirim* – Parnamirim

[gustavo.lima@ifrn.edu.br](mailto:gustavo.lima@ifrn.edu.br)

<sup>2</sup> Professor Dr. do Departamento de Engenharia de Computação e Automação – UFRN – Natal

[andres@dca.ufrn.br](mailto:andres@dca.ufrn.br)

ÁREA TEMÁTICA: Engenharias.

# Roteiro

Introdução

Referencial Teórico

Metodologia

Resultados e Discussões

Considerações Finais

Bibliografia

Fim

# Introdução

- ❑ Com o envelhecimento dos dutos industriais cresceu a necessidade de realizar inspeções e avaliações sobre as condições atuais dessas tubulações.
- ❑ Tem por objetivo detectar anomalias como vazamentos, corrosão, amassamentos, rachaduras, etc. e, assim, reduzir custos com manutenção.
- ❑ 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.
- ❑ Há basicamente dois tipos de PIGs, de acordo com sua função dentro do duto, podendo ser: de limpeza (*Cleaning PIGs*) ou instrumentado (*Smart PIGs*).

# Introdução



(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>

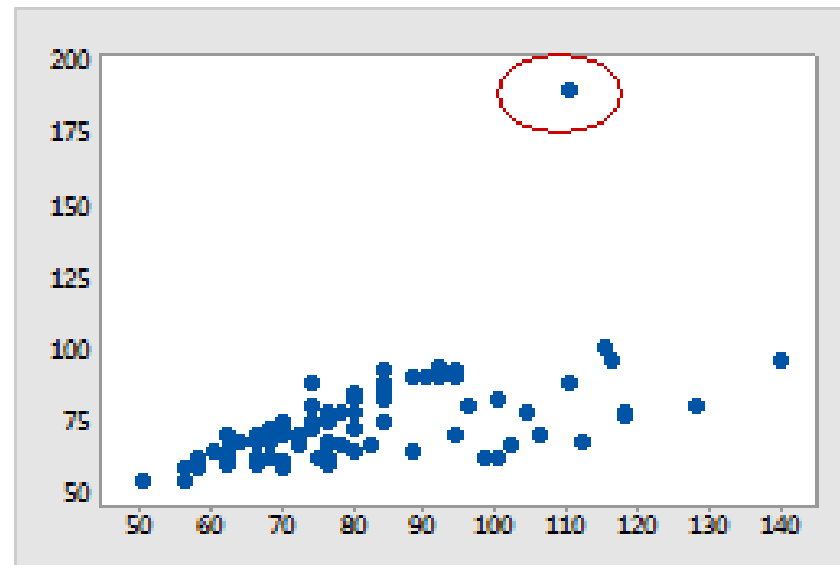
# Introdução

- ❑ Após uma revisão bibliográfica ficou evidente a necessidade da medição correta e confiável da velocidade atingida pelos PIGs durante a inspeção.
- ❑ O movimento inconstante do PIG dentro do duto provoca o surgimento de medições incorretas, que extrapolam a média dos demais pontos coletados.
- ❑ O objetivo deste trabalho é apresentar o estudo e a implementação de um código para detecção de pontos fora da curva (*outlier*) nos dados de velocidade.
- ❑ Durante o tratamento dos dados experimentais surgiram *outliers*, que foram substituídos por pontos de valor menor, eliminando variações bruscas no gráfico.

# Referencial Teórico

- ❑ Pontos fora da curva ou *outliers*, de acordo com [Machado \(2016\)](#), são 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 (...) 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.
- ❑ Algumas causas: 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.

# Referencial Teórico



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

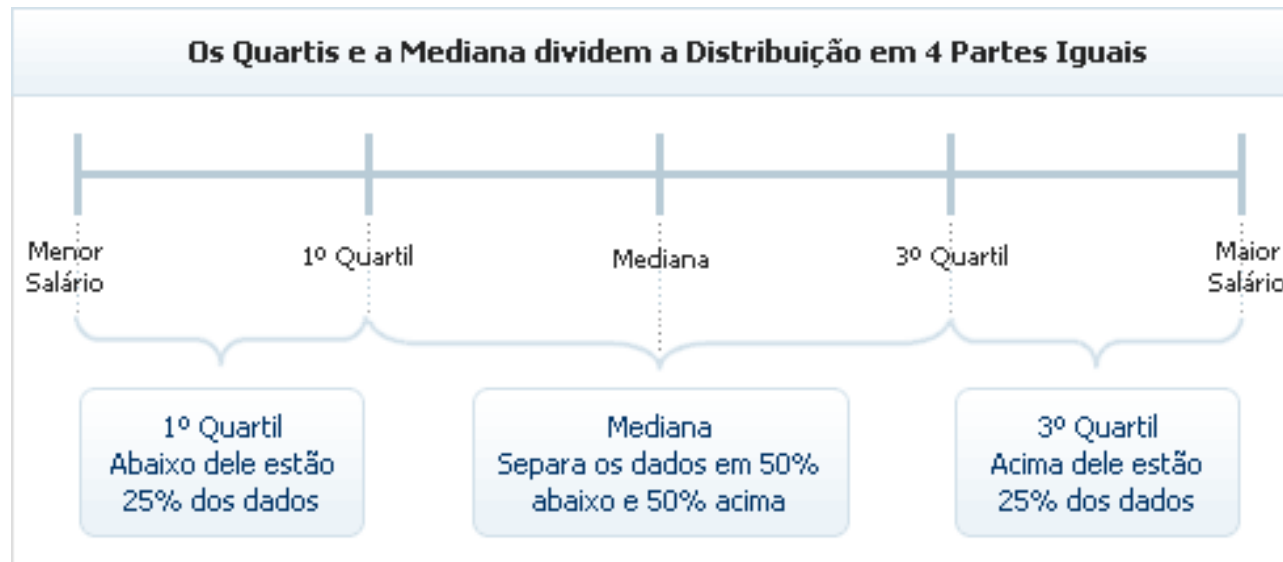
**Fonte** – [https://support.minitab.com/pt-br/minitab/18/outlier\\_scatterplot.png](https://support.minitab.com/pt-br/minitab/18/outlier_scatterplot.png).

# Referencial Teórico

- ❑ Um dos métodos para identificação de *outliers*, segundo [Andriotti \(2005\)](#), é com a utilização de estatística robusta (mediana : **Q2**, *inter quartile range* : **IQR = Q3 – Q1**).
- ❑ O 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**.
- ❑ Em outras palavras, *outliers* baixos estão abaixo de **Q1 – 1,5 · IQR** e *outliers* altos estão acima de **Q3 + 1,5 · IQR**.



# Referencial Teórico



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

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

# Referencial Teórico

- ❑ Os passos da utilização da “Regra do **1,5 · IQR**” para identificar *outliers* são:
  - Ordenar de forma crescente os dados de interesse;
  - Encontrar a mediana dos dados, o primeiro quartil e o terceiro quartil;
  - Calcular o intervalo interquartil (ou *inter quartile range*);
  - Calcular **Q1 – 1,5 · IQR** abaixo do 1º quartil e verificar a existência de *outliers*;
  - Calcular **Q3 + 1,5 · IQR** acima do 3º quartil e verificar a existência de *outliers*.

# Referencial Teórico

❑ O Código 1 desenvolvido no programa GNU/Octave (version 4.2.2) rodando no sistema operacional Ubuntu 16.04 LTS é apresentado abaixo:

```
clc; % limpa tela
veloc=csvread('vetordados.txt'); % recebe dados velocidade
notas=sort(veloc'); % ordena dados forma crescente
comprim=length(notas); % quant posicoes 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 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 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 outliers superior
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.

# Metodologia

## ❑ PIG Protótipo

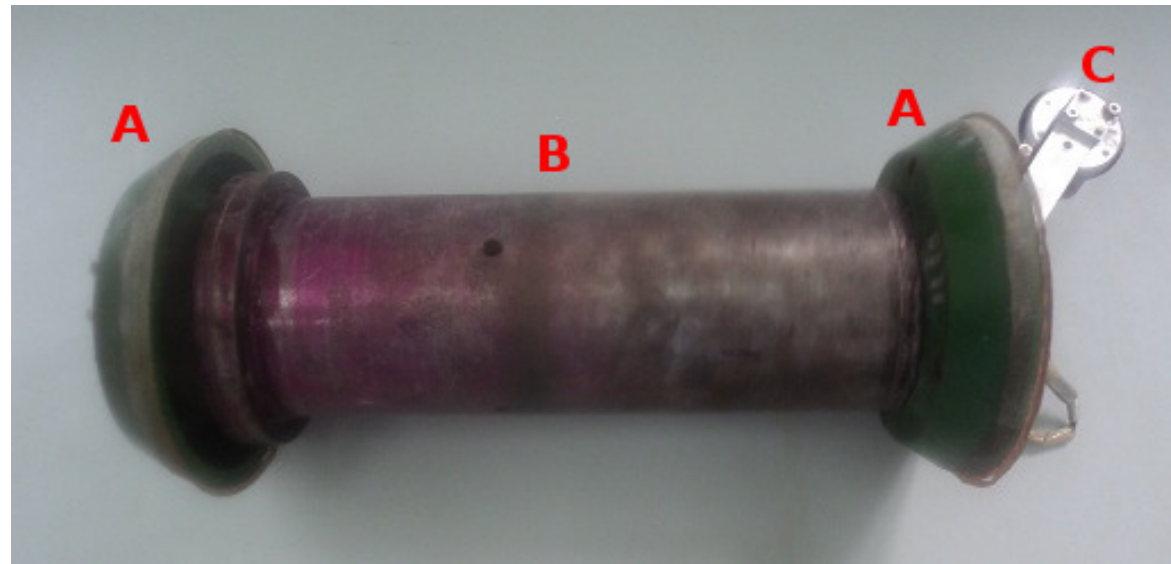
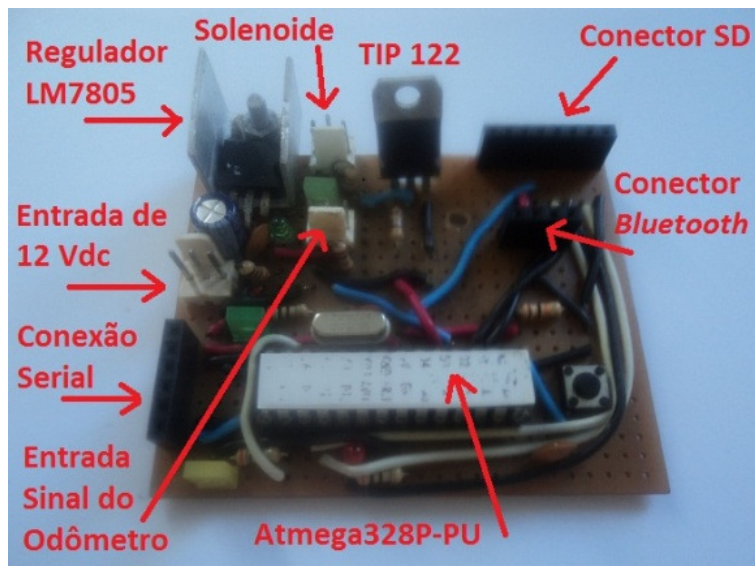


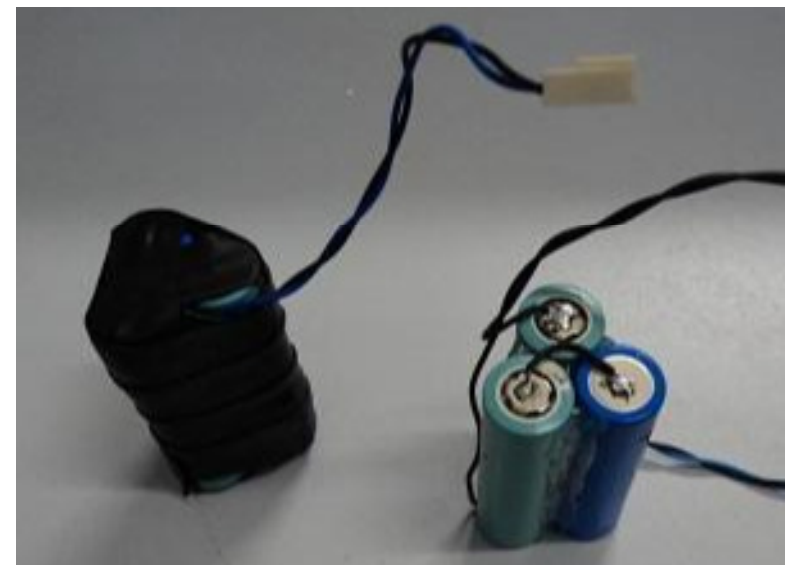
Figura 4 – PIG Protótipo.

# Metodologia

## ❑ Placa eletrônica e baterias



(a)



(b)

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

# Metodologia

## ❑ Tubulação de Testes

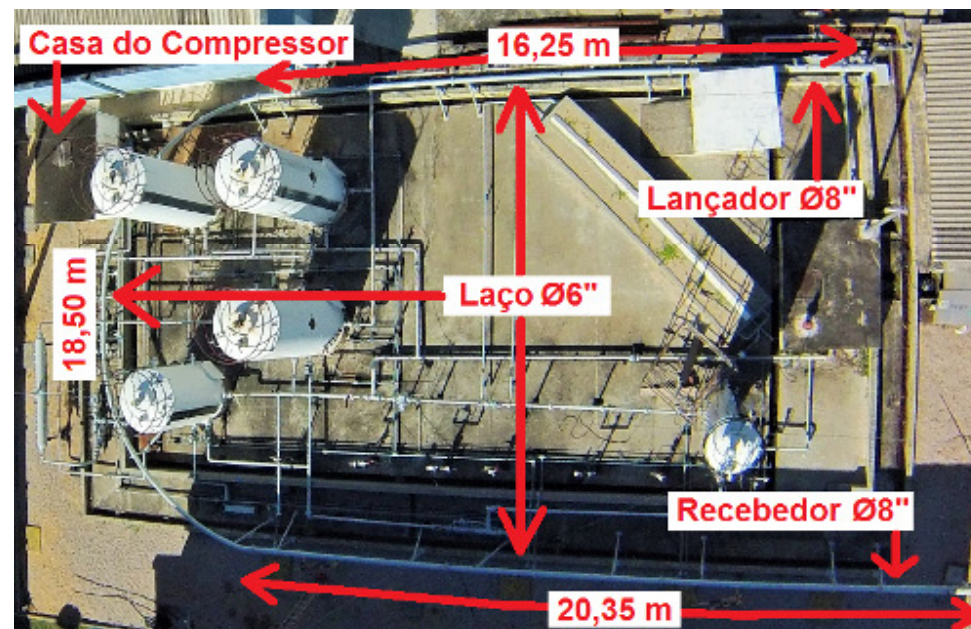


Figura 6 – Tubulação de teste vista do alto.

# Metodologia

## ❑ Lançador e Receptor de PIGs



(a)

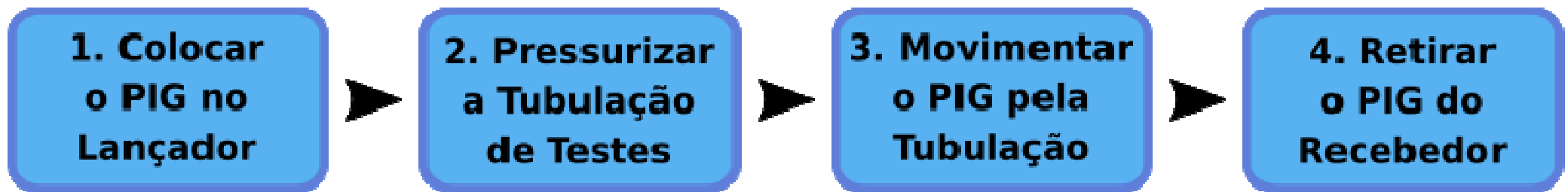


(b)

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

# Metodologia

## ❑ Procedimento para Obtenção de Dados de Velocidade



**Figura 8** – Sequência de funcionamento da Tubulação de teste.



# Metodologia

## ❑ Procedimento para Obtenção de Dados de Velocidade



(a)



(b)



(c)

**Figura 9** – Ensaio Experimental. (a) Colocação; (b) Chegada; (c) Retirada.

# Metodologia

## Gráfico de Distância e Velocidade de PIG

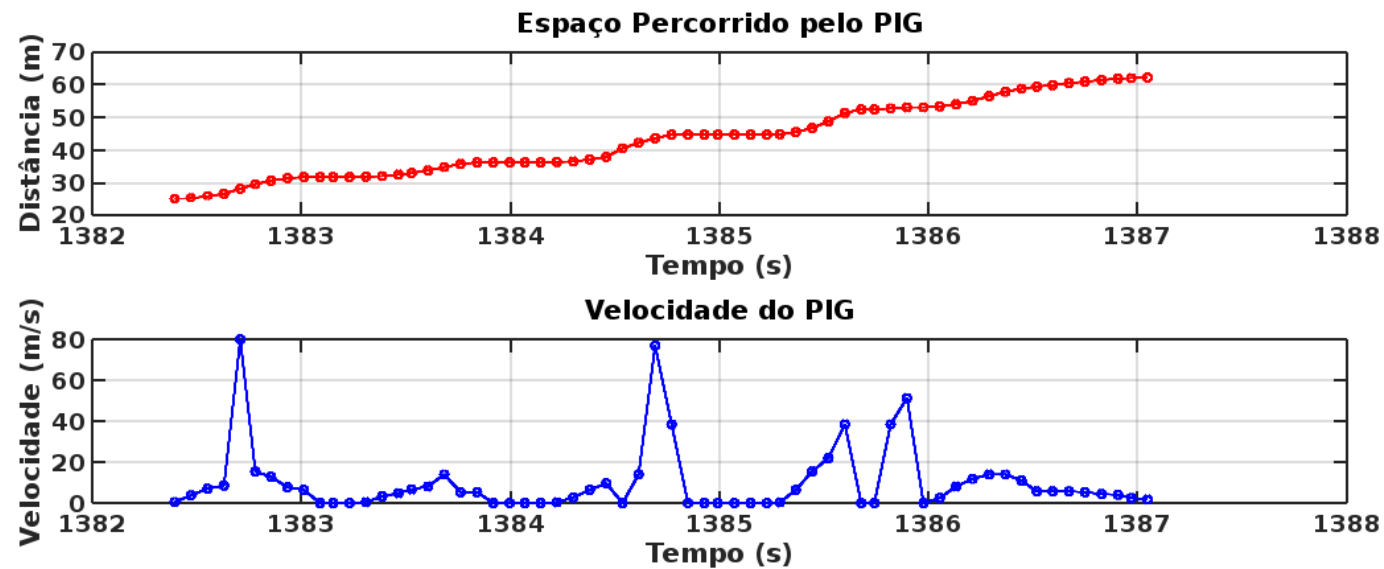


Figura 10 – Exemplo de gráfico com os dados de distância e velocidade do PIG Protótipo.

# Resultados e Discussões

```
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  
 31.0000   1384.692    76.970  
 32.0000   1384.771    38.480  
 43.0000   1385.600    38.480  
 46.0000   1385.819    38.480  
 47.0000   1385.897    51.310  
-----
```

```
>> |
```

Janela de Comandos Editor Documentação

**Figura 11** – Resultado do código implementado.

# Resultados e Discussões

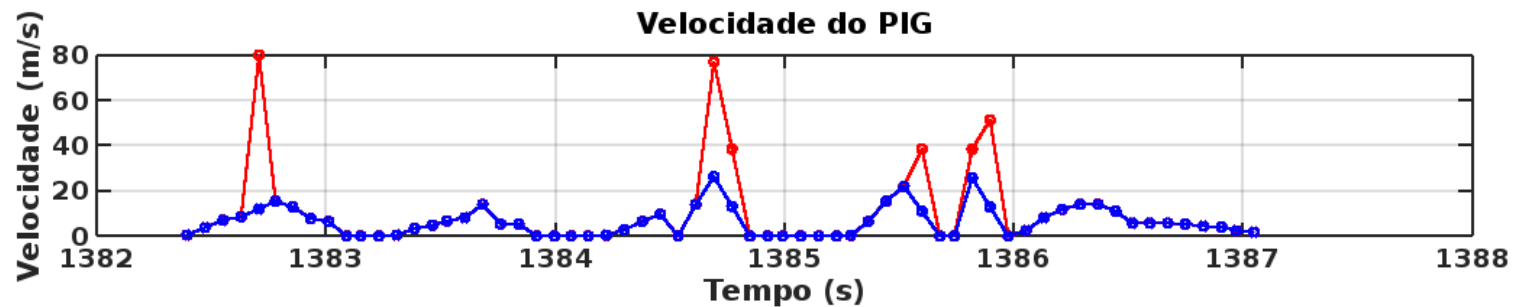


Figura 12 – Sobreposição dos dados depois da utilização do código implementado.

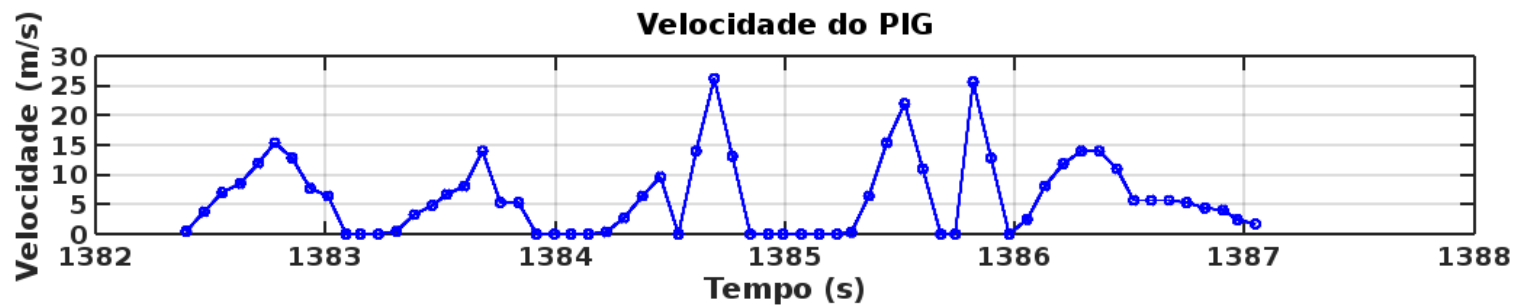


Figura 13 – “Novo” gráfico de velocidade sem os *outliers*.

# Resultados e Discussões

- ❑ Com a substituição dos *outliers* (cor vermelha) o “novo” gráfico é apresentado na Fig. 13. Nele, as maiores velocidades registradas são de 26,235 e 25,655 m/s, sendo consideradas aceitáveis.
- ❑ Outra vantagem, sem os *outliers*, os dados de velocidade ficam melhores de serem analisados, uma vez que a maioria dos pontos estão 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 mascarava os valores mais baixos.

# 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.
- ❑ Um PIG Protótipo e uma Tubulação de Testes são utilizados para obtenção de dados de velocidade. O código identifica 6 pontos fora da curva com sucesso e são substituídos por “novos” de menor valor. 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 é um sucesso. A principal contribuição deste trabalho é propor e apresentar um código capaz identificar e substituir *outliers* de dados utilizando ferramentas *open sources*.

# Bibliografia

- ❑ 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.
- ❑ 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.

Fim

O B R I G A D O

[gustavo.lima@ifrn.edu.br](mailto:gustavo.lima@ifrn.edu.br)

<http://docente.ifrn.edu.br/gustavolima>

Agradecimentos:

