Katiuski Pereira Nascimento

# Liquid level estimation and fluid identification using FBG temperature sensors via machine learning algorithms.

Vitória - ES

May - 2021

Katiuski Pereira Nascimento

# Liquid level estimation and fluid identification using FBG temperature sensors via machine learning algorithms.

Dissertation submitted to the Postgraduate Program in Electrical Engineering of the Technology Center of the Federal University of Espírito Santo as partial requirement for obtaining the Master's Degree in Electrical Engineering.

Federal University of Espirito Santo

Technology Center

Postgraduate Program in Electrical Engineering

Supervisor: Prof. Dr. Arnaldo Gomes Leal-Junior

Co-supervisor: Prof. Dr. Anselmo Frizera Neto

Vitória - ES

May - 2021

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

Nascimento, Katiuski Pereira, 1995-
N244l Liquid level estimation and fluid identification using FBG temperature sensors via machine learning algorithms. / Katiuski Pereira Nascimento. - 2021.
70 f. : il.

Orientador: Arnaldo Gomes Leal-Junior.
Coorientador: Anselmo Frizera-Neto.
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. FBG. 2. Sensores de temperatura. 3. Aprendizado de máquinas. I. Leal-Junior, Arnaldo Gomes. II. Frizera-Neto, Anselmo. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

---

Katiuski Pereira Nascimento

# Liquid level estimation and fluid identification using FBG temperature sensors via machine learning algorithms.

Dissertation submitted to the Postgraduate Program in Electrical Engineering of the Technology Center of the Federal University of Espírito Santo as partial requirement for obtaining the Master's Degree in Electrical Engineering.

Dissertation approved. Vitória - ES, May $5^{th}$, 2021:

---

**Prof. Dr. Arnaldo Gomes Leal-Junior**
Supervisor

---

**Prof. Dr. Anselmo Frizera Neto**
Co-supervisor

---

**Dra. Maria José Pontes**
(Federal University of Espírito Santo)

---

**Dr. Carlos Alberto Ferreira Marques**
(University of Aveiro, Portugal)

Vitória - ES
May - 2021

# Acknowledgements

FFirst, I thank my husband/best friend Thiago. He, who, with his understanding and unconditional love, was always beside me and never let me desist, even in the most challenging moments of our lives. Thiago, without your help I wouldn't get that far.

I thank my family: my brother Dyego, my mother Marli, my father Wilmes, and my in-laws Claudia, Luiz and Malu for all their care for me. I am so grateful for how much you are a part of my life.

I thank my friends and colleagues from the research group for the partnership, help, and for listening to all my doubts since I started working at LabTel. And for all the experience acquired in the lab.

I thank the Graduate Program in Electrical Engineering at the Federal University of Espirito Santo. I acknowledge funding received from Petróleo Brasileiro S. A. - PETROBRAS (2017/00702-6), the research project of Fundação de Amparo à Pesquisa e Inovação do Espirito Santo (FAPES) "PRONEM" (88336650).

Finally, I thank my supervisors and friends Arnaldo and Anselmo for all their patience and commitment to my learning. The best supervisors I could have.

"É chato chegar a um objetivo num instante.

Eu quero viver nessa metamorfose ambulante."

(Raul Seixas)

# Abstract

This dissertation proposes the use of the fiber Bragg grating (FBG) temperature sensors array to estimate the fluid level. A optical fiber sensor (OFS) level is ideal for evaluating oil tank level because it is a sensor that does not conduct electricity, is small size and resists corrosive areas. However, these sensors are complex to assemble, requiring several steps after fiber fabrication. Due to the temperature variation inside the tank, there is a need for a temperature sensor with no connection to these sensors, to measure the temperature. FBGs are intrinsically sensitive to temperature and strain. Therefore, level sensors also need a temperature sensor to reduce the temperature cross-sensitivity issues.

To demonstrate the possibility of using the FBG temperature sensor for liquid level estimation, the temperature distribution of an oil storage tank, 200 cm height and 40 cm in diameter, receiving solar radiation at the top, is simulated. Then, the presence of a 200 cm long and 125 $\mu m$ diameter fiber inside the tank with different amounts and distribution of FBGs along the fiber is simulated. In the simulation, due to the low variability of the classes, the Random Forest (RF) algorithm was chosen for classification. Starting with 200 FBG equidistant, decreasing to 6, with different distributions along the fiber. It was possible to classify the oil with an accuracy of 94.89% using 8 FBGs, using Tests for Two Proportions with a significance of 5%, the accuracy is equal to use 50 FBGs.

Using the results obtained in the simulation, we utlized a 22.5 cm beaker, with 3 FBGs inside. In the beaker, 3 different fluids are identified: water, mineral oil, and kyro oil. Afterwards their levels are estimated from the temperature distribution along the beaker (using the 3 FBGs). Furthermore, we keep the fluid inside the beaker heated by a peltier at the bottom of the beaker to 318.15 K during the entire experiment. We followed the same principle for the beaker experiment, using RF for both level identification, obtaining 100% accuracy in fluid identification, and fluid level measurement the mid RMSE was 0.2603.

After the simulation commented above, and the bench tests, using the beaker at constant temperature, we decided to expand the experiment. In this way we propose a full-scale experiment, using 9 FBGs distributed in this tank to estimate the liquid level. The tank is 100 cm in height and 30 cm in width, with 9 FBG sensors distributed along with the tank height. For the detection, we use the following Machine Learning (ML) algorithms: Logistic Regression (LogR), Decision Tree (DT) and Support Vector Machine (SVM). Initially the algorithm chosen was RF, but when using it we obtained RMSE of 16.32 cm. The algorithms chosen are Weighted Linear Regression (WLR), Support Vector Regression (SVR), SVR with kernel selection minimize cost (SVRmin). We propose the Mixed Model (MM), which selects the lowest Root Mean Square Error (RMSE) among the tested regression algorithms at each level, and associates it to it. The MM has RMSE of 3.56 cm,

which is approximately four times smaller than when using WLR. The SVM and SVMmin have RMSE of 6.28 cm and 6.14 cm, respectively.

**Keywords**: FBG, Temperature sensors, Machine Learning, Random Forest, Weighted linear regression, Support Vector Machine, Support Vector Regression, Decision Tree, Logistic Regression.

# Resumo

Esta dissertação propõe a utilização de grades de Bragg em fibra (FBG) para estimar o nível do fluido. O sensor de fibra óptica (OFS) de nível é ideal para avaliar o nível do tanque de óleo porque é um sensor que não conduz eletricidade, é de pequena dimensão e resiste a áreas corrosivas. No entanto, estes sensores são complexos de montar, exigindo várias etapas após a fabricação da fibra. Devido à variação de temperatura dentro do tanque, há a necessidade de um sensor de temperatura sem ligação a estes sensores, para medir a temperatura. As FBGs são intrinsecamente sensíveis à temperatura e a tensão. Portanto, os sensores de nível também precisam de um sensor de temperatura para reduzir os problemas de sensibilidade cruzada de temperatura. Assim, a estimativa do nível utilizando apenas a resposta de temperatura resulta em benefícios operacionais e econômicos, uma vez que há menos sensores e fácil montagem do conjunto de sensores. Este trabalho propõe a utilização de sensores de temperatura FBG para identificar os fluidos. A estimativa do nível utilizando apenas a temperatura resulta em benefícios operacionais e econômicos.

Para demonstrar a possibilidade de utilizar o sensor de temperatura FBG para a estimativa do nível de líquidos, é simulada a distribuição da temperatura de um tanque de armazenamento de petróleo, 200 cm de altura e 40 cm de diâmetro, recebendo a radiação solar no topo. Depois, é simulada a presença de uma fibra de 200 cm de comprimento e 125 $\mu m$ de diâmetro dentro do tanque com diferentes quantidades e distribuição de FBG ao longo da fibra. Na simulação, devido à baixa variabilidade das classes, foi escolhido o algoritmo Random Forest (RF) para a classificação. Começando com 200 FBG equidistantes, diminuindo para 6, com diferentes distribuições ao longo da fibra. Foi possível classificar o óleo com uma precisão de 94,89% usando 8 FBGs, usando Testes para Duas Proporções com um significado de 5%, a precisão é igual a usar 50 FBGs.

Utilizando os resultados obtidos na simulação, foi utilizado um béquer de 22,5 cm, com 3 FBGs no interior. No béquer, são inseridos 3 fluidos diferentes: água, óleo mineral, e óleo kyro. Posteriormente, os níveis são estimados a partir da distribuição da temperatura ao longo do béquer (utilizando os 3 FBGs). Além disso, mantemos o fluido no interior do copo aquecido por um peltier no fundo do copo a 318,15 K durante toda a experiência. Seguimos o mesmo princípio para a experiência do béquer, utilizando RF para a identificação do nível, obtendo uma precisão de 100% na identificação do fluido, e a medição do nível de fluido a RMSE média foi de 0,2603.

Após a simulação comentada acima, e os testes de bancada, utilizando o béquer a temperatura constante, decidimos expandir a experiência. Desta forma, propomos uma experiência à escala real, utilizando 9 FBGs distribuídos num tanque para estimar o nível de líquido. O tanque possui 100 cm de altura e 30 cm de largura, com 9 sensores FBG distribuídos juntamente com a altura do tanque.

Para a detecção, utilizamos os seguintes algoritmos de Aprendizado de máquina (ML): Regressão logística (LogR), Árvore de decisão (DT) e Máquina de vetor suporte (SVM). Escolhemos os algoritmos com base na sua usabilidade na literatura e consolidação teórica. O algoritmo com os melhores resultados entre os testados é o DT, resultando numa precisão média de 89,54%.

Inicialmente o algoritmo escolhido foi o RF, mas ao utilizá-lo obtivemos RMSE de 16,32 cm. Os algoritmos escolhidos são: Regressão linear ponderada (WLR), Máquinas de vetor suporte para regressão (SVR) e SVR com seleção de kernel que minimize o custo (SVRmin). Propomos o Modelo Misto (MM), que seleciona a menor Raiz do Erro Quadrático Médio (RMSE) entre os algoritmos de regressão testados em cada nível. O MM tem um RMSE de 3,56 cm, que é aproximadamente quatro vezes menor do que quando se utiliza o WLR. O SVM e o SVMmin têm RMSE de 6,28 cm e 6,14 cm, respectivamente.

**Palavras-chave**: FBG, Sensores de temperatura, Aprendizado de máquinas, Florestas aleatórias, Regressão linear ponderada, Máquinas de vetor suporte, Máquinas de vetor suporte para regressão, Árvore de decisão, Regressão logística.

# List of Figures

# List of Tables

# Contents

# 1 Introduction

## 1.1 Motivation and Justification

Liquid level measurement aids on cost reduction, increases productivity and contributes to quality and safety in many industrial processes (JIA et al., 2015). The use of sensors has become useful for monitoring these processes as they provide real-time monitoring of the production and feedback for control and management systems. There are several sensors for level measurement based on different technologies, for example: multi-electrode capacitance (SHI et al., 1991), ultrasound (BUKHARI; YANG, 2006), radiation (HJERTAKER; JOHANSEN; JACKSON, 2001), vibration (MERIBOUT et al., 2011), fluctuation (CHEN et al., 2018), radio (MAGNO et al., 2014) and light reflection (GU et al., 2020). However, due to their electromagnetic fields sensitivity, use of electrical signals and the risk corrosion, the use of such sensor are not suitable for classified areas (WRIGHT et al., 2019).

The FBG sensors are versatile, detect sensing physical quantities such as deformation, temperature, pressure, ultrasound, magnetic field and force (LV et al., 2019). FBGs consist of periodic modulation of the refractive index inscribed in the fiber optic core. Due to their multiplexing capabilities, these sensors have been increasingly used (WRIGHT et al., 2019). Multiplexing of FBG sensors is an effective technology to provide multi-point measurement over a single fiber cable (CHEN et al., 2018). FBG temperature sensors are present in healthcare (CHEN et al., 2018), construction and general industry (RAO et al., 1997; LEAL-JUNIOR et al., 2018a).

Hydrostatic pressure sensors measure fluid levels based on their densities (ZHANG; KAHRIZI, 2007; AMEEN et al., 2016). However, these sensors need incorporation in different structures or suffer fiber modifications for the pressure sensing (ZHANG; KAHRIZI, 2007). Thus, FBG level sensors need additional steps after the inscription of the FBG, different than an FBG temperature sensor. Ameen et al. (2016) uses diaphragms, fixing the FBG grating region in the center of the diaphragm and inserts a second diaphragm covering the grating. The sensors based on the Archimedes' law of buoyancy consist of suspending a mass at one end of an FBG liquid level sensor while holding the other fixed on a rigid support (CONSALES et al., 2018). Similarly, there is also the integration of an FBG in a cantilever rod connected to a float (SOHN; SHIM, 2009). In literature, several other level sensors in FBG are proposed (BUKHARI; YANG, 2006; LEAL-JUNIOR et al., 2018c; DIAZ et al., 2019a).

The use of only temperature response results in operational and economic benefits,

since there are fewer sensors and easy assembly of the sensor array due to FBG inherent sensitivity to temperature variations. Thus the use of an FBG temperature sensor that measures level and temperature simultaneously reduces the number of sensors used in an oil tank or oil storage facility. In addition, because it is a quasi-distributed sensor, depending on the number of FBGs inscribed in the optical fiber, there is an increase in accuracy.

The application of FBG temperature sensors can be found in the area of health (MAJEED; MURDAS, 2019), construction (LV et al., 2019), and industry in general (RAO et al., 1997). More specifically, this sensor was previously used in the oil industry, due to the temperature variations observed on crude oil tanks (LEAL-JUNIOR et al., 2018c). Refractive index (ZHANG; KAHRIZI, 2007) and hydrostatic pressure sensors in which the fluid density (LEAL-JUNIOR et al., 2018c) can also be used for the fluid classification. However, these sensors need incorporation in different structures (LEAL-JUNIOR et al., 2018c) or modifications in the fiber for the refractive index sensing (ZHANG; KAHRIZI, 2007). FBGs are naturally sensitive to temperature and strain (LEAL-JUNIOR et al., 2018a). Thus, oil classification using only temperature response results in operational and economic benefits, since there are fewer sensors and easy assembly of the sensor array due to FBG inherent sensitivity to temperature variations (SILVA et al., 2007).

The proposed sensor for level estimation and fluid identification consists of using a single mode fiber (SMF) with Bragg gratings, in short an FBG temperature sensor. Thus for its creation, the addition of any other materials and components is not necessary. Thus, the challenge lies in understanding how the level affects the temperature gradient along the fiber is related to the thermal to the thermal dynamics of the system. Due to non-linear or second order effects, the need arises for a survey of techniques for the estimation and classification of the fluid level, seeking a solution to this problem we use ML. The possibility of level estimation using FBG temperature sensors is shown in the throughout this dissertation. This dissertation numerically and experimentally demonstrate the proposed approach. Thus, numerically it presents a water-oil interface classification for level detection. In the experiment, we used level estimation based on the predicted location of the FBGs. And then, detection is performed for different fluids: water, mineral oil and Kryo 51 oil (silicone oil cover a wide temperature range), where subsequently occurs the estimation of the level for each fluid.

The level effect on fiber temperature variation is related to the thermal dynamics of the system. It may have nonlinear or second-order effects (KHOSRAVI; AZIZIAN; BOUKHERROUB, 2019). Therefore, to estimate the level from the temperature it may be necessary to use different data processing techniques. As one of these processing techniques, ML algorithms are divided into three learning groups: supervised, semi-supervised, and unsupervised (SOLOMON et al., 2020). Supervised ML algorithms are those that the

possible responses of the problem are known, for example: linear regression (LR) (AL-YASEEN; OTHMAN; NAZRI, 2017), support vector machines (SVM) (ZHU et al., 2019) and random forests (RF) (SU; CHEN; ZHANG, 2019). In contrast, unsupervised ML algorithms, the response of the problem is unknown, the so-called clustering techniques, for example: hierarchical clustering (MANIE et al., 2020), principal components analysis (PCA) (DJURHUUS et al., 2019). Finally, semi-supervised ML algorithms are those where learning is done with supervised and unsupervised inputs, seeking to find grouping and response criteria, for example: extreme learning machine (ELM) (MANIE et al., 2018), deep Boltzmann machine (DBM) (YAO et al., 2020).

## 1.2 Objectives

This dissertation proposes the estimation of the level through FBG temperature sensors. In order to detect/estimate the level via FBG temperature sensors, we selected the supervised ML, as the data has known response obtained from the tests conditions. The proposed estimation consists of two steps. This work is divided into two steps: a simulation study and the application to the problem. The application, is done first using a beaker and the second using a storage tank, 100cm.

The specific objectives of this work are listed as follow:

1. To demonstrate the possibility of level estimation using FBG temperature sensors via simulation;
2. To study Machine Learning techniques and measuring their capability for level estimation and fluid detection.
3. To demonstrate the possibility of identifying different fluids using FBG temperature sensors.
4. To demonstrate the possibility of level estimation using temperature FBG in benchtop and storage tanks prototypes;

## 1.3 Dissertation Structure

This dissertation is organized as follows: the theoretical background of the history of optical fiber and FBG sensors is presented. In addition we list the ML methods used in the course of the dissertation. Then the problem is presented in its simulated process as well as in the experimental setup. After presenting the simulation and the two studied setups, we present the results for each of them, including the simulation. At the end, we present the conclusions gained from the work, as well as future work for improvements and continuation of the study.

# 2 Theoretical background

## 2.1 Optical fiber sensors

At the end of the 18th century, John Tyndall sought the transmission of light through curvilinear paths (LUO et al., 1999). Only in the following century, in the year 1950, the term "optical fiber" (GRATTAN; SUN, 2000) appeared. However, for Kapany, Eyer e Keim (1957) the optical fiber was possible to be used only in medicine. In the following decade, the use of optical fibers in communication was already foreseen due to the reduction of the optical attenuation in the waveguides (GRATTAN; SUN, 2000). It was in the 1970's, with the improvement of the emitters, optical fiber transmission systems (GRATTAN; SUN, 2000) appeared. In the 1980s, optical fibers revolutionized communication systems worldwide, providing large amounts of information transmission with high speeds between thousands of kilometers (GRATTAN; SUN, 2000). According to Diaz et al. (2019a), this progress was made possible by technological advances in the production of optical fibers together with technological developments to manufacture compact optical sources, new optical amplification schemes and new modulation techniques.

Besides the use of optical fiber in communications, it has also been widely used in the area of sensing (DIAZ et al., 2019a). Due to its immunity to electromagnetic interference, electrical insulation, small size, large bandwidth, light detection heads, high precision and resolution, easy multiplexing and remote monitoring capability the optical sensors have been widely used in the field of sensing (XU et al., 1993). In addition, these advantages make optical fibers suitable for long-term and reliable level measurements under special conditions (GRATTAN; SUN, 2000). Optical fiber sensing has stimulated research in different areas such as industrial, medical and civil (SILVA et al., 2019). In these areas there is a need for monitoring several parameters, such as liquid level, temperature, tension, pressure, deformation, acceleration, refractive index.

Optical Fiber Sensors (OFS) are immune to electromagnetic interference, are corrosion resistant and no electrical power is needed at the measuring point (WRIGHT et al., 2019). The OFS can be classified as based on intensity, phase modulated and wavelength modulated. The first class refers to sensors based on signal intensity variation as a function of a measurand, such as the microbend sensor (LUO et al., 1999). Then, the type phase-modulated OFS, that use of light phase change, the Michelson sensor is an example (SWART, 2004). The last class is the wavelength-modulated OFS. Sensors in this class use changes in the wavelength of light for detection. A widely used sensor in this class is the Fiber Bragg Grating (FBG) sensor (HILL; MELTZ, 1997). FBG sensors are a potential alternative to conventional electrical sensors for level utilization (VORATHIN et

al., 2019). Its main advantage is the multiplexing capabilities, allowing numerous sensors in a singular optical fiber cable (VORATHIN et al., 2019; DIAZ et al., 2019b). The FBGs consist of periodic or nonperiodic refractive index modulation inscribed in the optical fiber's core using continuous wave or pulsed lasers (ZHANG; KAHRIZI, 2007).

### 2.1.1 Fiber Bragg grating

In 1978, Kawasaki et al. (1978) reported the observation of photosensitivity in the central region of Ge-doped silica optical fibers, and this effect was used to produce narrow-band-waveguide reflection filters for the construction of a distributed-feedback gas laser. Thus, with the Bragg reflectors, the filters are permanently induced in the fibers, without any decay or degradation of the optical fibers (KAWASAKI et al., 1978).

FBGs were discovered in 1978, by Hill et al. (1978), and have become a good option among optical sensors (VORATHIN et al., 2019; LEAL-JUNIOR et al., 2018b). These gratings are formed as periodic modulation in the refractive index along the longitudinal direction. The existence of this modulation causes a selective reflection in part of the spectrum of a broadband optical signal transmitted through that fiber. The light is guided along the core of the optical fiber being reflected from each plane of the network, and centered at a specific wavelength (HILL; MELTZ, 1997). FBG sensor technology acts as a wavelength selective mirror, such as a filter tuned to a narrow band (HILL et al., 1978). In this way, a broadband light source when injected into the optical fiber where the FBG is located, only light that is in that narrow spectral region centered on the Bragg wavelength reflected (HILL et al., 1978). And so the remaining light continue its path through the fiber with no representative losses.

FBGs sensors have significant advantages such as small size, low cost, long term stability and robustness (LEAL-JUNIOR et al., 2018a). They have been widely applied to perform for temperature and strain detection in classified areas, composite defect inspection, and structural health monitoring (OU et al., 2017). Used also, places of dense sensing point demand, for example, structural health monitoring, such as bridges and buildings. According to Wang et al. (2011) a large number of sensing FBGs are distributed to the sensing region, and the FBGs are multiplexed with various methods to reduce the unit cost of the sensor.

FBGs are naturally sensitive to temperature and strain. Equation 2.1, shows the wavelength variation ($\Delta\lambda_B$) caused by temperature and strain effects on the FBG.

$$\frac{\Delta\lambda_B}{\lambda_B} = (1 - P_e)\epsilon(p) + (\alpha + \zeta)\Delta T \tag{2.1}$$

where $P_e$ is the effective photoelastic constant, $\epsilon(p)$ is the strain on the fiber, $\alpha$ is the thermal expansion coefficient, $\zeta$ is the fiber thermo-optic coefficient, and $\Delta T$ the temperature variation. According to Diaz et al. (2019a), FBGs are widely used due to their intrinsic

detection response to stress and temperature, inherent capacity for self-referencing, making them independent of fluctuating levels of light and other light noises. The information collected can be encoded in wavelength. These sensors measure the liquid level based on the length of the immersed sensor, which leads to a specific and effective refractive index according to (MARTINS et al., 2019).

## 2.2 Machine Learning Process

In the literature, there are several ML supervised algorithms, ranging from the most traditional such as linear regression (HSIEH; BLOCH; LARSEN, 1998) to the most modern such as convolutional neural networks (YAO et al., 2020). In classification, Ye (2019) and Gola et al. (2019) suggested the use of SVM due to high precision and low false-positive rate. Ing et al. (2019) recommended the use of LogR, a statistical technique already consolidated. According to Sarkar et al. (2019), DT are that it can build smaller trees, better memory capability. In regression, a statistical technique as popular as Logistic Regression (LogR) is LR. The Weighted Linear Regression (WLR) has lower assumptions than the LR and presents similar characteristics (JOHANSEN, 1980). When used for regression problems, the SVM is named SVR (SMOLA; SCHOLKOPF, 2004). To understand the differences between regression and classification, basically in classification we have a qualitative analysis of the data. In regression, the analysis is quantitative.

### 2.2.1 Classification Process

The level detection or the fluid classification is done to decrease the errors related to the liquid level estimation. Johansen (1980) recommended use Logistic Regression (LogR), as it is appropriate for a binary or categorical response. Moreover, it allows estimating the probability of occurrence of the event of interest. While the SVM due to its convex solution is not local minimum affected (ZHU, 2014; PŁAWIAK; ABDAR; ACHARYA, 2019; JIA et al., 2018). Rodriguez-Galiano et al. (2012) recommended Decision Tree (DT) due to its easy implementation and outliers resistance.

LogR is a statistical technique useful in binary data. It is a resource that allows us to estimate the probability associated with the occurrence or not of an event based on the variables studied (SOCIETY, 2017; ING et al., 2019), which enable calculate the probabilities of an event belonging, and not belonging, to the class of interest.

SVM is used to solve both classification and regression problems (JIA et al., 2019). In this work, we use the response of each FBG temperature sensor. This algorithm consists of dividing the input data into two categories and finding a hyperplane that best separates the data (NIU; HA; CHI, 2018).

DT is an ML algorithm discovered by Quinlan (SARKAR et al., 2019). DT consists of mapping the possible results of a series of related choices. The idea of the algorithm is to divide the dataset into "pure" subsets as to class. The algorithm establishes the criteria that best separate the classes. The tree starts with a single node with branches. Each of these results leads to additional nodes, which branch out into other possibilities. This branching is done until you find groups that contain only one class or a minimum number of class mixes.

RF instead of using a single DT, makes use of several trees, which explains its association with forests. RF is an ensemble-learning algorithm (ALI et al., 2012). Algorithms with this learning methodology are more robust and precise than those in which unique learning is considered (LEE; ULLAH; WANG, 2020). This method is similar to decision trees. Its hierarchical form allows considering nonlinear relations in the data, to generate classification regions (PRONIEWSKA; PREGOWSKA; MALINOWSKI, 2020).

#### 2.2.1.1 Logistic Regression

In LR, the probability of occurrence of an event can be estimated directly. In the case of the dependent variable $Y$ is a binary variable, in this case the problem can be summarized to: is the fluid above or below the observed FBG. For an LR model there is a need for the data to be mutually exclusive, so we assume that Y has a Bernoulli distribution, with probability of success $P(Yi = 1) = \pi$ and probability of failure $P(Yi = 0) = 1 - \pi$, thus:

$$Y_i \sim B(1, \pi_i). \tag{2.2}$$

However, when $n$ independent samples are observed and let $p \geq 1$ independent variables $X_1$ , $X_2$ , ..., $X_p$, which would be the $\Delta\lambda$, we have that Y has binomial distribution of parameters $p$ and $\pi$:

$$Y_i \sim Bin(p, \pi_i). \tag{2.3}$$

This form, the logistic regression model can be expressing the probability of $Y$.

$$\begin{aligned} \pi_i &= \frac{exp\{-g(X_i)\}}{1 + exp\{-g(X_i)\}} \\ &= \frac{1}{1 + exp\{-g(X_i)\}}, \end{aligned} \tag{2.4}$$

where $g(X_i) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$. Thus, we can rewrite Equation 2.4 as follows:

$$\begin{aligned} \ln\left(\frac{\pi_i}{1 + \pi_i}\right) &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \\ &= \beta_0 + \sum_{i=1}^{p} \beta_i X_i. \end{aligned} \tag{2.5}$$

According to Jr, Lemeshow e Sturdivant (2013), the coefficients $\beta_i$, for $i = 1, 2, \ldots, p$ are estimated from the data set using the maximum likelihood method (JR; LEMESHOW; STURDIVANT, 2013), then a combination of coefficients is made that maximizes the probability that the sample was observed. Based on the observed probability ($\pi_i$), using cross-validation to associated $Y_i$ in 1 or 0.

#### 2.2.1.2 Support Vector Machine

SVM consists of finding the minimum structural risk (JIA et al., 2019). Thus, SVM is an algorithm whose goal is to classify a given set of data points that are mapped to a multidimensional feature space using a kernel function, an approach used to classify problems (CORTES; VAPNIK, 1995). In it, the decision boundary in the input space is represented by a hyperplane in higher dimension in space.

A classification problem of order $k$, for SVM can be viewed as an optimization problem as shown in the Equation 2.6.

$$\Phi(w, \xi) = \frac{1}{2} \sum_{m=1}^{k} (w_m \times w_m). \tag{2.6}$$

Using to theory of Lagrange multiplications (CORTES; VAPNIK, 1995) the optimization problem represent in Equation 2.7.

$$J(w, b, \alpha) = \frac{1}{2}(w_m \times w_m) - \sum_{i=1}^{N} \alpha_i \left[ y_i((w_m \times x_i) + b) - 1 \right], \tag{2.7}$$

subject to:

$$(w_{y_i} \times w_i) + b_{y_i} \geq (w_m \times x_i) + b_m + 2 - \xi_i^m,$$

$\xi \geq 0$, $i = 1, 2, \ldots, l$, and, $m$. Getting the decision function:

$$g(x) = argmax_n[(w_m \times x) + b_n], \qquad n = 1, 2, \ldots, k. \tag{2.8}$$

In resume, the SVM creates a new n-dimensional space, to find the hyperplane that best differentiates the classes, which in our problem would be the dichotomous variable: fluid level is above or below the observed FBG (CORTES; VAPNIK, 1995).

#### 2.2.1.3 Decision tree

DTs are maps of possible outcomes of a series of choices concerning observed data. A decision tree usually begins with a single node, which is divided into possible outcomes. Each of these outcomes leads to additional nodes, which branch into other possibilities(SARKAR et al., 2019).

The procedure of a DT is described as follows: based on information from the data set, the process is started through the initial node of the tree (SARKAR et al., 2019).

Depending on the result of the logic test used by the node, the tree branches to one of the child nodes and this procedure is repeated until an end node is reached. The repetition of this procedure characterizes the recursion of the decision tree. Thus, DT is a recursive algorithm that is easy to implement (ALI et al., 2012).

#### 2.2.1.4 Random Forest

In RF, different bootstrap samples are considered for each classification tree randomly selected (PRONIEWSKA; PREGOWSKA; MALINOWSKI, 2020). All bootstrap samples are identically distributed, and the average of all samples is equal to the average of each sample (PRONIEWSKA; PREGOWSKA; MALINOWSKI, 2020). The model bias of the aggregation trees is equivalent to that observed for each tree (RODRIGUEZ-GALIANO et al., 2012). The RF algorithm tries to keep the low bias of each classify, since it reduces the total variance. Thus, the RF construction for each tree is given by:

$$h(x, \theta_k), \tag{2.9}$$

where $x$ is the input vector and $\theta_k$ are the independent and identically distributed random vectors (RODRIGUEZ-GALIANO et al., 2012). When RF increases the nodes of the classification tree, a better division of a random subset of input observations or predictive variables are chosen for the divisions of its nodes (PRONIEWSKA; PREGOWSKA; MALINOWSKI, 2020). It should be noted that, as the tree growth is done without pruning, computationally an RF classify requires lower processing costs compared to other methods (REIS; BARON; SHAHAF, 2018).

The RF classify uses the Gini index as the attribute selection measure. This index measures the degree of randomness of an attribute for within a partition (RODRIGUEZ-GALIANO et al., 2012). For an arbitrary set of T formation, randomly selected and belonging to class $C_i$, the Gini index can be written as:

$$\sum\sum_{j \neq i} = \left(\frac{f(C_i, T)}{|T|}\right)\left(\frac{f(C_j, T)}{|T|}\right), \tag{2.10}$$

where $f(Ci, T)/|T|$ is the probability that the selected case belongs to class $C_i$. To classify a new data set, each new observation passes through each of the previously created $N$ trees. The forest chooses a class with the highest number of votes and classifies it according to the vote.

### 2.2.2 Regression Process

Regression consists of modeling to obtain real values of the level. In this technique, we use the results obtained in the classification and . The algorithms selected were Weighted Linear Regression (WLR), Support Vector Regression (SVR) and SVR select the that

results in minimum cost (SVRmin) and RF for Regression (RFR). LR is a traditional statistical model for the use of regression. There is an adaptation of the LR for problems that do not meet the assumption of normality, the WLR (JOHANSEN, 1980). Due to its consolidated theory and the non-requirement of normality, this algorithm was chosen for this paper. Another algorithm used in this paper due to its statistical theory is the SVM (SMOLA; SCHOLKOPF, 2004). It is called SVR for regression problems. The SVRmin is an adaptation of the SVR.

LR studies the relationship between a variable called response (level) and others, called independent variables (from the FBGs studied and the level detection). It consists of a mathematical model that associates the response variable with the independent variables (JOHANSEN, 1980). However, LR has assumptions that must be fulfilled to reduce errors, among them, the normality in the data, linearity, among others [25]. When this assumption is not satisfied, the WLR is used (JOHANSEN, 1980). Similar to LR, WLR establishes a linear relationship between the data but using parameter functions obtained in LR to weight its parameters. This weighting results in a decrease of the estimation errors.

For regression problems, there is an adaptation, of the SVM method, the SVR. In the SVR, it does not matter from which side of the surface the points are located. The interest of the method is that the distance to the surface is the minimum possible. The errors are allowed as long as they do not exceed the specified threshold (SMOLA; SCHOLKOPF, 2004). The SVR depends on the gamma ($\gamma$) and cost ($C$) parameters. Forming an ordered pair space $(\gamma_i, C_i)$, using exhaustive search for the pair that results in higher accuracy. The SVRmin, like the previous algorithm, works with the SVR, but instead of exhaustively searching for the pair, we select the that results in minimum C.

#### 2.2.2.1 Weighted Linear Regression

LR is a mathematical equation that provides a linear relationship, between $U$, referring to the estimated fluid level, with the other variables $X_i$ are $\Delta\lambda$ of the observed FBGs. This linear function is expressed as follows:

$$Y_i = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \epsilon_i, \tag{2.11}$$

where $\epsilon_i$ is error. Similar to 2.2.1.1, $\beta_i$ are the coefficients of the model, however in LR, they are estimated by the least squares method. For the application of an LR algorithm, some assumptions are required:

1. Error has zero expected value and unknown variance.

$$E(\epsilon) = 0 \qquad \text{and} \qquad Var(\epsilon) = \sigma^2$$

2. Furthermore, the errors are not correlated with with one another.

$$\rho(\epsilon_i, \epsilon_j) = 0, \qquad i \neq j$$

3. The error must have a normal distribution.

$$\epsilon \sim N(0, \sigma^2)$$

4. The variable $X_i$, for $i = 1, \ldots, p$, assumes fixed value.

A test frequently used to identify heteroscedasticity is the Breusch-Pagan (BP) test (HALUNGA; ORME; YAMAGATA, 2017). Testing the following hypotheses:

- $H_0$: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_n^2$.
- $H_1$: At least one $\sigma_i^2$ is different, for $i = 1, 2, 3, \ldots, n$

Each standardized error can be defined as:

$$u_i = \frac{e_i^2}{(\sum_{i=1}^{n} e_i^2)/\, n} \tag{2.12}$$

where $e$ is the difference between the predicted and expected value. Using $u$, and the predicted values in the linear regression $\hat{y}$, to create a new model, where $u$ is the response variable and $\hat{y}$ is the explanatory variable. By calculating the Regression Sum of Squares (RSS), we obtain the statistic $\chi^2_{BP}$ (see Equation 2.13) with chi-square distribution and $n - 1$ degrees of freedom.

$$\chi^2_{BP(studentized)} = \frac{\chi^2_{BP}}{\lambda} \tag{2.13}$$

where $\lambda = Var(e)^2/(Var(e))^2$.

The $\chi^2_{BP(studentized)}$ has a chi-square distribution of 1 degree of freedom under the null hypothesis. Thus, if there is no heteroscedasticity, it is expected that the squared residuals not change with increasing predicted value, $\hat{y}$, obtaining test statistic with insignificant value. Thus, a significant p-value is obtained. However, when the p-value is not significant, we can guarantee the non-normality of the data, and thus the non-possibility of using LR.

To solve this problem we propose the WLR. The WLR loosens the assumptions, and to reduce the estimated error generated by this loosening, it weights the estimated $\beta_i$ for each $X_i$. In this way we find the heteroscedasticity of the data (JOHANSEN, 1980).

Thus, the variance of the least squares estimate depends on the variance of the experimental errors. The variance of the experimental error is not constant, but can be

represented as $Var(\epsilon) = \sigma^2 V$, where $V$ is a known diagonal matrix. Let $V$ be a non-singular, positive definite matrix. Then, there exists a matrix $K$ that:

$$K^T K = V. \tag{2.14}$$

So, we can say $Z = K^{-1}Y$, $\phi = K^{-1}\beta$ and $\xi = K^{-1}\epsilon$. In this way the WLR can be expressed as:

$$Z_i = \phi_0 + \sum_{i=1}^{p} \phi_i X_i + \xi_i. \tag{2.15}$$

#### 2.2.2.2 Support Vector Regression and Support Vector Regression minimum

The formulation of the SVR uses principles similar to the classification algorithm (SMOLA; SCHOLKOPF, 2004). However, unlike the SVM, the SVR proposes to determine an optimal hyperplane in which the training samples are as close as possible. In this case, the method is called support vector regression. First, needing to find a predictor that has the lower error among a set of predictors. To do this we need to define what the error of a predictor is, and a measure to evaluate its performance. A natural way to measure the error is with a function $C$, which measures the error of the predictor on the sample set.

$$Y = \omega X + \tau, \tag{2.16}$$

where $\omega \in X$ is normal vector of a hyperplane $\tau/||\omega||$ through in origin relation, where $\tau \in \mathbb{R}$. Based in 2.16, SVR minimize $\omega$, thus:

$$\min_{\omega,\tau} \frac{1}{2}||\omega||^2, \tag{2.17}$$

subject to:

$$\begin{cases} y_i - \omega x_i - \tau & \leq \quad \eta_i \\ \omega x_i + \tau - y_i & \leq \quad \eta_i. \end{cases} \tag{2.18}$$

As this is a regression model, it is assumed that for the estimation, there are outliers and even noise in the data, creating a range to ensure the acceptance of the predicted value ($\hat{y}$) with the observed (y), thus:

$$|\hat{y}_i - y_i| < \zeta. \tag{2.19}$$

Therefore, we have:

$$\min_{\omega,\tau,\zeta} \frac{1}{2}||\omega||^2 + C\left(\sum_{i=1}^{n} \zeta_i\right). \tag{2.20}$$

The parameter $C \leq 0$ determines the penalty for cases in which the regression function accepts objects with deviations larger than the value $\zeta$. Thus, the robustness of the regression model depends on the choice of $C$ value.

According to SMOLA e SCHOLKOPF (2004), using the Lagrange multiplier ($\alpha$ and $\alpha^*$) method a dual problem can be obtained. Obtaining the objective function as follows:

$$\hat{y} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x_j) + \tau. \tag{2.21}$$

$K$ is the function, a function that takes two points ($x_i$ and $x_j$) in the original input space and computes the scalar product of these objects in the high dimensional feature space. There are innumerable $K$ functions (SMOLA; SCHOLKOPF, 2004), however due to its popularity, easy fit to diverse data and having only one parameter to be estimated (RAMEDANI et al., 2014), the Radial Basis Function (RBF) was chosen:

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2), \tag{2.22}$$

where $\gamma$ controls the flexibility of the function $K$. The choice of the ordered pair $(\gamma, C)$ is done exhaustively via cross validation. We propose via evaluation of the quadratic function $|x_i - x_j||^2$, to find a $\gamma$ that generates smaller $C$, called SVRmin. This way there is lower computational effort if we compare SVRmin to SVR.

#### 2.2.2.3 Random Forest for Regression

For regression, RF has the same algorithm as in 2.2.1.4. However, when it comes to classification, the final prediction is decided selecting the class that most of the trees identified as correct. In regression, the decision is based on the arithmetic mean of the forest tree results (LIAW; WIENER et al., 2002).

# 3 Material and methods

## 3.1 Simulation process

A simulation study is carried out to demonstrate the possibility of using an FBG temperature sensor for liquid level identifier. A tank with dimensions of 200 cm height and 40 cm diameter is simulated, assuming the presence of solar radiation. Inside the tank, the oil, emulsion and water layers were simulated. The thermal conductivity for water is $6.13\times10^{-3}\ W/mK$ and for oil $1.2\times10^{-3}\ W/mK$. For emulsion, we assumed similar quantities of water and oil, thus choose the arithmetic mean of the thermal conductivities of water and oil under the same temperature, $3.665\times10^{-3}\ W/mK$. The solar angulation adopted is 0, to represent midday, period usually with higher radiation, known as sun at zenith (INCROPERA et al., 2007). The boundary conditions of the storage tank are shown in Figure 1. The center of the tank and the bottom are adiabatic, the top receives constant solar radiation, and the tank's external walls suffer convection with the external environment. The fluid inside the tank is considered to be in a stationary state. Axisymmetric distribution is adopted for the temperature profile.

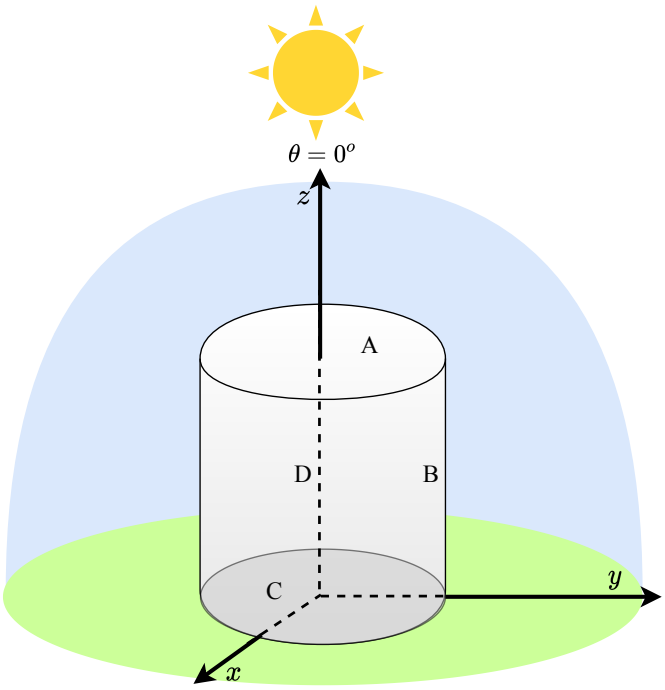



Figure 1 – Oil storage tank layout. A = Tank top (constant heat flow), B = Outside wall (convection with the external environment), C = Bottom (adiabatic wall), D = Axisymmetric (adiabatic).

Outside the tank, we assumed white paint on metallic substrate, which the emissivity of the tank surface ($\varepsilon$) shown in Table 1. Following$K$irkchhoff's law of thermal radiation, the atmosphere absorptivity ($\alpha_{atm}$) is also defined as $\varepsilon$ (INCROPERA et al., 2007). The net heat flow of the radiation ($q''_{rad}$) in the tank is calculated as follows:

$$q''_{rad} = \alpha_s G_s + \alpha_{atm} G_{atm} - E \tag{3.1}$$

where, $\alpha_s$ and $\alpha_{atm}$ are presented in Table 1, $G_s$ is solar irradiation, $G_s$ is defined in (3.2), $f$ is a correction factor to take into account the eccentricity of Earth's orbit around the Sun ($0.97 \leq f \leq 1.03$) (INCROPERA et al., 2007), assuming $f = 1$ without loss of generality. $G_{atm}$ is the atmospheric irradiation, defined in (3.3), and $E$ in (3.4), is the emissive power of the tank surface (INCROPERA et al., 2007).

$$G_s = S_c f \cos(\theta) \tag{3.2}$$

$$G_{atm} = \sigma_{SB} T_{atm}^4 \tag{3.3}$$

$$E = \varepsilon \sigma_{SB} T_{Sf}^4 \tag{3.4}$$

Table 1 – Parameters employed to simulate the temperature distribution of the oil storage tank.

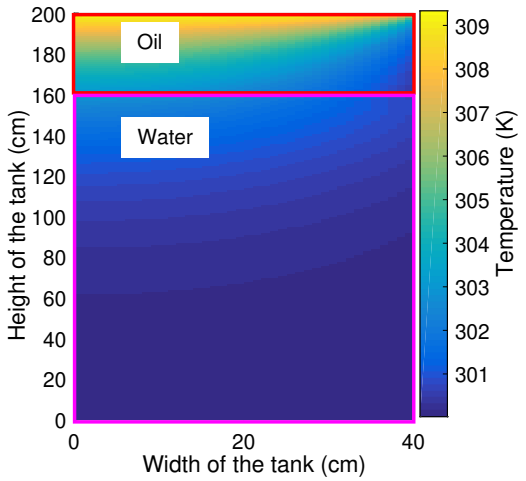
| Parameter | Symbol | Value | Unit |
|---|---|---|---|
| Emissivity of the tank surface | $\varepsilon$ | 0.96 | |
| Tank surface temperature | $T_{sf}$ | 323.00 | $K$ |
| Atmospheric temperature | $T_{atm}$ | 273.15 | $K$ |
| Atmospheric absorptivity | $\alpha_{atm}$ | 0.96 | |
| Solar absorptivity | $\alpha_s$ | 0.21 | |
| Solar constant | $S_c$ | 1368.00 | $W/m^2$ |
| Convection at the top of the tank | $h$ | 5.06 | $W/m^2K$ |
| Stefan-Boltzmann constant | $\sigma_{SB}$ | $5.67\times10^{-8}$ | $W/m^2K^4$ |

To identify the fluid level with FBG temperature sensors, we simulate different layers heights inside the storage tank, where the total level is always 200 cm. The upper and lower bounds of level variations are specified as follows:

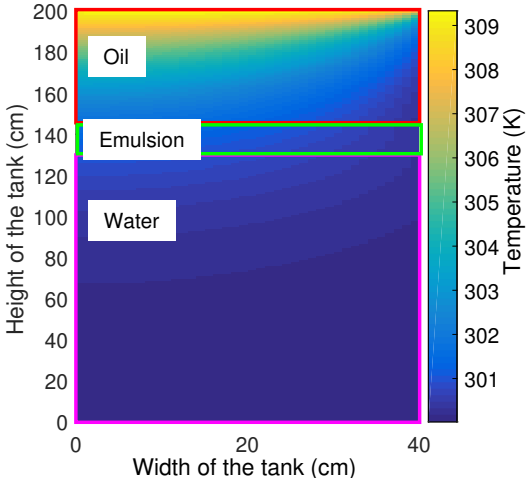
- Water varies from 40 cm to 160 cm;
- Emulsion varies from 0 cm to 20 cm;
- Oil varies from 40 cm to 160 cm.

To understand the temperature behavior in the different fluids, measured at the temperature distribution along the tank is presented in Figure 2(a), 2(b) and 2(c). At 140

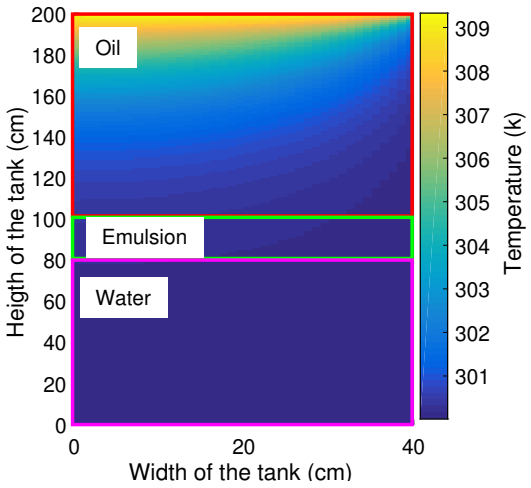
cm in Figure 2(a) we observe water, in Figure 2(b) emulsion and Figure 2(c) oil, these temperatures are presented in Table 2. It is possible to observe in Table 2 that water at the same height as other fluids has a higher temperature, this is due to its higher thermal conductivity.




(a) Oil= 40 cm, Emulsion= 0 cm and Water= 130 cm.




(b) Oil= 55 cm, Emulsion= 15 cm and Water= 80 cm.




(c) Oil= 100 cm, Emulsion= 20 cm and Water= 80 cm.

Figure 2 – Oil tank temperature distribution with different fluid levels.

Based on the simulation the temperature distribution inside the tank, we simulate the presence of FBGs temperature sensors inside it. This sensor is 200 cm long and 125 $\mu m$ in diameter, with different FBGs distributions along the fiber. The objective is to correctly identify the oil according to the collected temperature variations. For this, we analyze $\Delta\lambda$ and FWHM variations. These two parameters vary with the temperature changes in which the fiber is exposed. $\Delta\lambda$ changes due to the effect of thermo-optic and thermal expansion, while FWHM changes due to the effect of chirp in FBG when it is not submitted to a

Table 2 – Position and temperature in relation to the top of the tank for the fluids studied.

| **Fluid** | **Position** $(cm)$ | **Temperature** $(K)$ |
|---|---|---|
| Water | 140 | 301.1314 |
| Emulsion | 140 | 301.2189 |
| Oil | 140 | 301.3564 |

uniform temperature along the grating region (SILVA et al., 2007). In this analysis, we consider that there is temperature distribution in intervals of the physical length of the FBG (1 cm), and that there is no temperature distribution in the fiber diameter.

For numerical analysis of the FBG temperature sensor into the tank, coupled mode equations with a modified transfer matrix formulation (T-matrix) were used to solve coupled mode equations for a large number of grid segments (PRABHUGOUD; PETERS, 2004). We also consider that the temperature is distributed on the $z$-axis. The temperature distribution on the $y$ and $x$-axes were ignored, as the fiber diameter (125 $\mu m$) is smaller than the tank diameter (40 cm) and fiber length (200 cm). The FBG is defined as a periodical disturbance to the refractive index of the optical fiber core (PRABHUGOUD; PETERS, 2004):

$$\delta_{eff} = \bar{\delta}_t eff \left[1 + \nu \cos\left(\frac{2\pi}{\Lambda_0} z + \Phi(z)\right)\right] \tag{3.5}$$

where $\bar{\delta} n_{eff}$ is the average index variation over a grating period, $\nu$ is the fringe visibility of the index change, $\Lambda_0$ is the grating period, $z$ is the cartesian coordinate in the direction of the light propagation in the fiber and $\Phi(z)$ is the grating chirp, which is constant for uniform gratings (LEAL-JUNIOR et al., 2019). Using the coupled mode theory for FBG spectra, the following first order differential equations are obtained (LEAL-JUNIOR et al., 2019):

$$\frac{\partial R(z)}{\partial z = i} = i\sigma R(z) + ikS(z) \tag{3.6}$$

$$\frac{\partial S(z)}{\partial z = i} = i\sigma S(z) - ikR(z) \tag{3.7}$$

where (3.6) and (3.7) describe the propagation of the mode with amplitude $R(z)$ into a counter-propagating mode $S(z)$ (LEAL-JUNIOR et al., 2019). For both equations, $\sigma$ is a 'dc' self-coupling coefficient defined, as presented in (3.8), whereas $k$ is 'ac' coupling coefficient, as shown in (3.9) (LEAL-JUNIOR et al., 2019):

$$\begin{aligned} \sigma &= 2\pi n_{eff}\left(\frac{1}{\lambda} - \frac{1}{(2n_{eff}\Lambda_0)}\right) + \\ &+ 2\frac{\pi}{\lambda}\bar{\delta} n_{eff} f - \frac{1}{2}\frac{\Phi(z)}{z} \end{aligned} \tag{3.8}$$

$$k = \frac{\pi}{\lambda}\nu\bar{\delta} n_{eff} \tag{3.9}$$

The reflectivity of FBG ($r(\lambda)$) can be estimated as:

$$r(\lambda) = \left| \frac{S(-L/2)}{R(-L/2)} \right|^2 \tag{3.10}$$

where, $L$ is the physical length of the grating. The limit conditions for calculating reflectivity are:

$$R(-L/2) = 1 \qquad \text{and} \qquad S(L/2) = 0 \tag{3.11}$$

To solve these equations the T-matrix approximation is employed. In this case, a grid with length $L$ is divided into sections $M$ in the $z$-axis, the field amplitudes in the $i$-th section, $R_i$ and $S_i$, obtaining the following relationship with the T-matrix ($F_i$) (PRABHUGOUD; PETERS, 2004):

$$\begin{bmatrix} R_i \\ S_i \end{bmatrix} = F_i \begin{bmatrix} R_{i-1} \\ S_{i-1} \end{bmatrix} \tag{3.12}$$

$$F_i = \begin{bmatrix} \varphi_1 - i(\sigma/\gamma)\varphi_2 & -i(k/\gamma)\varphi_2 \\ i(k/\gamma)\varphi_2 & \varphi_1 + i(\sigma/\gamma)\varphi_2 \end{bmatrix} \tag{3.13}$$

$$\varphi_1 = \cosh(\gamma \Delta z) \tag{3.14}$$

$$\varphi_2 = \sinh(\gamma \Delta z) \tag{3.15}$$

where, $\Delta z$ as the ratio between $L$ and $M$, and . The main advantage of the T-matrix method is its computational efficiency when compared with direct numerical integration (LEAL-JUNIOR et al., 2019). In Equation 15, the shift Bragg is presented. This FBG sensor is not shifted due to strain, only to temperature.

$$\begin{aligned} \lambda_B &= 2n_{eff}\Lambda \\ &= \lambda_0[1 + (\alpha + \zeta)\Delta T]) \end{aligned} \tag{3.16}$$

where, $\lambda_0$ is the initial value of the Bragg wavelength, alpha is the fiber thermal expansion coefficient $\zeta$ is the thermo-optical coefficient, $\Delta T$ is the temperature variation. The parameters used to simulate the spectra are listed in Table 3.

The simulated spectra presented in 3 show that the different temperatures cause differences in the FBG spectrum profile. The observed variations in FWHM are less significant than the $\Delta\lambda$ shifts. However, as we observe variations in both, we choose to use both parameters. The FBG exposed to a temperature of $300.05K$ is represented in Figure 3 by the green curve. If compared to the initial spectrum, in the FWHM there was a decrease in the fourth decimal place. The pink curve, the FBG was exposed to $302.50K$, and besides the detachment of the wavelength, there was an increase in the FWHM. In the case of the red curve ($309.17K$), besides the displacement, we observed an increase in FWHM compared to the initial spectrum, but lower than the FWHM

Table 3 – Parameters employed in the simulation of the FBG spectrum for temperature and fluid variations inside the tank.

| **Parameter** | **Symbol** | **Value** | **Unit** |
|---|---|---|---|
| Nominal period | $\Lambda_0$ | 556 | $nm$ |
| Fringe visibility | $\nu$ | 1 | |
| Initial Bragg wavelength | $\lambda_0$ | 1540 | $nm$ |
| Effective refractive index | $n_{eff}$ | 1.39 | |
| Number of grating | $M$ | 200 | |
| Grating length | $L$ | 1 | $cm$ |
| Average index variation | $\bar{\delta}n_{eff}$ | $1\times10^{-4}$ | |
| Thermal expansion coefficient | $\alpha$ | $277.25\times10^{-7}$ | $K^{-1}$ |
| Thermal-optic coefficient | $\zeta$ | $280.95\times10^{-4}$ | $K^{-1}$ |

observed at $302.50K$. In all curves no reduction of reflectivity is observed. The interest is the correct classification in oil, so the other fluids are classified as non-oil. We started the selection process at 200 equidistant points, then 6, 8, 10, 12, 25, 50 and 100, with different distributions along the fiber and without the requirement of equidistant. The oil and non-oil classification were done by RF, using cross validation, dividing the data in training (65% data, randomly selected) and test (35% remaining). Using the training, we taught the classification criteria to the machine. The forecast or classification were performed at each condition of oil/water mixture of the test samples (35% of the samples). The RF forecast is simple, the test data is evaluated in each of the decision trees created by the RF in the training, and associated with more classes with similar characteristics.

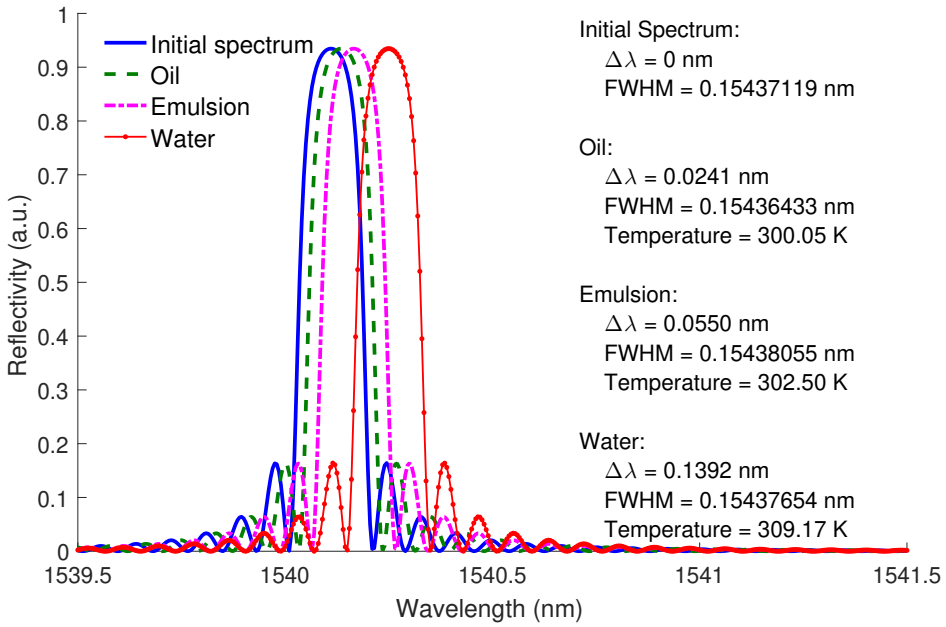



Figure 3 – FBG spectra for three simulated temperature conditions. The initial spectrum is also presented for comparison purposes.

## 3.2 Experimental setup

### 3.2.1 Experimental setup for fluid detection and level estimation

In the experimental analysis, the following fluids were added in 22.5 cm height beaker: water, mineral oil (Oil), or Kryo 51. Then, the fluids were heated using a peltier at the bottom of the beaker, until it reaches a temperature of 313.15 K. Inside the container there are 3 FBGS, located at 3 cm with respect to the bottom of the beaker, 10.5 cm, and 21.5 cm, is near the top, as shown in Figure 4. The cycle is defined as level reduction of 5.5 cm at every 3 minutes until it reaches 3 cm, and then the increasing level in the same way, respecting the descent time and heigth. A total of three cycles per fluid were made. In this way, it is possible to identify the fluid that is in contact with the FBGs, and also to estimate its level.

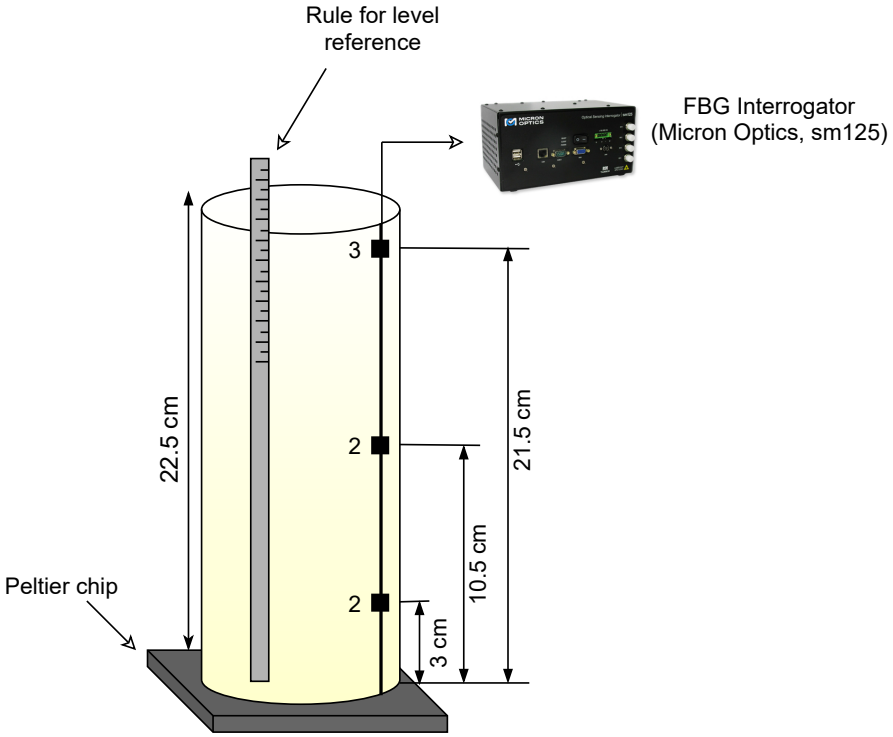



Figure 4 – Experimental setup employed of the three FBGs for level measurement and fluid identify.

### 3.2.2 Experimental setup for liquid storage tank

For the experimental setup, we used a tank of 100 cm height and 30 cm width. We positioned 9 FBG sensors inscribed in photosensitive single-mode fiber (SMF) in the tank, as shown in Figure 5. Although the distances between the FBGs are close to 10 cm, they are not uniform, especially between FBGs 3 and 4 as well as FBGs 1 and 2. It is worth noting that this non-uniform distribution does not interfere on the liquid level estimation, as the distances between the FBGs are already known. Furthermore, these differences in the FBGs distances are related to the FBG arrangements in the fiber, since there is a splice between 3 SMFs (with FBGs inscribed), which resulted in this higher distances between some FBGs. The sensor array was installed by gluing one end of the optical fiber in the top of the tank, whereas the optical fiber is attached to small grooves in the tank wall to avoid lateral movements of the sensor array. The FBG sensors responses are acquired by the optical interrogator sm125 (Micron Optics, USA) with a sampling rate of 0.5 Hz. The tank was filled with water at room temperature. Then, we insert an electric boiler in the water for 10 minutes, which was positioned close to FBG 7. Thereafter, we started data acquisition and, after each 5 minutes interval, the liquid level is reduced by opening the tank drain valve at its bottom. Each cycle is defined as a level reduction of 10 cm every 5 minutes until it reaches 20 cm and 3 cycles were performed. The mean temperatures observed after removal of the electric boiler, as well as their standard deviations, can

be seen in Table 4. From Table 4, we can state that the high temperature of FBG 7 is explained because the electric boiler was close to it.

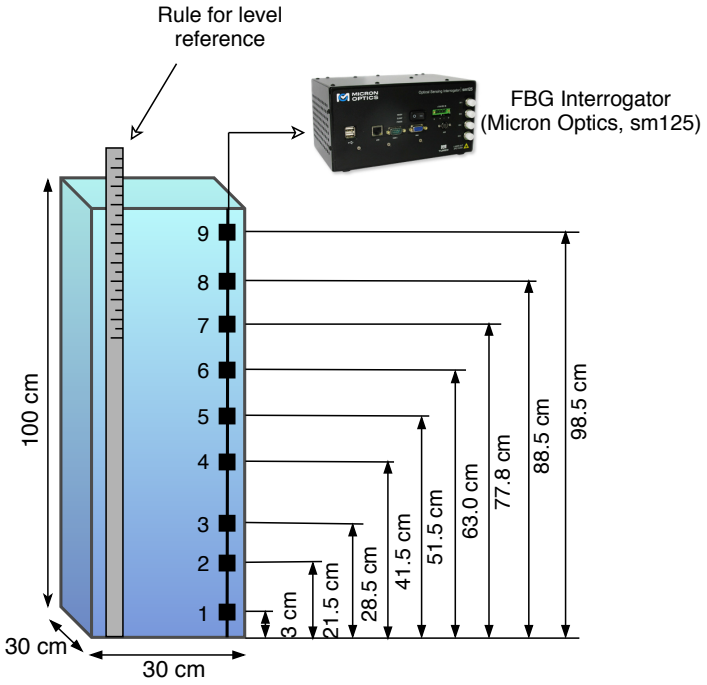



Figure 5 – Experimental setup employed of the nine FBGs for level measurement.

Table 4 – Temperature after removing the electric boiler, and standard deviation.

| Sensor | Mean | Standard deviation |
|---|---|---|
| **FBG 1** | 30.12 | 0.0057 |
| **FBG 2** | 30.49 | 0.0174 |
| **FBG 3** | 30.62 | 0.0126 |
| **FBG 4** | 31.82 | 0.0148 |
| **FBG 5** | 32.05 | 0.0133 |
| **FBG 6** | 32.87 | 0.0129 |
| **FBG 7** | 34.4 | 0.0031 |
| **FBG 8** | 33.8 | 0.0118 |
| **FBG 9** | 34.08 | 0.0547 |

The goal of the analysis is level identification through a SMF with 9 FBG temperature sensors. The technique that best fits the classification is the most accurate, sensitive and specific among the tests. Models created have a single input variable, $\Delta\lambda$. The FBGs 1 and 9 are taken from the model. The FBG 1, located 3 cm from the bottom of the tank, is always covered. The FBG 9, on the other hand, is always uncovered. Its removal is done to avoid the insertion of noise in the analysis. These data are unbalanced by nature and to correct this problem we use oversampling. Oversampling consists of sampling with

replacement of the least representative class to balance both classes (LI et al., 2016). This technique is useful for the elaboration of models based on ML, because they bring more equity of selection probability between classes. In this way, we ensure that both classes have the similar number of observations, providing similar sampling. We use cycles 1 and 2 as training set and cycle 3 as test set of the algorithm.

The need for both experiments was to demonstrate an experiment similar to a real problem, shown in Figure 5, and one with controlled temperature, see Figure 4. As the objective is to estimate the fluid level, we use ML techniques to estimate the level through the wavelength shift of each sensor. In order to emulate an operation condition in many industrial processes, the initial temperature is not constant, as in industrial processes the fluid generally reaches the tanks with a higher temperature than the liquid inside it.

# 4 Results and discussion

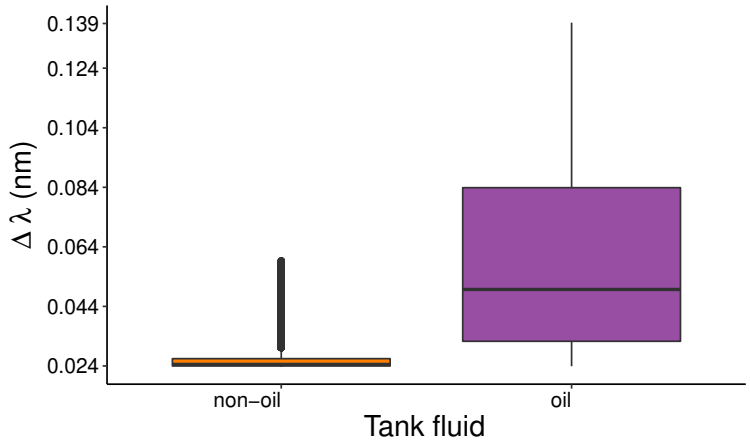
## 4.1 Simulation Results

Figure 3 presents the wavelength shifts and FWHM changes that occur with temperature variation. In Table 2 and Figures 2, it is shown that under the same conditions and height we identify differences in water, emulsion and oil temperature. Based on this temperature difference between the fluids, we want to find the ideal number and location of FBGs. For ensuring the identification of oil with accuracy higher than 95%, sensor construction economy and higher spectral efficiency.

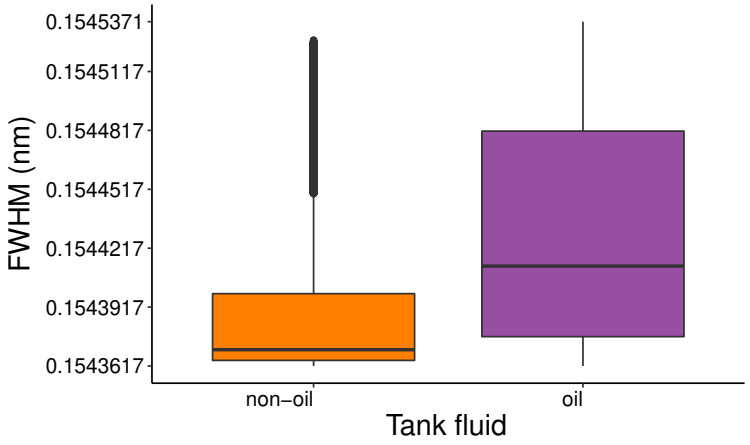
The RF input variables were $\Delta\lambda$, FWHM, and the location of the FBG relative to the tank. The expected output in the algorithm, based on these inputs, was the fluid classification: oil or non-oil. In terms of the location of each FBG (a RF input, as mentioned above), it is also possible.

The variables for RF initialization were: $\Delta\lambda$ , FWHM and the height of the FBG on the fiber in relation to the tank bottom. The RF have as output the oil and non-oil classification. Indirectly, it is possible to estimate the oil level based on the classification. Figure 6(a) shows the association of $\Delta\lambda$ with the classes. Note that only some values above the third quartile of the non-oil class intersect with the oil class. Thus, $\Delta\lambda$ separates almost correctly the two classes. For Figure 6(b) the FWHM cannot correctly divide two classes, thus the high intersection between classes compared to 6(a). Thus, we can assume that the FWHM is less significant than $\Delta\lambda$ for separation. For 6(a) and 6(b), due to the lack of symmetry between the median and the first and third quartiles, representing the non-normality of the data. To guarantee that we are not inserting multicollinearity in the model, we calculate the correlation ($\rho$) between FWHM and $\Delta\lambda$ , obtaining $\rho$=0.29. Based on the observed $\rho$ we verify that there is no increase of variability due to repeated information in RF.

To find the ideal number of FBGs and their locations, different points were tested based on RF accuracy. As there are a large number of localization possibilities for each case, Table 5 presents the location that resulted in the highest accuracy. In general, the FBGs very close to the top of the tank had high temperatures due to radiation. We assume that the FBGs along the fiber are distributed below 10 cm, except for 200 FBGs. For fibers with 6, 8, 10, 12, 25, 50 and 100 simulated FBG sensors, because there are a large number of localization possibilities for each case, we randomly selected 10 sensor distributions. Then, the location that returned the highest accuracy in each fiber was selected. Fibers with 6, 8, 10 and 12 FBGs were selected based on the points that classified correctly for




(a) Box plot of $\Delta\lambda$ for oil and non-oil classification




(b) Box plot of FWHM for oil and non-oil classification

Figure 6 – Tank fluids as function of $\Delta\lambda$ and FWHM, to find characteristics that facilitate the classification of fluids.

more times in the previous cases. The location of FBGs for fibers with lower FBGs was:

- 6 FBGs: 8 cm, 44 cm, 80 cm, 116 cm, 152 cm, 188 cm;
- 8 FBGs: 15 cm, 40 cm, 65 cm, 90 cm, 115 cm, 140 cm, 165 cm, 190 cm;
- 10 FBGs: 11 cm, 29 cm, 49 cm, 69 cm, 89 cm, 100 cm, 130 cm, 151 cm, 171 cm, 190 cm;
- 12 FBGs: 10 cm, 20 cm, 39 cm, 50 cm, 70 cm, 89 cm, 110 cm, 130 cm, 149 cm, 170 cm, 190 cm.

Table 5 shows the number of trees chosen and the accuracy. The second column shows the number of trees used in the algorithm. And, we guarantee that each class would have contact with at least 1 FBG.

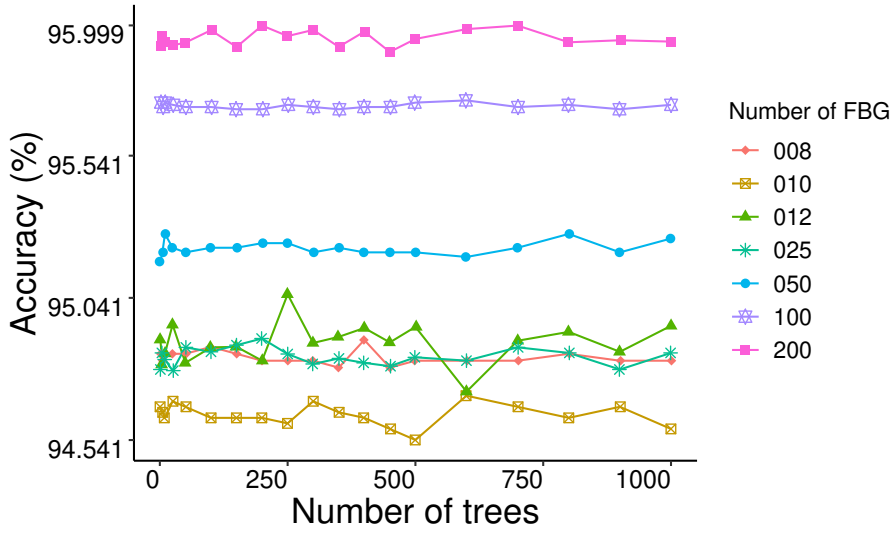
Table 5 – Oil classification using RF algorithm.

| **Number of FBG** | **Oil tank prediction using RF algorithm** | |
|---|---|---|
| | **Number tree** | **Average accuracy (%)** |
| 200 | 200 | 95.95 |
| 100 | 250 | 95.71 |
| 050 | 200 | 95.21 |
| 025 | 600 | 94.83 |
| 012 | 800 | 94.88 |
| 010 | 600 | 94.63 |
| 008 | 400 | 94.83 |
| 006 | 200 | 81.89 |

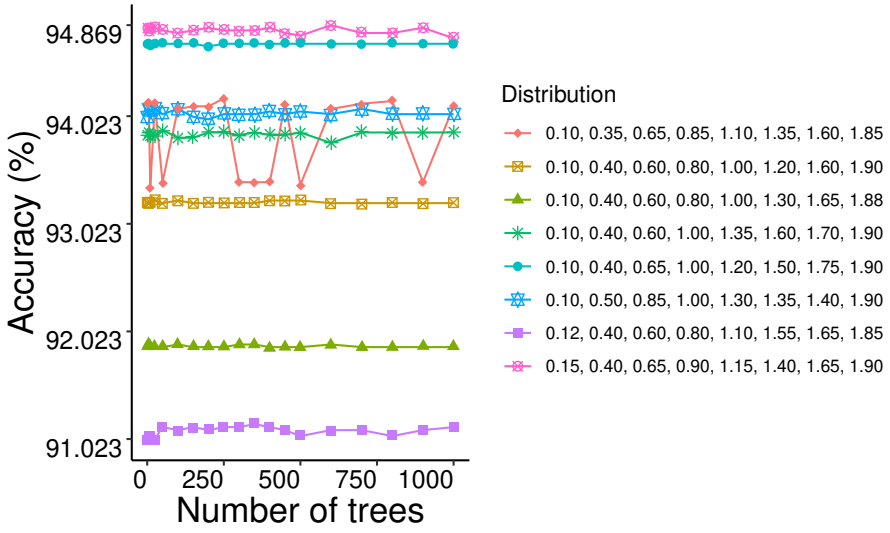
Figure 7(a) shows the accuracy by the number of RF trees with different quantities and distributions of FBGs. The highest accuracy was found with the highest number of FBGs. However, inscribing many FBGs into the fiber is expensive and time consuming, and interrogators have limited wavelength ranges. Thus a high number of sensors implies fewer tanks to be monitored simultaneously. A low number of sensors ensures higher spectral efficiency. Thus, a high number of sensors, such as: 200, 100 or 50 FBGs, are not among the most efficient options. Reducing the FBGs to 12, we find the average accuracy of 94.88%, with 10 we observe 94.63% and 8 FBGs 94.83%. The difference between 8, 10 and 12 FBGs is lower than 0.2%. In reducing to 6 FBGs there was a 10% lower accuracy, so we chose 8 FBGs as the optimal number for this simulation.

Based on the accuracy, the classification in oil and non-oil, is mutual exclusive, thus, has Bernoulli distribution. Hence, the accuracy can be tested with 8 and 200, 25, 50 and 100, using the Tests for Two Proportions. The p-values with a significance of 5% are presented to compare the 8 FBGs array with the other ones. Comparing the accuracy of a 8-FBGs array with the one of 200 FBGs, we obtain a p-value= 0.0003742, at a 5% significance level, which indicates a higher accuracy of the 200 FBGs array. The same occurs with 100 FBGs, with a p-value of 0.007578. However, when comparing the 8 FBG sensor with a 50 FBG sensors and a 25 FBG sensors, we can state that both are significantly equal, with p-values of 0.292 and 0.9888 respectively. Thus, a temperature sensor with 8 FBGs has the same accuracy as sensors with 25 or 50 FBGs. The choice of the distributions in Figure 7(b), was made by analyzing the scenarios with 25, 50, 100 and 200 FBGs, in the places with FBGs that classified correctly more often. The distributions with less accuracy in Figure 7(b), have irregular spacing between the FBGs, unlike the distributions with greater accuracy. The location of the FBG influences the observed accuracy.

In the Table 6, we have the measures of mean accuracy, and number of trees, by using 8 FBGs with different distributions presented in Figure 7(b).




(a)




(b)

Figure 7 – **(a)** Number of trees per accuracy in the best prediction based on the RF algorithm. **(b)** Number of trees per accuracy when using 8 FBGs in different distributions.

Table 6 – Location in relation to the bottom of the tank, accuracy and number of trees, for sensors tested with 8 FBGs in different distributions via RF for oil classification

| **Number FBGs Distribution** | **Location (cm) of the FBGs** | | |
|---|---|---|---|
| | **Location (m)** | **Accuracy** | **Tree** |
| 1 | 15, 40, 65, 90, 115, 140, 165, 190 | 94.83 | 600 |
| 2 | 10, 50, 85, 100, 130, 135, 140, 190 | 94.05 | 700 |
| 3 | 10, 40, 60, 100, 135, 160, 170, 190 | 93.85 | 50 |
| 4 | 10, 40, 60, 80, 100, 120, 160, 190 | 93.32 | 25 |
| 5 | 10, 40, 65, 100, 120, 150, 175, 190 | 93.86 | 300 |
| 6 | 10, 40, 60, 80, 100, 130, 165, 188 | 91.88 | 350 |
| 7 | 12, 40, 60, 80, 110, 155, 165, 185 | 91.09 | 400 |
| 8 | 10, 35, 65, 85, 110, 135, 160, 185 | 94.79 | 700 |

## 4.2 Temperature Characterization

As we estimate the liquid level from FBG temperature sensors, only temperature variations are considered, disregarding any stress effect. In a FBG the Bragg wavelength $\lambda_B$ is given by:

$$\lambda_B = \lambda_0[1 + (\alpha + \zeta)\Delta T] \tag{4.1}$$

where, $\lambda_0$ is the initial value of the Bragg wavelength, is the thermo-optical coefficient, $\Delta_T$ is the temperature variation.

Prior to the liquid level estimation, the temperature characterization, shown in Figure 8, is performed without transverse force acting on the fiber, for the temperatures of 298.15 K, 303.15 K, 308.15 K, 313.15 K, 318.15 K and 323.15 K. The results presented in Figure 3 show the offset caused by the temperature in the sensor response. The error bars represent the standard deviation and the dots the mean value of wavelength shift. This FBG has sensitivity of 10.93 pm/K in increasing and decreasing of 10.97 pm/K. Measurements were taken independently in a total of 6 temperature cycles (from 298.15°C to 323.15°C) to ensure the repeatability of the sensor. The repeatability of the increase was 0.091 and the decrease 0.109. We observed a hysteresis of 0.95 %. The measured uncertainty is 0.105% to increasing and 0.106% to decreasing. In addition, both coefficients (R2) for the linear regression higher than 99%.

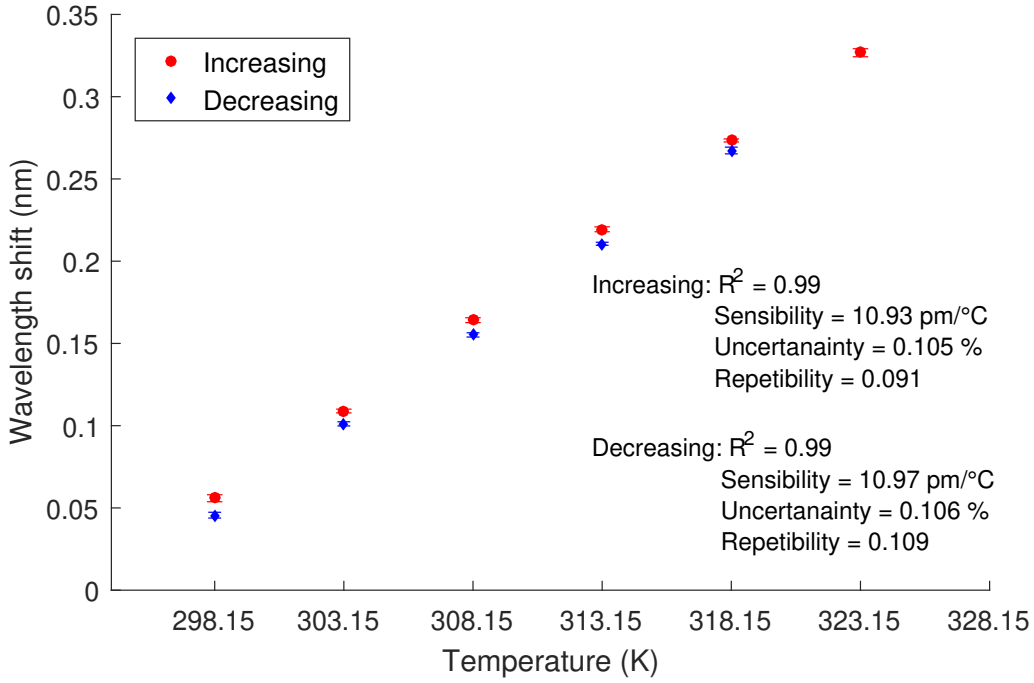



Figure 8 – Increasing and decreasing temperature response of FBGs.

## 4.3 Experimental results for fluid detection and level estimation

The experimental setup consists of a 22.5 cm beaker, used for the detection of three distinct fluids (water, mineral oil and Kryo 51) and the estimation of their levels. Similar to the previously commented Section 3.1, we used RF. The RF was used both for the identification/classification of the fluids and for the estimation of their levels.

In Figures 9(a) and 9(b), the 4 FBGs of the experiment are presented, we note the stability of the FBG Air, as expected, since it does not suffer with the change of heat of water, that has higher thermal conduction coefficient. Figures 9(c) and 9(d) show the wavelength variation when measuring the levels for Kryo 51. Similar to Figure 9(a), the FBG Air shows stability. In the other FBGs we notice the gradual increases, as the level goes lower in 9(a) or higher, in 9(b). In contrast to Figures 9(c) and 9(d), which show the ascent (9(e)) and descent (9(f)) levels of the mineral oil have unusual behavior. Even when exposed to the same process as the previous fluids.

To understand the relationship between the FBGs studied for each fluid, we understand the linear dependence relationship between them. Thus, for each fluid, a table was created to understand how linearly related the variables are. This study is important to avoid multicollinearity of the data. Multicollinearity causes the variance of the model to increase, causing errors in prediction. In general, the literature recommends that the variables be evaluated two on the basis of Pearson correlation coefficient ($\rho$). If $\rho$ greater

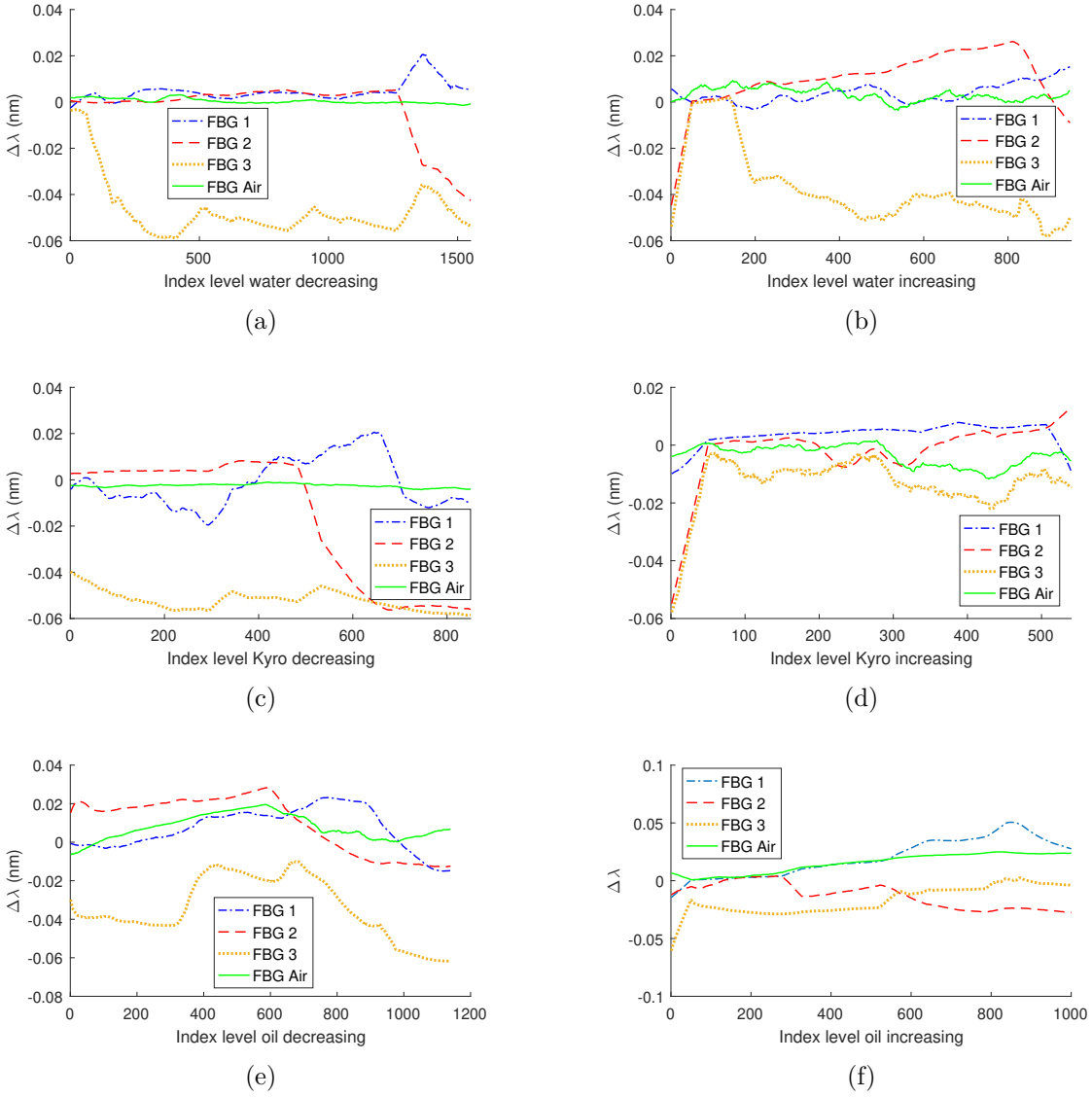



Figure 9 – **(a)** FBG temperature sensors array response on the cycle 1 in level decreasing of water, with standard error (SE) of 0.14% in cycle 2 and 0.097% in cycle 3. **(b)** FBG temperature sensors array response on the cycle 1 in level increasing of water. With SE of 0.1% in cycle 2 and 0.12% in cycle 3, with SE FBG temperature sensors array response on the cycle 1 in level decreasing of mineral oil, with SE of 0.155% in cycle 2 and 0.147% in cycle 3. **(d)** FBG temperature sensors array response on the cycle 1 in level increasing of mineral oil, with SE of 0.074% in cycle 2 and 0.08% in cycle 3. **(e)** FBG temperature sensors array response on the cycle 1 in level decreasing of Kryo 51 oil, with SE of 0.108% in cycle 2 and 0.112% in cycle 3. **(f)** FBG temperature sensors array response on the cycle 1 in level increasing of Kryo 51 oil, with SE of 0.099% in cycle 2 and 0.13% in cycle 3.

than |0.85| one of the variables is removed, so as not to add duplicate information to the model, the term multicollinearity (SCHROEDER; LANDER; LEVINE-SILVERMAN, 1990).

Table 7 presents Pearson correlation for water. In general, all correlations calculated for the FBGs in the water experiment are low, except when comparing FBG 1 with FBG Air, both associated with level decrease.

Table 7 – Pearson correlation between the observed FBGs for both water level increase and decrease.

| **Water** | **Level Decreasing** | | | |
|---|---|---|---|---|
| | **FBG 1** | **FBG 2** | **FBG 3** | **FBG Air** |
| **FBG 1** | 1.00 | -0.53 | -0.13 | 0.6 |
| **FBG 2** | | 1.00 | 0.01 | -0.13 |
| **FBG 3** | | | 1.00 | -0.02 |
| **FBG Air** | | | | 1.00 |
| | **Level Increasing** | | | |
| | **FBG 1** | **FBG 2** | **FBG 3** | **FBG Air** |
| **FBG 1** | 1.00 | -0.28 | -0.17 | 0.60 |
| **FBG 2** | | 1.00 | 0.04 | -0.18 |
| **FBG 3** | | | 1.00 | -0.13 |
| **FBG Air** | | | | 1.00 |

Analyzing the correlations of Kryo 51, Table 8, we observed a high variability (in modulus), of the FBG 1 with the other variables of the study. However, it is lower than the pre-established limit. In Table 9, for the mineral oil, we note that, due to the variation of the FBG Air, we observed a correlation percentage higher than 30% with the other variables. FBG Air as it is a variable that has no contact with the fluids, we chose to remove it so as not to add errors to the modeling and use it as a reference sensor for room temperature.

Table 8 – Pearson correlation between the observed FBGs for both mineral oil level increase and decrease.

| **Mineral Oil** | **Level Decreasing** | | | |
|---|---|---|---|---|
| | **FBG 1** | **FBG 2** | **FBG 3** | **FBG Air** |
| **FBG 1** | 1.00 | -0.36 | 0.60 | 0.48 |
| **FBG 2** | | 1.00 | 0.03 | -0.17 |
| **FBG 3** | | | 1.00 | 0.15 |
| **FBG Air** | | | | 1.00 |
| | **Level Increasing** | | | |
| | **FBG 1** | **FBG 2** | **FBG 3** | **FBG Air** |
| **FBG 1** | 1.00 | -0.25 | -0.61 | 0.53 |
| **FBG 2** | | 1.00 | -0.34 | 0.09 |
| **FBG 3** | | | 1.00 | 0.03 |
| **FBG Air** | | | | 1.00 |

Table 9 – Pearson correlation between the observed FBGs for both Kryo 51 oil level increase and decrease.

| **Kryo 51 Oil** | **Level Decreasing** | | | |
|---|---|---|---|---|
| | **FBG 1** | **FBG 2** | **FBG 3** | **FBG Air** |
| **FBG 1** | 1.00 | -0.02 | 0.11 | 0.31 |
| **FBG 2** | | 1.00 | 0.69 | 0.53 |
| **FBG 3** | | | 1.00 | 0.77 |
| **FBG Air** | | | | 1.00 |
| | **Level Increasing** | | | |
| | **FBG 1** | **FBG 2** | **FBG 3** | **FBG Air** |
| **FBG 1** | 1.00 | 0.08 | 0.42 | 0.77 |
| **FBG 2** | | 1.00 | 0.82 | 0.62 |
| **FBG 3** | | | 1.00 | 0.86 |
| **FBG Air** | | | | 1.00 |

In this analysis, we first identified which fluid the interface level is positioned. Then, we calculate its level. For the classification, using RF we obtained 100% accuracy in detecting the fluids. Table 10 shows a identity matrix resulting from the correct classification of the studied fluids. Therefore, when using an FBG temperature sensor, there were no errors in the identification of the fluids.

Table 10 – Confusion matrix for fluid identification.

| | **Water** | **Mineral Oil** | **Kryo 51** |
|---|---|---|---|
| **Water** | 1 | 0 | 0 |
| **Mineral Oil** | 0 | 1 | 0 |
| **Kryo 51** | 0 | 0 | 1 |

The RFReg is used to estimate the level. Thus, once the fluids are classified, we create a model for each of them, dividing them into rising and falling. We take into account two situations, one in which the FBGs have no interaction with each other, called Model 1, and another, Model 2, which does not take them into account. Using RF to construct the trees, each model uses the following variables:

- Model 1: FBG 1, FBG 2, FBG 3, (FBG 1)×(FBG 2), (FBG 1)×(FBG 3), (FBG 2)×(FBG 3) and (FBG 1)×(FBG 2)×(FBG 3).
- Model 2: FBG 1, FBG 2 and FBG 3.

Table 11 shows the RMSE for each fluid, separating them into Model 1 and 2, both for increasing and decreasing. In order to reduce the computational effort in the choice of the model, we prioritized Model 2, due to its parsimony. Thus, even though it has a lower RMSE, Model 1 is only chosen when the difference in its RMSE with the compared model is higher than 5%, otherwise Model 2 is selected.

Table 11 – RMSE (in cm) for the proposed models, separated by fluid and level incresing or decreasing.

| Fluid | RMSE level decreasing | | RMSE level increasing | |
|---|---|---|---|---|
| | Model 1 | Model 2 | Model 1 | Model 2 |
| **Water** | **0.367** | 1.620 | **0.382** | 1.712 |
| **Oil** | 0.136 | **0.155** | 0.166 | **0.156** |
| **Kryo** | **0.247** | 0.446 | **0.255** | 0.444 |

Figure 10 shows the estimation of the water level as a function of the observed level based on Model 1, for both level decrease and level increase. Model 1 was chosen, due to its lower RMSE than Model 2.

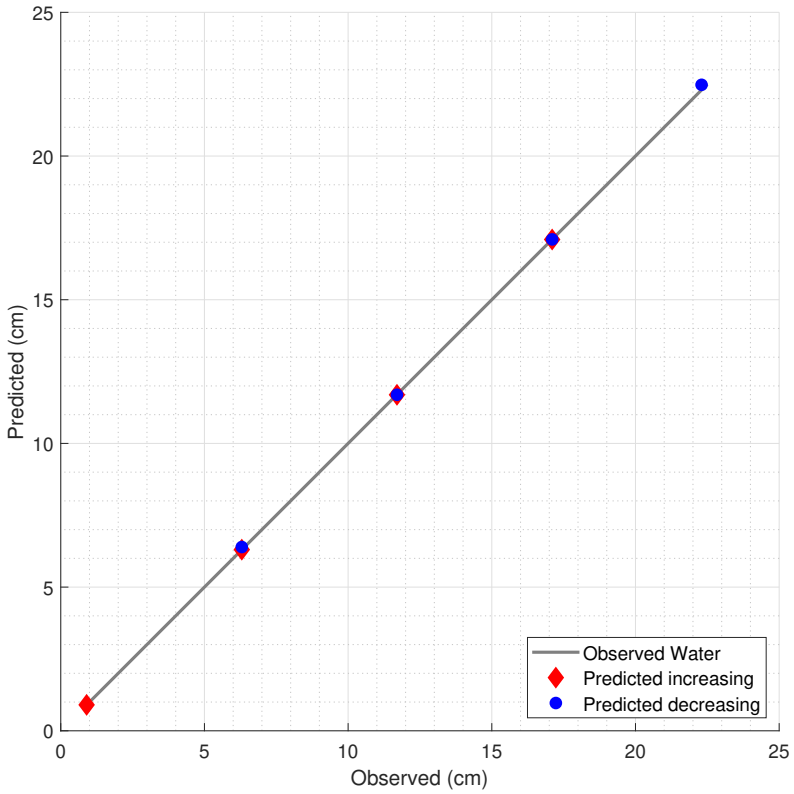



Figure 10 – Prediction of water level variation.

For mineral oil, Figure 11, Model 2 was selected, both for the decrease and the increase of the level. The choice of Model 2 was made due to iat parsimony of the model and avoiding overfitting, since, even with the RMSE of Model 1 lower for the decrease, the difference was only 1.9%. Thus, the choice of a model without interaction,which requires fewer estimated coefficients, compared to the model with interactions, becomes feasible.

Figure 12 presents the prediction of Model 1, both for level decrease and increase, and the expected straight line values. We note that, for all observed fluids, the RF has

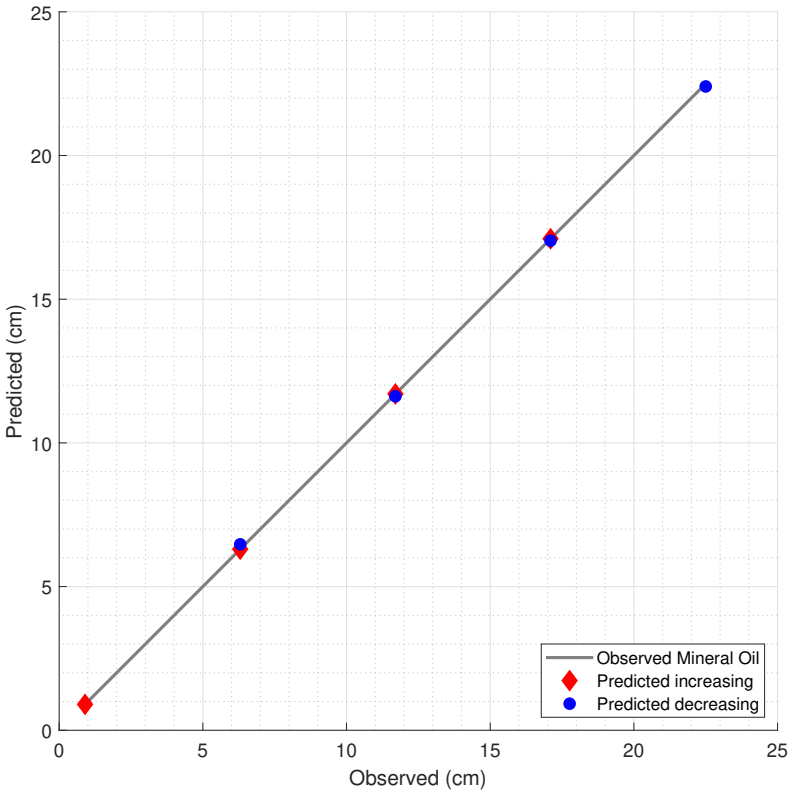



Figure 11 – Prediction of mineral oil level variation.

high predictive power, generating a maximum RMSE of 0.4 cm for level decreasing and 2.4 cm for level increasing.

## 4.4 Experimental Results for liquid storage tank

To expand the experiment, we use a 100cm height tank. Once the possibility of fluid detection has been proven, this experiment uses only one fluid, water. In this way, we can measure the accuracy of the level estimation, in the setup similar to storage tanks in real environment.

### 4.4.1 Classification

One of the three cycles is shown in Figure 13. It is possible to note that as the level reduces, the FBGs exposed to the liquid, present an increase in the wavelength shift and, then, decline. It is worth pointing out that FBG 1 and FBG 9, present irregular pattern when compared to the other gratings. The FBG 1 is completely submerged during the whole experiment. On the other hand, FBG 9 is only submerged in the last level step, the sensor does not detect level changes, and it is used as the temperature reference for the air,

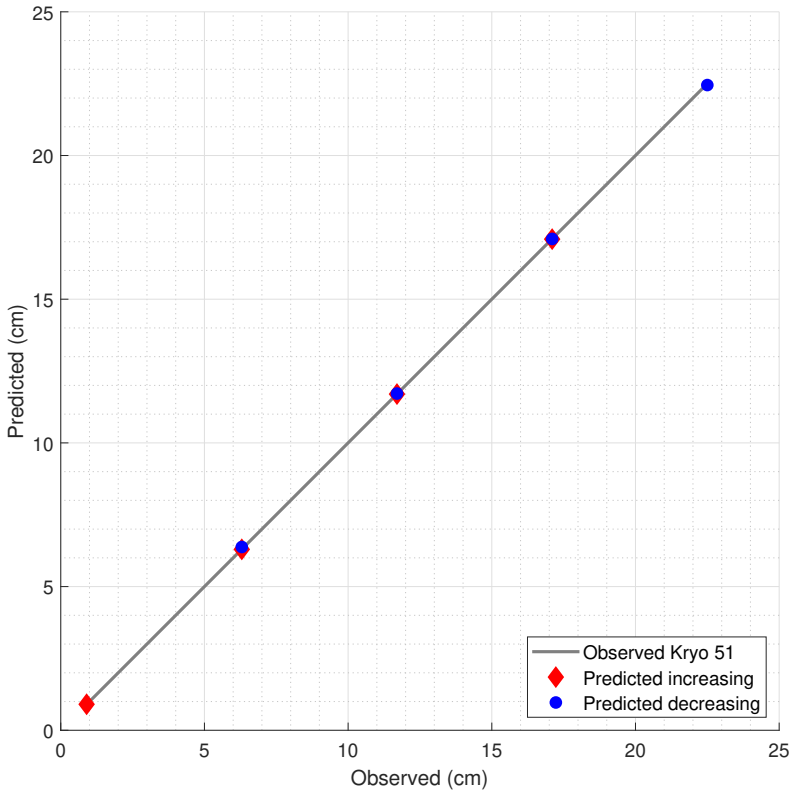



Figure 12 – Prediction of Kryo 51 level variation.

whereas the FBG 1 is a temperature reference of the water. For the experiment, we did three cycles. The goal of this analysis is the level estimation using only FBG temperature sensors.

Figure 14(a) shows the accuracy of the models studied in each FBG. The best performances in terms of accuracy comes from DT and SVM, with an average accuracy of 89.54% and 85.04%, respectively. We also observed higher errors in the initial levels when LogR was used. Sensibility and specificity are secondary measures of accuracy, and consist of quality measures of the adjustment in binary classification problems (HU et al., 2010). For this dissertation, sensibility is defined as the conditional probability that the FBG is in the water, given the FBG actually is. Specificity is the conditional probability that the FBG is not covered in water, given it was classified that it was not. Thus, among the models studied, the one with the highest Sensibility and Specificity is the one with the best suitability for the data. Figure 14(b) shows the average sensibility and specificity of the models, where once again the LogR presented the worst performance (among the ones tested). On the other hand, DT presented superior performance when compared to the others, regarding its Sensibility and Specificity.

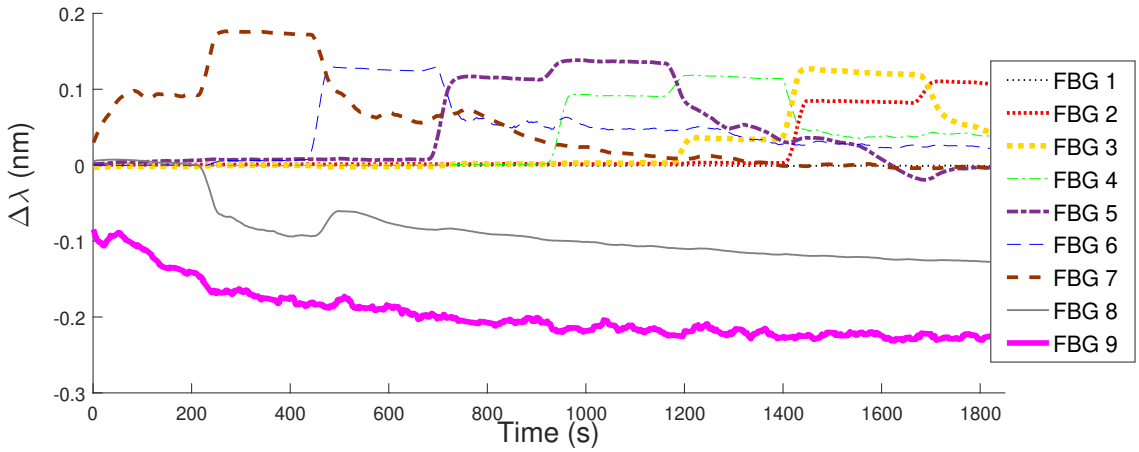
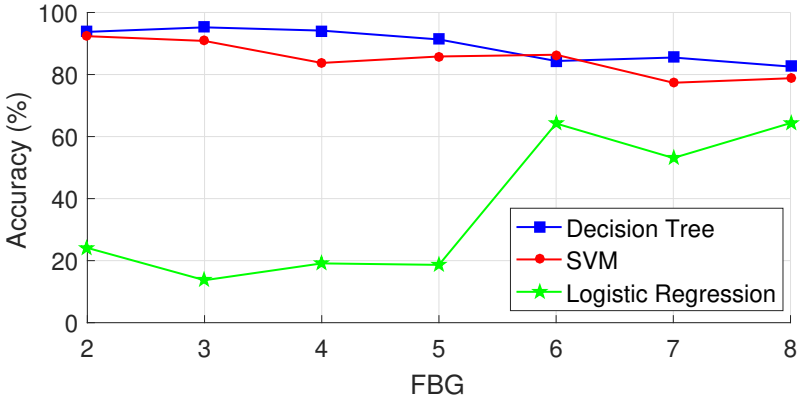



Figure 13 – FBG temperature sensors array response on the cycle 1. With standard error (SE) of 0.0021% in cycle 2 and 0.0025% in cycle 3.

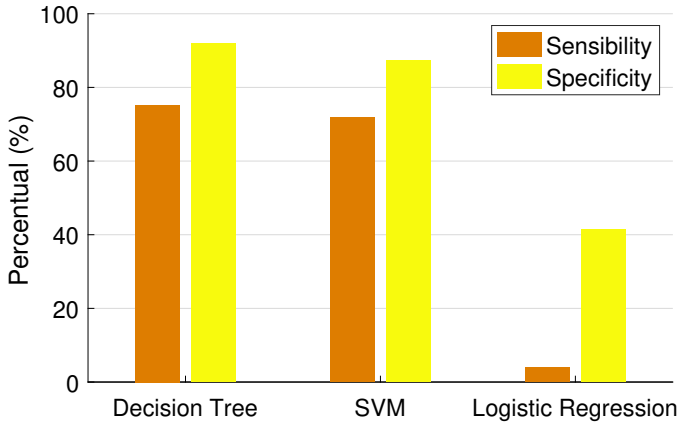
### 4.4.2 Regression

In the experiment, the level varied from 90 cm to 20 cm at 10 cm steps. Thus, FBG 1 is always covered, and FBG 9 almost always uncovered by the liquid. In addition, we removed the transition period between level steps, analyzing only when the level stabilizes. To analyze if the model fits the assumptions of linear regression, the BP test is performed (HALUNGA; ORME; YAMAGATA, 2017). The BP test uses the Lagrange multiplier, useful to test the null hypothesis ($H_0$) that the variances of errors are equal (homoscedasticity) against the alternative hypothesis ($H_1$). Applying the test to the model, we obtained p-value of approximately 0, so we assume that the errors are not equal, heteroscedastic. Since LR is a model not recommended for non-constant variance cases, we propose the WLR, SVM and SVMmin models. Similar to classification, we divide the data into training (cycles 1 and 2) and test (cycle 3). It is also pointed out that as this is an experiment where the temperature input occurs before the level decrease, its decrease is not studied. Thus, it was not possible to detect temperature variations on level increase.

When using RF we observe an RMSE of 16.32 cm, with this we suspect that RF for a problem where the temperature decays, is not the best option. Thus we suggest WLR, SVR and SVRmin. When using just WLR we observed an RMSE of 13.06 cm, with SVR the RMSE was 6.28 cm, and with SVRmin we got 6.18 cm. Figure 15 shows the test group prediction for the proposed algorithms. We notice that the prediction via SVMmin is closer to the line of values Observed at initial levels (20 cm and 30 cm) if compared to the others. Between 40 cm and 60 cm, the proximity to the Observed line occurs when using SVMmin. At 70 cm interval, both SVMmin and SVM present similar proximity to the interest line. While the WLR approaches the Observed at the intervals 80 cm and 90 cm.




(a)




(b)

Figure 14 – **(a)** Accuracy in the level of the models in each FBG based on the algorithms used. **(b)** Sensibility and Specificity in the level of the models based on the algorithms use.

We propose the so-called mixed model (MM), that consists of the selection through cross-validation of the best regression model in each group of levels. In order to cross-validate, we select the model with the minimum RMSE. The MM model is based on the smallest RMSE observed among the proposed algorithms, selects, during training, the algorithm that results in the smallest RMSE for each interval. Then, the MM assigns the selected algorithm to the interval in question. The RMSE obtained for the proposed model is observed in Figure 16.

Figure 16 shows that the smallest RMSE on levels 20 cm, 30 cm and 70 cm is observed when using the SVR, on levels 40 cm, 50 cm and 60 cm, when using the SVRmin and on levels 80 cm and 90 cm the smallest RMSE is observed when using the WLR algorithm. For the MM level prediction, the RMSE information observed in Figure 16 is introduced intuitively.

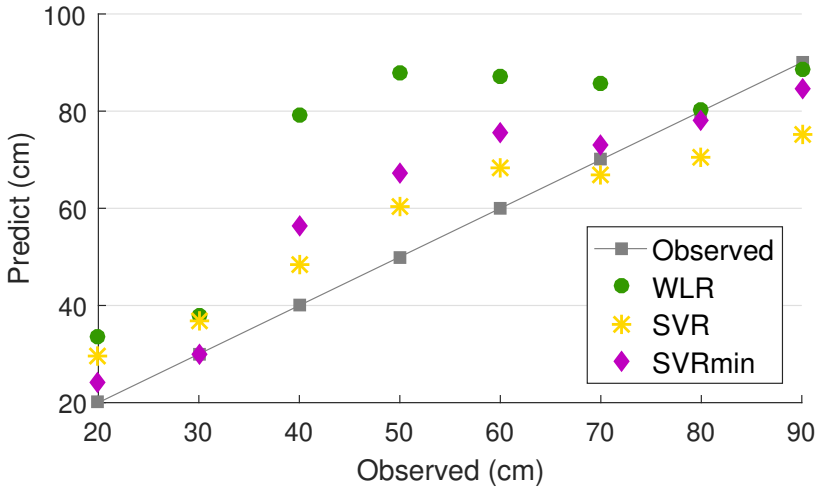



Figure 15 – Level estimation via WLR, SVR and SVRmin.

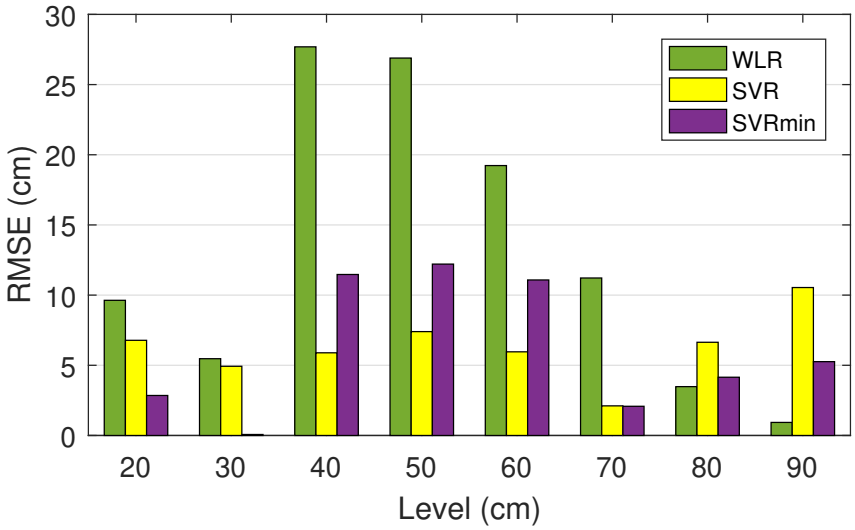



Figure 16 – RMSE for the studied levels of each regression model for the measured data.

Figure 17 shows the RMSE of the algorithms in the intervals selected in Figure 16. This Figure, relates the interval to be used as cutoff point with the RMSE of the prediction. According to Figure 17, RMSE line represents are the minimum RMSE of the algorithms, according to the intervals.

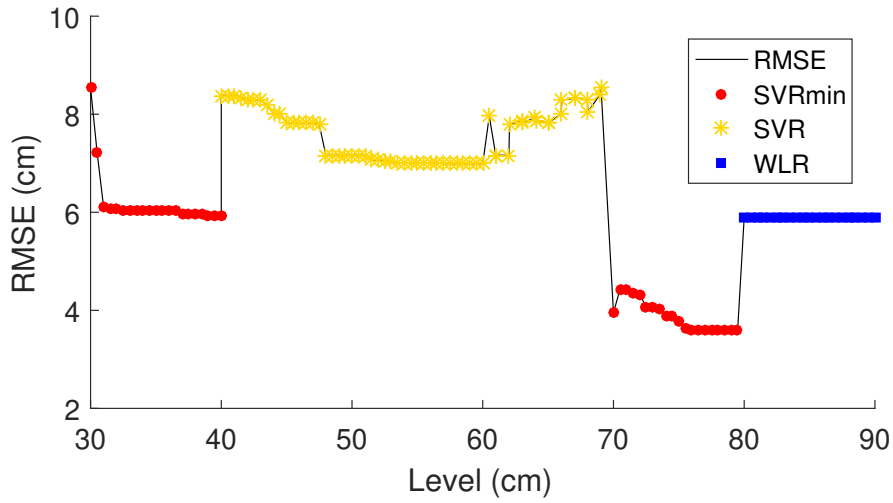



Figure 17 – Cutoff points in level for the RMSE.

Thus, we have obtained the following cutoff points:

- Cutoff 1: G1< 39 cm assumes SVRmin;
- Cutoff 2: 39 cm, G2< 61.5 cm assumes SVR;
- Cutoff 3: 61.5 cm≤ G3< 76 cm assumes SVRmin;
- Cutoof 4: G4≥ 76 cm assumes WLR.

Table 12, presents the RMSE of each algorithm for the data collected. We calculate the RMSE for the other algorithms, where for WLR a RMSE of 13.06 cm was achieved. The RMSE for SVM is 6.28 cm, whereas the one for SVMmin is 6.14 cm. The use of MM results a RMSE of 3.58 cm, in a reduction of RMSE of more than 2.5 cm compared to the algorithms used.

Table 12 – RMSE for each algorithm for the observed data.

| **Level (cm)** | **Algorithms** | | | | |
|---|---|---|---|---|---|
| | **WLR** | **SVR** | **SVRmin** | **RF** | **MM** |
| **20** | 9.63 | 6.78 | 2.85 | 9.09 | 3.35 |
| **30** | 5.47 | 4.93 | 0.07 | 7.29 | 0.24 |
| **40** | 27.69 | 5.89 | 11.47 | 6.43 | 6.28 |
| **50** | 26.90 | 7.41 | 12.21 | 4.90 | 7.42 |
| **60** | 19.23 | 5.96 | 11.07 | 6.26 | 5.23 |
| **70** | 11.22 | 2.11 | 2.08 | 13.64 | 1.96 |
| **80** | 3.48 | 6.64 | 4.15 | 25.79 | 3.08 |
| **90** | 0.93 | 10.54 | 5.26 | 28.64 | 0.94 |

The Figure 18 compares MM and RF. We see that RF does not behave well for problems with a larger temperature range. However, for a problem with low temperature

variation, MM is a more appropriate fit, resulting in a lower RMSE. We also note that both have similar behavior up to the 50 cm level. From the 60 cm level on, RF had no detection sensitivity.

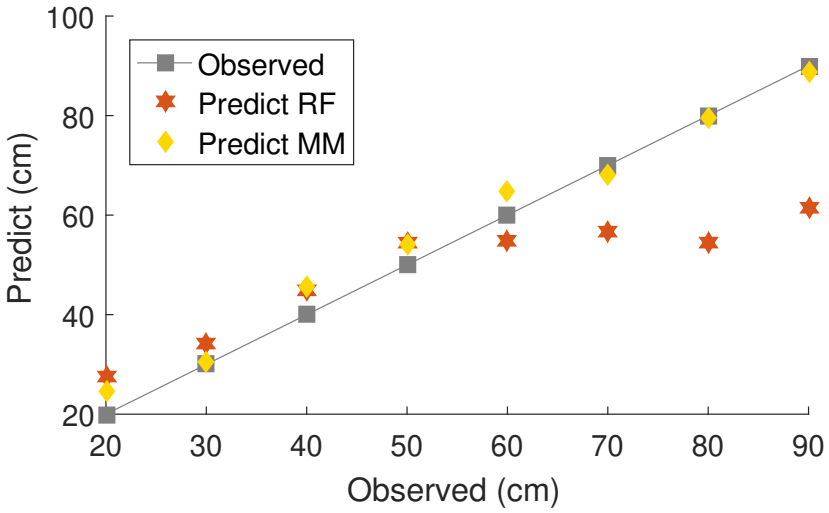



Figure 18 – RMSE comparison between RF and MM.

# 5 Conclusions and future works

The use of FBG sensors in conjunction with ML have been widely employed (YE, 2019; GOLA et al., 2019; ING et al., 2019). For the separation of overlapping spectra generated at wavelength division multiplexing, Manie et al. (2018), applied an ELM. Djurhuus et al. (2019) used FBG temperature sensors' measurements, and in order to improve the sampling frequency they implemented Gaussian process regression (GPR). Jia et al. (2015) featured FBG strain sensor in conjunction with SVM, for prospective in pipeline accident monitoring and safety evaluation.

This dissertation proposes the use of FBG temperature sensors for the identification inside the oil tank. Indirectly, it is possible to estimate the oil level. The simulated tank is 200 cm height and 40 cm diameter, with influence on solar radiation at a zenith angle. For the analysis, we dichotomize the fluids in oil and non-oil. The algorithm inputs are: $\Delta\lambda$, FWHM and the location of the FBG relative to the tank. $\Delta\lambda$ and FWHM were extracted during the simulation. We observed that FWHM is less influential for classification than $\Delta\lambda$. As these parameters are poorly correlated with each other, we chose to keep the FWHM. The parameters were observed based on the temperature variations fluids inside the tank. The RF algorithm was applied to the classification is indicated for data that have low distinction between the classes due to its. To find the ideal number and location, we simulated that inside the tank, in a fiber of 200 cm in length and 125 $\mu m$, there can be from 200 to 6 FBGs with different distributions. The selection of the ideal number was based on the accuracy observed, with respect to the number of FBGs in the fiber. The ideal number of FBGs for the simulation was 8. For the choice we considered that there is lower production cost and higher spectral efficiency if compared to 200 or 100 sensors in the fiber. Then through the Tests for Two Proportions, it was found that when using 8 FBGs at a 5% significance level, its accuracy is equal to using 25 or 50 FBGs. In future investigations, fluid classification is planned based on the FBG found, but without dichotomization and with experimental applications in oil tanks. By using RF in Section, as proposed in the simulation, we obtained accuracy of 100% for fluid identification, and RMSE less than 0.4 cm.

When using the tank, we saw that the RF did not have the ability to detect, for this the MM is proposed. In the experiment, we used a tank 100 cm height with heat insertion via boiler at the top of the tank. We performed 3 cycles of measurement, each cycle is defined as reducing the liquid level at 10 cm steps, every 5 min. We started the experiment with 90 cm and finished with 20 cm. In the analysis, we dichotomize the fluid in if the level is under the observed FBG (class 1), or over the observed FBG (class 0). In this step, we used the algorithms LogR, DT and SVM. The DT gave us the best results. Then, we

used the results obtained via DT and to estimate the level. For the estimation, we used the algorithms: WLR, SVR and SVRmin, instead of selecting the best, we proposed the MM, a model that is the junction of all 3 models previously suggested. With the MM we obtained RMSE 3.58 cm, against 13.06 cm of WLR, while SVM and SVMmin obtained RMSE of 6.28 cm and 6.14 cm respectively. Using MM algorithm reduces the RMSE by 57.93%, on average, if compared to the proposed other algorithms.

Thus, the use of FBG temperature sensors for level measurement in conjunction with the MM algorithm results in operational and economic benefits, adequate for classified areas, and an average RMSE of under 6 cm. Future work includes fluid classification and estimation of its fluid level interface in an oil storage tank under multiphase flow. Future works also include the sensor fusion of different sensor approaches for more accurate assessment of the interface level and fluid densities.

Thus, this work contributes to the advancement of research in the production of economical and easily constructed devices for measuring different fluids. For the setup simulation, seen in Section 3.1, we published the paper "Simulation of FBG Temperature Sensor Array for Oil Identification via Random Forest Classification", in ECSA 7. This paper will also be submitted to Special Issue "Selected Papers from the $7^th$ International Electronic Conference on Sensors and Applications" in journal Sensors, as a complementary result the bench setup seen in Section 3.2.1 will be added. In Section 3.2.2, a second setup was presented, this in turn resulted in the article "Machine Learning techniques for level estimation using with FBG temperature sensor array", for the journal Optical Fiber Technology/Elsevier, and is in the correction phase. As a complementary result the bench setup seen in section A will be added.

# Bibliography

AL-YASEEN, W. L.; OTHMAN, Z. A.; NAZRI, M. Z. A. Multi-level hybrid support vector machine and extreme learning machine based on modified k-means for intrusion detection system. *Expert Systems with Applications*, Elsevier, v. 67, p. 296–303, 2017. Citado na página 21.

ALI, J. et al. Random Forests and Decision Trees. *International Journal of Computer Science Issues*, v. 9, n. 5, p. 272–278, 2012. ISSN 1694-0784. Citado 2 vezes nas páginas 26 and 28.

AMEEN, O. F. et al. Graphene diaphragm integrated fbg sensors for simultaneous measurement of water level and temperature. *Sensors and Actuators A: Physical*, Elsevier, v. 252, p. 225–232, 2016. Citado na página 19.

BUKHARI, S. F. A.; YANG, W. Multi-interface level sensors and new development in monitoring and control of oil separators. *Sensors*, v. 6, n. 4, p. 380–389, 2006. ISSN 14248220. Citado na página 19.

CHEN, B. D. et al. Water wave energy harvesting and self-powered liquid-surface fluctuation sensing based on bionic-jellyfish triboelectric nanogenerator. *Materials Today*, v. 21, n. 1, p. 88–97, 2018. ISSN 18734103. Disponível em: <https://doi.org/10.1016/j.mattod.2017.10.006>. Citado na página 19.

CONSALES, M. et al. A Fiber Bragg Grating Liquid Level Sensor Based on the Archimedes' Law of Buoyancy. *Journal of Lightwave Technology*, v. 36, n. 20, p. 4936–4941, 2018. ISSN 07338724. Citado na página 19.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 27.

DIAZ, C. A. et al. Optical fiber sensing for sub-millimeter liquid-level monitoring: a review. *IEEE Sensors Journal*, v. 19, n. 17, p. 7179–7191, 2019. ISSN 15581748. Citado 3 vezes nas páginas 19, 23, and 24.

DIAZ, C. A. R. et al. Temperature cross-sensitivity compensation in liquid level sensor using Mach-Zehnder interferometers. In: . [S.l.]: SPIE-Intl Soc Optical Eng, 2019. p. 56. ISBN 9781510624702. ISSN 1996756X. Citado na página 24.

DJURHUUS, M. S. et al. Machine learning assisted fiber bragg grating-based temperature sensing. *IEEE Photonics Technology Letters*, IEEE, v. 31, n. 12, p. 939–942, 2019. Citado 2 vezes nas páginas 21 and 61.

GOLA, J. et al. Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels. *Computational Materials Science*, Elsevier, v. 160, n. January, p. 186–196, 2019. ISSN 09270256. Disponível em: <https://doi.org/10.1016/j.commatsci.2019.01.006>. Citado 2 vezes nas páginas 25 and 61.

GRATTAN, K.; SUN, T. Fiber optic sensor technology: an overview. *Sensors and Actuators A: Physical*, Elsevier, v. 82, n. 1-3, p. 40–61, 2000. Citado na página 23.

GU, P. et al. Narrowband light reflection resonances from waveguide modes for high-quality sensors. *Nanomaterials*, v. 10, n. 10, p. 1–11, 2020. ISSN 20794991. Citado na página 19.

HALUNGA, A. G.; ORME, C. D.; YAMAGATA, T. A heteroskedasticity robust breusch–pagan test for contemporaneous correlation in dynamic panel data models. *Journal of econometrics*, Elsevier, v. 198, n. 2, p. 209–230, 2017. Citado 2 vezes nas páginas 30 and 55.

HILL, K. O. et al. Photosensitivity in optical fiber waveguides: Application to reflection filter fabrication. *Applied Physics Letters*, v. 32, n. 10, p. 647–649, 1978. ISSN 10773118. Citado na página 24.

HILL, K. O.; MELTZ, G. Fiber Bragg grating technology fundamentals and overview. *Journal of Lightwave Technology*, v. 15, n. 8, p. 1263–1276, 1997. ISSN 07338724. Citado 2 vezes nas páginas 23 and 24.

HJERTAKER, B. T.; JOHANSEN, G. A.; JACKSON, P. Recent developments in hydrocarbon separator interface imaging. *Process Imaging for Automatic Control*, v. 4188, p. 81, 2001. ISSN 0277-786X. Citado na página 19.

HSIEH, F. Y.; BLOCH, D. A.; LARSEN, M. D. A simple method of sample size calculation for linear and logistic regression. *Statistics in medicine*, Wiley Online Library, v. 17, n. 14, p. 1623–1634, 1998. Citado na página 25.

HU, J. S. et al. GP73, a resident Golgi glycoprotein, is sensibility and specificity for hepatocellular carcinoma of diagnosis in a hepatitis B-endemic Asian population. *Medical Oncology*, v. 27, n. 2, p. 339–345, 2010. ISSN 13570560. Citado na página 54.

INCROPERA, F. P. et al. *Fundamentals of heat and mass transfer*. [S.l.]: Wiley, 2007. Citado 2 vezes nas páginas 33 and 34.

ING, E. et al. Support Vector Machines and logistic regression to predict temporal artery biopsy outcomes. *Canadian Journal of Ophthalmology*, Elsevier Inc., v. 54, n. 1, p. 116–118, 2019. ISSN 17153360. Disponível em: <https://doi.org/10.1016/j.jcjo.2018.05.006>. Citado 2 vezes nas páginas 25 and 61.

JIA, Z. et al. Pipeline leakage identification and localization based on the fiber Bragg grating hoop strain measurements and particle swarm optimization and support vector machine. *Structural Control and Health Monitoring*, v. 26, n. 2, p. 1–23, 2019. ISSN 15452263. Citado 2 vezes nas páginas 25 and 27.

JIA, Z. et al. The indoor localization and tracking estimation method of mobile targets in three-dimensionalwireless sensor networks. *Sensors (Switzerland)*, v. 15, n. 11, p. 29661–29684, 2015. ISSN 14248220. Citado 2 vezes nas páginas 19 and 61.

JIA, Z. et al. Pipeline abnormal classification based on support vector machine using FBG hoop strain sensor. *Optik*, Elsevier, v. 170, n. January, p. 328–338, 2018. ISSN 00304026. Disponível em: <https://doi.org/10.1016/j.ijleo.2018.05.103>. Citado na página 25.

JOHANSEN, S. The welch-james approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, Oxford University Press, v. 67, n. 1, p. 85–92, 1980. Citado 3 vezes nas páginas 25, 29, and 30.

JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. [S.l.]: John Wiley & Sons, 2013. v. 398. Citado na página 27.

KAPANY, N. S.; EYER, J. A.; KEIM, R. E. Fiber optics. part ii. image transfer on static and dynamic scanning with fiber bundles. *JOSA*, Optical Society of America, v. 47, n. 5, p. 423–427, 1957. Citado na página 23.

KAWASAKI, B. S. et al. Narrow-band Bragg reflectors in optical fibers. *Optics Letters*, v. 3, n. 2, p. 66, 1978. ISSN 0146-9592. Citado na página 24.

KHOSRAVI, M.; AZIZIAN, S.; BOUKHERROUB, R. Efficient oil/water separation by superhydrophobic CuxS coated on copper mesh. *Separation and Purification Technology*, Elsevier, v. 215, n. January, p. 573–581, 2019. ISSN 18733794. Disponível em: <https://doi.org/10.1016/j.seppur.2019.01.039>. Citado na página 20.

LEAL-JUNIOR, A. et al. Material features based compensation technique for the temperature effects in a polymer diaphragm-based FBG pressure sensor. *Optics Express*, v. 26, n. 16, p. 20590, 2018. ISSN 1094-4087. Citado 3 vezes nas páginas 19, 20, and 24.

LEAL-JUNIOR, A. G. et al. FBG-Embedded Oblong Diaphragms with Extended Dynamic Range. *IEEE Sensors Letters*, Institute of Electrical and Electronics Engineers (IEEE), v. 2, n. 2, p. 1–4, 4 2018. ISSN 2475-1472. Citado na página 24.

LEAL-JUNIOR, A. G. et al. Multi-interface level in oil tanks and applications of optical fiber sensors. *Optical Fiber Technology*, Academic Press Inc., v. 40, p. 82–92, 1 2018. ISSN 10685200. Citado 2 vezes nas páginas 19 and 20.

LEAL-JUNIOR, A. G. et al. Simultaneous measurement of axial strain, bending and torsion with a single fiber bragg grating in CYTOP fiber. *Journal of Lightwave Technology*, v. 37, n. 3, p. 971–980, 2019. ISSN 07338724. Citado 2 vezes nas páginas 36 and 37.

LEE, T.-H.; ULLAH, A.; WANG, R. Bootstrap aggregating and random forest. In: *Macroeconomic Forecasting in the Era of Big Data*. [S.l.]: Springer, 2020. p. 389–429. Citado na página 26.

LI, J. et al. Adaptive swarm cluster-based dynamic multi-objective synthetic minority oversampling technique algorithm for tackling binary imbalanced datasets in biomedical data classification. *BioData Mining*, BioData Mining, v. 9, n. 1, p. 1–15, 2016. ISSN 17560381. Disponível em: <http://dx.doi.org/10.1186/s13040-016-0117-1>. Citado na página 42.

LIAW, A.; WIENER, M. et al. Classification and regression by randomforest. *R news*, v. 2, n. 3, p. 18–22, 2002. Citado na página 32.

LUO, F. et al. Fiber optic microbend sensor for distributed sensing application in the structural strain monitoring. *Sensors and Actuators, A: Physical*, v. 75, n. 1, p. 41–44, 1999. ISSN 09244247. Citado na página 23.

LV, J. et al. Research on new FBG displacement sensor and its application in Beijing Daxing Airport project. *Optik*, Elsevier, v. 178, n. 122, p. 146–155, 2019. ISSN 00304026. Disponível em: <https://doi.org/10.1016/j.ijleo.2018.09.117>. Citado 2 vezes nas páginas 19 and 20.

MAGNO, M. et al. Wake-up radio receiver based power minimization techniques for wireless sensor networks: A review. *Microelectronics Journal*, Elsevier, v. 45, n. 12, p. 1627–1633, 2014. ISSN 00262692. Disponível em: <http://dx.doi.org/10.1016/j.mejo.2014.08.010>. Citado na página 19.

MAJEED, A. F.; MURDAS, I. A. Modern System for Blood Pressure and Temperature Monitoring Based on FBG Sensor. *NICST 2019 - 1st Al-Noor International Conference for Science and Technology*, IEEE, n. Nicst, p. 7–13, 2019. Citado na página 20.

MANIE, Y. C. et al. Using a machine learning algorithm integrated with data de-noising techniques to optimize the multipoint sensor network. *Sensors (Switzerland)*, v. 20, n. 4, 2020. ISSN 14248220. Citado na página 21.

MANIE, Y. C. et al. Intensity and wavelength division multiplexing FBG sensor system using a Raman amplifier and extreme learning machine. *Journal of Sensors*, v. 2018, 2018. ISSN 16877268. Citado 2 vezes nas páginas 21 and 61.

MARTINS, J. et al. Low-Cost and High-Performance Optical Fiber-Based Sensor for Liquid Level Monitoring. *IEEE Sensors Journal*, v. 19, n. 13, p. 4882–4888, 2019. ISSN 1530437X. Citado na página 25.

MERIBOUT, M. et al. Interface layers detection in oil field tanks: a critical review. *Expert Systems for Human, Materials and Automation*, InTech, p. 181–208, 2011. Citado na página 19.

NIU, W.; HA, Y.; CHI, N. Novel Phase Estimation Scheme Based on Support Vector Machine for Multiband-CAP Visible Light Communication System. *Asia Communications and Photonics Conference, ACP*, IEEE, v. 2018-Octob, p. 1–3, 2018. ISSN 2162108X. Citado na página 25.

OU, Y. et al. Large WDM FBG Sensor Network Based on Frequency-Shifted Interferometry. *IEEE Photonics Technology Letters*, IEEE, v. 29, n. 6, p. 535–538, 2017. ISSN 10411135. Citado na página 24.

PŁAWIAK, P.; ABDAR, M.; ACHARYA, U. R. Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Applied Soft Computing Journal*, Elsevier B.V., v. 84, p. 105740, 2019. ISSN 15684946. Disponível em: <https://doi.org/10.1016/j.asoc.2019.105740>. Citado na página 25.

PRABHUGOUD, M.; PETERS, K. Bragg Grating Strain Sensors. *October*, v. 22, n. 10, p. 2302–2309, 2004. Citado 2 vezes nas páginas 36 and 37.

PRONIEWSKA, K.; PREGOWSKA, A.; MALINOWSKI, K. P. Identification of Human Vital Functions Directly Relevant to the Respiratory System Based on the Cardiac and Acoustic Parameters and Random Forest. *Irbm*, Elsevier Masson SAS, v. 1, p. 5–10, 2020. ISSN 18760988. Disponível em: <https://doi.org/10.1016/j.irbm.2020.02.006>. Citado 2 vezes nas páginas 26 and 28.

RAMEDANI, Z. et al. Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 39, p. 1005–1011, 2014. Citado na página 32.

RAO, Y. J. et al. In-fiber bragg-grating temperature sensor system for medical applications. *Journal of Lightwave Technology*, v. 15, n. 5, p. 779–784, 1997. ISSN 07338724. Citado 2 vezes nas páginas 19 and 20.

REIS, I.; BARON, D.; SHAHAF, S. Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets. *The Astronomical Journal*, IOP Publishing, v. 157, n. 1, p. 16, 2018. ISSN 0004-6256. Disponível em: <http://dx.doi.org/10.3847/1538-3881/aaf101>. Citado na página 28.

RODRIGUEZ-GALIANO, V. F. et al. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), v. 67, n. 1, p. 93–104, 2012. ISSN 09242716. Disponível em: <http://dx.doi.org/10.1016/j.isprsjprs.2011.11.002>. Citado 2 vezes nas páginas 25 and 28.

SARKAR, S. et al. An optimization-based decision tree approach for predicting slip-trip-fall accidents at work. *Safety Science*, Elsevier, v. 118, n. January, p. 57–69, 2019. ISSN 18791042. Disponível em: <https://doi.org/10.1016/j.ssci.2019.05.009>. Citado 3 vezes nas páginas 25, 26, and 27.

SCHROEDER, M. A.; LANDER, J.; LEVINE-SILVERMAN, S. Diagnosing and dealing with multicollinearity. *Western journal of nursing research*, SAGE PUBLICATIONS, INC. 2111 W. Hillcrest Dr., Newbury Park, CA 91320, v. 12, n. 2, p. 175–187, 1990. Citado na página 50.

SHI, T. M. et al. Capacitance-based instrumentation for multi-interface level measurement. *Measurement Science and Technology*, v. 2, n. 10, p. 923–933, 1991. ISSN 09570233. Citado na página 19.

SILVA, A. A. et al. Hollow-core negative curvature fibers for application in optical gas sensors. *Optical Engineering*, International Society for Optics and Photonics, v. 58, n. 7, p. 072011, 2019. Citado na página 23.

SILVA, J. C. C. da et al. Dynamic analysis and temperature measurements of concrete cantilever beam using fibre Bragg gratings. *Optics and Lasers in Engineering*, v. 45, n. 1, p. 88–92, 2007. ISSN 01438166. Citado 2 vezes nas páginas 20 and 36.

SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression. *Statistics and Computing*, v. 14, p. 199–222, 2004. ISSN 0960-3174. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=1CAD92EF8CCE726A305D8A41F873EEFC?doi=10.1.1.114.4288&rep=rep1&type=pdf%0Ahttp://download.springer.com/static/pdf/493/art%3A10.1023%2FB%3ASTCO.0000035301.49549.88.pdf?auth66=1408162706_8a28764ed0fae9>. Citado 4 vezes nas páginas 25, 29, 31, and 32.

SOCIETY, I. B. Best Subsets Logistic Regression Author ( s ): David W . Hosmer , Borko Jovanovic and Stanley Lemeshow Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2531779 REFERENCES Linked references are available on JSTOR for . v. 45, n. 4, p. 1265–1270, 2017. Citado na página 25.

SOHN, K. R.; SHIM, J. H. Liquid-level monitoring sensor systems using fiber Bragg grating embedded in cantilever. *Sensors and Actuators, A: Physical*, v. 152, n. 2, p. 248–251, 2009. ISSN 09244247. Citado na página 19.

SOLOMON, O. et al. Deep unfolded robust PCA with application to clutter suppression in ultrasound. *IEEE Transactions on Medical Imaging*, IEEE, v. 39, n. 4, p. 1051–1063, 2020. ISSN 1558254X. Citado na página 20.

SU, H.; CHEN, T.; ZHANG, Y. On the visual reality and monitoring of gantry cranes using FBG sensing. *Australian Journal of Mechanical Engineering*, Taylor & Francis, v. 00, n. 00, p. 1–8, 2019. ISSN 14484846. Disponível em: <https://doi.org/10.1080/14484846.2019.1602018>. Citado na página 21.

SWART, P. L. Long-period grating Michelson refractometric sensor. *Measurement Science and Technology*, v. 15, n. 8, p. 1576–1580, 2004. ISSN 09570233. Citado na página 23.

VORATHIN, E. et al. FBG water-level transducer based on pvc-cantilever and rubber-diaphragm structure. *IEEE Sensors Journal*, Institute of Electrical and Electronics Engineers Inc., v. 19, n. 17, p. 7407–7414, 9 2019. ISSN 15581748. Citado na página 24.

WANG, Y. et al. A quasi-distributed sensing network with time-division-multiplexed fiber bragg gratings. *IEEE Photonics Technology Letters*, IEEE, v. 23, n. 2, p. 70–72, 2011. ISSN 10411135. Citado na página 24.

WRIGHT, R. F. et al. Corrosion sensors for structural health monitoring of oil and natural gas infrastructure: A review. *Sensors (Switzerland)*, v. 19, n. 18, 2019. ISSN 14248220. Citado 2 vezes nas páginas 19 and 23.

XU, M. et al. Optical in-fibre grating high pressure sensor. *Electronics letters*, IET, v. 29, n. 4, p. 398–399, 1993. Citado na página 23.

YAO, P. et al. Fully hardware-implemented memristor convolutional neural network. *Nature*, Springer US, v. 577, n. 7792, p. 641–646, 2020. ISSN 14764687. Disponível em: <http://dx.doi.org/10.1038/s41586-020-1942-4>. Citado 2 vezes nas páginas 21 and 25.

YE, K. Key feature recognition algorithm of network intrusion signal based on neural network and support vector machine. *Symmetry*, v. 11, n. 3, 2019. ISSN 20738994. Citado 2 vezes nas páginas 25 and 61.

ZHANG, B.; KAHRIZI, M. High-Temperature Resistance Fiber Bragg Grating. *Sensors (Peterborough, NH)*, v. 7, n. 4, p. 586–591, 2007. Citado 3 vezes nas páginas 19, 20, and 24.

ZHU, J. Research on Classification of Fiber Intrusion Signal Based on Supported Vector Machines. n. 2, p. 163–167, 2014. Citado na página 25.

ZHU, L. et al. Forming a new small sample deep learning model to predict total organic carbon content by combining unsupervised learning with semisupervised learning. *Applied Soft Computing Journal*, Elsevier B.V., v. 83, p. 105596, 2019. ISSN 15684946. Disponível em: <https://doi.org/10.1016/j.asoc.2019.105596>. Citado na página 21.